

2013

EFFECTS OF POSITION AND ORIENTATION ON PAX6 UPSTREAM ENHANCERS IN *CIONA INTESTINALIS*

Thomas Andrew Bregnard
University of Rhode Island, thomas_bregnard@my.uri.edu

Follow this and additional works at: <https://digitalcommons.uri.edu/theses>

Terms of Use

All rights reserved under copyright.

Recommended Citation

Bregnard, Thomas Andrew, "EFFECTS OF POSITION AND ORIENTATION ON PAX6 UPSTREAM ENHANCERS IN *CIONA INTESTINALIS*" (2013). *Open Access Master's Theses*. Paper 166.
<https://digitalcommons.uri.edu/theses/166>

This Thesis is brought to you by the University of Rhode Island. It has been accepted for inclusion in Open Access Master's Theses by an authorized administrator of DigitalCommons@URI. For more information, please contact digitalcommons-group@uri.edu. For permission to reuse copyrighted content, contact the author directly.

EFFECTS OF POSITION AND ORIENTATION ON PAX6 UPSTREAM
ENHANCERS IN *CIONA INTESTINALIS*

BY

THOMAS ANDREW BREGNARD

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
CELL AND MOLECULAR BIOLOGY

UNIVERSITY OF RHODE ISLAND

2013

MASTER OF SCIENCE THESIS
OF
THOMAS ANDREW BREGNARD

APPROVED:

Thesis committee:

Major Professor

Steven Irvine

Niall Howlett

Becky Sartini

Nasser H. Zawia

DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND

2013

ABSTRACT

The classical definition of genetic *cis*-regulatory modules such as enhancers identifies these elements as insensitive to changes in position and orientation relative to the transcription start site (TSS). While this theory is well supported, some recent studies have uncovered examples of constraint in the position or orientation of particular enhancers. This study examines the position and orientation sensitivity of two upstream enhancer elements found in the gene *Pax6*, which regulates development of the eyes and central nervous system in vertebrates. Late embryonic stages of the organism used in this study, *Ciona intestinalis*, show expression of *Pax6* in the nerve cord and sensory vesicle.

For this investigation *Pax6* constructs were created in which the enhancers were repositioned or inverted relative to the TSS. These constructs were electroporated into embryos of *C. intestinalis*, and expression of the *Pax6-GFP* fusion gene was measured both by total fluorescence and by number of positive embryos. Alteration of either position or orientation of these enhancers was found to cause a strong decrease in measured *Pax6* expression. These results support the conclusion that the upstream enhancers must be in the proper location and arrangement to be fully functional, a finding that is consistent with the results of many other studies of *cis*-regulatory elements.

ACKNOWLEDGEMENTS

I would like to thank Dr. Steven Irvine for allowing me to work in his lab and for mentoring me throughout this research project. I would also like to thank Dr. Niall Howlett and Dr. Becky Sartini for serving on my thesis committee and for helping me to improve both my research goals and this thesis, as well as Dr. Hany Alashwal for agreeing to serve as the chairperson of my defense committee. In addition, I would like to acknowledge Paul Johnson and the rest of the staff at the Rhode Island Genomics and Sequencing Center, who assisted me with the use of equipment at the GSC and who performed DNA sequencing on the constructs used in this project. The GSC is supported in part by grants from the National Science Foundation and the U.S. Department of Agriculture and by the University of Rhode Island.

TABLE OF CONTENTS

| | |
|--|-----|
| ABSTRACT..... | ii |
| ACKNOWLEDGEMENTS..... | iii |
| TABLE OF CONTENTS..... | iv |
| LIST OF TABLES..... | v |
| LIST OF FIGURES..... | vi |
| INTRODUCTION..... | 1 |
| Enhancers and other <i>cis</i> -regulatory elements..... | 1 |
| The developmental gene <i>Pax6</i> | 5 |
| The model organism <i>Ciona intestinalis</i> | 7 |
| Hypothesis..... | 8 |
| METHODS..... | 11 |
| Production of test vectors..... | 11 |
| Collection and quantification of DNA..... | 13 |
| Fertilization and transformation of embryos..... | 13 |
| Embryo fixation..... | 15 |
| Photography of embryos..... | 16 |
| Data analysis..... | 16 |
| RESULTS..... | 20 |
| DISCUSSION..... | 36 |
| BIBLIOGRAPHY..... | 50 |

LIST OF TABLES

| | |
|--|----|
| Table 1. Quantification of embryos expressing <i>GFP</i> | 21 |
|--|----|

LIST OF FIGURES

| | |
|--|----|
| Figure 1. Conserved functional regions in <i>CiPax6</i> | 9 |
| Figure 2. Maps of experimental and intermediate vectors..... | 14 |
| Figure 3. Representative photographs from trials with each plasmid..... | 22 |
| Figure 4. Raw fluorescence values from whole embryo analysis..... | 25 |
| Figure 5. Normalized fluorescence of controls, whole embryo analysis..... | 26 |
| Figure 6. Normalized fluorescence of experimental embryos, whole embryo analysis..... | 28 |
| Figure 7. Raw fluorescence values from trunk based analysis..... | 32 |
| Figure 8. Normalized fluorescence of controls, trunk based analysis..... | 33 |
| Figure 9. Normalized fluorescence of experimental embryos, trunk based analysis..... | 34 |

INTRODUCTION

Enhancers and other cis-regulatory elements

The expression of genetic information as traits relies on a group of DNA elements collectively referred to as *cis*-regulatory elements. These DNA elements are located on the same strand of DNA as the genes they regulate; the function of these elements is to control the degree to which their target genes are expressed as well as when during development and in what cells these genes are expressed.

One type of *cis*-regulatory element is the enhancer, so named because it increases the expression of the target gene (reviewed in Atchison, 1988). An enhancer is typically found upstream from the gene it regulates, although some enhancers are found downstream of their target genes or within introns of the target genes.

Regulation of gene expression is accomplished through transcription factors (TFs), which are proteins that bind particular DNA elements in order to promote or suppress gene expression. Specific TFs bound to the enhancer element will form a complex with general TFs bound to the basal promoter, a DNA element that marks the start site for the process of transcription (Ptashne, 1986; Ptashne, 1988). This interaction of specific and general TFs will influence the extent to which the target gene is expressed.

In the 1980s a number of studies of enhancers and other *cis*-regulatory elements established the theory that these elements are flexible in their position and orientation relative to the transcription start site (TSS). These studies

examined enhancers from diverse systems, such as Moloney murine sarcoma virus (Laimins *et al.*, 1984), hepatitis B virus (Shaul *et al.*, 1985), yeast (Elion and Warner, 1986), and the mouse (Jaynes *et al.*, 1988). For each study the enhancers for particular genes were repositioned relative to the TSS, reversed in their directional orientation relative to the TSS, or both; the amount of expression of these genes was then measured and compared to the expression level using the default enhancer setup. These studies concluded that the enhancers still functioned in an altered position or orientation.

The results of studies such as these established the idea that enhancer function is unaffected by changes to the position and orientation of these elements, an idea that eventually became a criterion for establishing a DNA element as an enhancer. By the end of the same decade a mechanism had been discovered that explained the basis for this flexibility. In this model (Ptashne, 1986; Ptashne, 1988) the non-coding DNA between an enhancer and the basal promoter forms a loop that brings these two elements into close proximity to one another. TFs bound to each of these elements are then in the correct position to interact and stimulate transcription. Since the size of the loop does not matter, a change in enhancer position would affect the size of the loop formed but would preserve crucial TF interactions. Similarly, a change in enhancer orientation would cause a slightly different loop structure to form but would not affect the interactions of TFs. Because the necessary TF interactions are preserved in this

model, the level of transcription is not affected by alterations to the enhancer position or orientation.

In many of these same early studies, however, the altered position and orientation of the enhancers did appear to reduce the level of expression of these genes. Expression from the murine muscle creatine kinase promoter dropped to approximately 20% when the upstream enhancer was either moved to a position downstream of the reporter gene or inverted (Jaynes *et al.*, 1988). Similarly, an enhancer from the long terminal repeat of Moloney murine sarcoma virus was found to be only 30 to 40% active when moved downstream of the reporter gene (Laimins *et al.*, 1984). An enhancer from hepatitis B virus (HBV) gave values in the range of 50 to 300 in the forward orientation as measured by a chloramphenicol acetyltransferase (CAT) assay, while expression using the inverted enhancer ranged from 25 to 50 (Shaul *et al.*, 1985). Such results indicate that while an enhancer may still be able to function if its position or orientation is affected, such alterations to the enhancer do reduce the level of expression to some extent. The reduction in enhancer activity found in these studies was generally overlooked at the time of these investigations.

In more recent years there have also been a few studies that have uncovered enhancers that do not behave in the traditional way. One such study (Kim *et al.*, 2008) examined two *cis*-regulatory elements in mice. One, the CSE2 element, is an enhancer that promotes expression of the TF-encoding gene *Peg3*. The other, the CSE1 element, is a silencer for two co-regulated genes, *Peg3* and

Usp29. Silencers differ from enhancers in that silencers suppress expression of their target genes. Silencers also show the same position and orientation flexibility as enhancers. This study, however, revealed a near complete loss of regulation of the target gene *Peg3* when either element was inverted. Inversion of the enhancer CSE2 caused a drop in expression from 1.7 fold to 1.1 fold. These numbers are in reference to the 1.0 fold expression observed when CSE2 was deleted from the DNA. In the case of the silencer CSE1, inversion caused expression to increase to 2.5 fold. By comparison, deletion of CSE1 gave 2.7 fold expression of *Peg3*. The effect of CSE1 orientation on silencing of *Usp29* was not examined in this study.

Another study in the fruit fly *Drosophila melanogaster* (Small *et al.*, 1993) examined the expression of the *even-skipped* (*eve*) gene, which controls patterns of striping. The *eve* gene is controlled by several upstream enhancers. Correct expression of this gene was found to be dependent on a minimum length of sequence separating two of the upstream enhancers; various abnormalities result when they are placed too close together. It is important to point out that in the *eve* gene investigation the enhancers are constrained by their positions to each other rather than to the TSS.

In addition to functional studies, several sequence conservation studies have identified genes in which the position and orientation of enhancers are evolutionarily conserved. Sequence conservation typically indicates that a region of DNA has some important function, such as a gene or a *cis*-regulatory element. Similarly, if the orientation of an enhancer or its distance from the TSS is

conserved, this may indicate that the observed position or orientation is necessary for correct gene expression. One recent sequence conservation study (Goode *et al.*, 2005) compared human chromosomal region 7q36 to the genome of the pufferfish *Fugu rubripes* and identified a number of conserved non-coding sequence elements in which the position and orientation were conserved. In a subsequent functional assay, most of these conserved elements were found to function as enhancers.

The developmental gene Pax6

The gene *Pax6* plays an important role in development of the eyes and nervous system in both vertebrates and invertebrates (reviewed in Simpson and Price, 2002, and in Thompson and Ziman, 2011). The Pax6 protein encoded by this gene functions as a transcription factor. The Pax family TFs contain an N-terminal paired domain used in binding target genes and a C-terminal transactivation domain that mediates protein-protein interactions; Pax6 also includes a DNA-binding homeodomain (Thompson and Ziman, 2011). The functions of Pax6 are highly diverse, including roles in patterning of the central nervous system (CNS), control of cell migration, expression of cell surface adhesion molecules, and promoting both proliferation and differentiation of neurons in the eyes and CNS at different stages of development (Simpson and Price, 2002; Thompson and Ziman, 2011). The functions of Pax6 differ between cell types, due in part to at least two isoforms of the Pax6 protein with different

specificities for target genes (Simpson and Price, 2002; Thompson and Ziman, 2011).

Mutations in *PAX6* are associated with a variety of phenotypic abnormalities in humans. One of these is the autosomal dominant condition aniridia, reviewed in (Lee *et al.*, 2008). Aniridia is caused by mutations in *PAX6* that cause premature termination of translation; these are commonly nonsense mutations or frameshift insertions or deletions. The reduced level of functional *PAX6* protein results in improper development and maintenance of the cornea, lens, optic nerve, and other components of the eye. This condition can include aniridia-associated keratopathy, a corneal disorder that results in dry or red eyes, photophobia, increased watering of the eyes, and vision loss. Aniridia can also lead to glaucoma due to improper drainage. *PAX6* mutations have been implicated in other eye disorders such as keratitis, certain cataracts, and Peter's anomaly (Simpson and Price, 2002; Thompson and Ziman, 2011). Animal models with *Pax6* mutations show various eye and brain disorders such as an aniridia-like small eye, also known as microphthalmia (Simpson and Price, 2002; Thompson and Ziman, 2011). Abnormal phenotypes may result from loss of expression or overexpression of *Pax6*, indicating that proper *Pax6* function is concentration-dependent (Simpson and Price, 2002; Thompson and Ziman, 2011).

Pax6 was chosen as the gene of interest in this project for several reasons. Since the function of *Pax6* is widespread and very sensitive to concentration of the protein, any change in *Pax6* expression as a result of an altered *cis*-regulatory

module could cause a harmful phenotypic change. There have also been many studies of *cis*-regulation of *Pax6* and its homologs in various systems such as the mouse (Xu *et al.*, 2002; Kleinjan *et al.*, 2004; Kleinjan *et al.*, 2006; Kammandel *et al.*, 1999), fruit fly (Adachi *et al.*, 2003; Hauck *et al.*, 1999), and others (Griffin *et al.*, 2002; Kammandel *et al.*, 1999; Plaza *et al.*, 1999). These studies allow comparisons of effects on *cis*-regulation to be made across diverse species. In addition, the Irvine lab has previously identified the major *cis*-regulatory regions that appear to be necessary for *Pax6* expression in *Ciona intestinalis* (Irvine *et al.*, 2008).

The model organism Ciona intestinalis

The organism used in this study is the sea squirt *Ciona intestinalis*, discussed in (Satoh, 1994). Sea squirts, or ascidians, are ubiquitous in marine environments worldwide (Satoh, 1994). The adults are sessile filter feeders with a tube- or barrel-shaped body. The body has two tubular openings: an oral siphon that takes in food and oxygenated water, and an atrial siphon that expels waste and deoxygenated water (Satoh, 1994). *Ciona* is a hermaphroditic organism but is self-sterile; fertilization requires gametes from different individuals. The embryos of *C. intestinalis* include tailbud stages that resemble tadpoles; these embryos are the focus of this study.

C. intestinalis is an ideal choice for studying expression of transgenes for several reasons (reviewed in Satoh *et al.*, 2003). First, a simple electroporation method has been developed to introduce a single transgene into large numbers of

fertilized *C. intestinalis* eggs (Corbo *et al.*, 1997). In addition, a second draft of the *C. intestinalis* genome has been published (Dehal *et al.*, 2002). Comparison of sequences between *C. intestinalis* and the closely related species *C. savignyi* has proven to be an effective tool for locating conserved non-coding DNA elements such as enhancers (Johnson *et al.*, 2004; Irvine *et al.*, 2008; see Figure 1).

Furthermore, embryos of *C. intestinalis* are transparent and have a relatively small number of cells; the cells are large and easily distinguished under a microscope.

The developmental fates of all cells in the early embryonic stages of this organism have been well documented. This means that expression patterns of transgenes introduced by electroporation can be directly observed and measured at any of the various embryonic stages, and that these results can be collected in a cell- or tissue-specific manner.

Hypothesis

Since many previously studied enhancers show some sensitivity to changes in position or orientation, it is worth investigating whether this phenomenon is common among *cis*-regulatory DNA elements. This investigation focuses on two enhancer elements, named U_B and U_A, located approximately 1.8 kb upstream of the TSS of the *C. intestinalis Pax6* (*CiPax6*) gene; Figure 1 shows the locations of *cis*-regulatory elements identified in *CiPax6*. These two enhancers work synergistically to form a single *cis*-regulatory module. Here I present evidence that the U_B-U_A enhancer cassette loses much of its regulatory function when inverted or moved closer to the TSS. These results are similar to

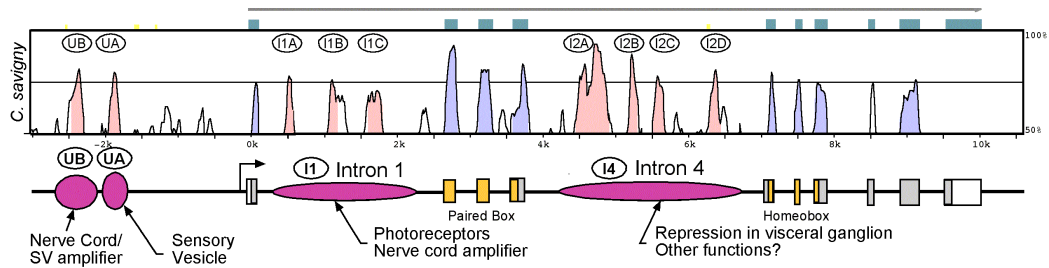


Figure 1. Conserved functional regions in *CiPax6*. A VISTA plot of conserved sequence regions between *Ciona intestinalis* and *Ciona savignyi* is shown above a diagram of major functional regions identified in the *CiPax6* gene. *CiPax6* contains four regions with experimentally confirmed *cis*-regulatory functions, shown in this diagram as pink ovals. The upstream U_B and U_A enhancer elements promote expression of *Pax6* in the nerve cord and in the sensory vesicle of the brain. A third enhancer within the first intron promotes expression in the photoreceptor cells and the nerve cord. The element within the fourth intron acts as a silencer. Image taken from (Irvine *et al.*, 2008).

observations from other enhancer elements, suggesting that this phenomenon of position and orientation sensitivity is more common in enhancers than previously recognized.

METHODS

Production of test vectors

Vector Lig 1 was produced from the experimental vector CiP6-2.5UI1 used and discussed in Irvine *et al.*, 2008. CiP6-2.5UI1 contains a portion of the *CiPax6* gene beginning 2.5 kb upstream of the TSS and ending within the first exon. This fragment includes the upstream U_B and U_A enhancer elements followed by approximately 1.8 kb of non-coding DNA separating the enhancers from the TSS. The intron 1 enhancer, which promotes expression in the nerve cord and photoreceptor cells (Irvine *et al.*, 2008), is also included in this vector. CiP6-2.5UI1 also contains a nuclear localization signal (NLS), *lacZ* reporter gene sequence, and a SV40 polyadenylation signal. Lig 1 was created by replacing the *lacZ* reporter sequence with *GFP* cDNA.

Lig 2 was intended to be a vector with the upstream enhancer cassette inverted in orientation. This vector was not successfully produced, however; therefore the discussion of this vector is skipped in this thesis.

Lig 3 was produced by digesting vector 1 with *ClaI* and *NruI* to remove a 1.15-kb piece of DNA from between the upstream enhancers and TSS. The piece of DNA removed from this clone was a piece of non-coding, non-conserved DNA that was believed to exclude any TF binding sites from the enhancers or the basal promoter. The *ClaI* overhang was blunted with T4 DNA polymerase, and the vector was ligated.

GW 1, GW 2, and GW 3 were produced using the Gateway recombination system (Invitrogen). These vectors were designed to be similar to Lig 1, Lig 2, and Lig 3, respectively; however, the vectors produced through the recombination method include *attB* recombination sites at both ends of the enhancer cassette. The production of comparable vectors by two different methods was used to check that any effects observed was not an artifact of the way in which the plasmids were made. Prior to the production of these vectors, the intermediate vector Aux 1 was produced. Aux 1 was prepared by digesting Lig 1 with *AscI* and *HindIII*, which removed a 19 base pair fragment immediately upstream of the enhancers. A Gateway recombination cassette was ligated into this site to produce Aux 1.

Next the entire vector except the enhancer region was amplified by PCR using primers that bind outside the enhancer region. This PCR product was then phosphorylated with T4 polynucleotide kinase and ligated. The product of this reaction, Aux 2, was used as the destination vector in the generation of vectors GW 1 and GW 2.

To produce GW 1, a copy of the enhancer cassette was first amplified with *attB* recombination sites at both ends. This product was used in a BP reaction with the vector pDONR221 (Invitrogen) to produce an entry vector with the enhancer cassette in the forward orientation. This entry vector was then used in a LR reaction with Aux 2, resulting in production of GW 1. GW 2 was produced

through a similar strategy, but the entry vector contained the enhancer cassette in the flipped orientation.

The destination vector for production of GW 3 was generated by digesting Aux 1 with *AscI* and *NruI*. This reaction removed a segment that was 1922 bases in length on the coding strand. The 5' overhang left by *AscI* was filled using T4 DNA polymerase, and the vector was ligated to produce Aux 3. This new destination vector was used in a LR reaction with the previously mentioned entry vector containing the enhancers in the forward orientation, resulting in the production of vector GW 3.

The upstream promoter regions of all plasmids tested in this investigation are diagrammed in Figure 2.

Collection and quantification of DNA

Plasmids were transformed into chemically competent TOP10 *E. coli*. DNA was harvested from overnight cultures by midiprep. For quantification, samples of known volumes of each plasmid were digested with *XhoI*, which cuts at a single site upstream of the enhancers. Concentrations of the linear vectors were determined on agarose gels using the GeneTools software from SynGene.

Fertilization and transformation of embryos

Adult animals were obtained locally from the Point Judith Marina at Snug Harbor, Rhode Island, or purchased from M-Rep, Carlsbad, California. Gametes were collected from adult animals by dissection, and fertilization was performed *in vitro*.

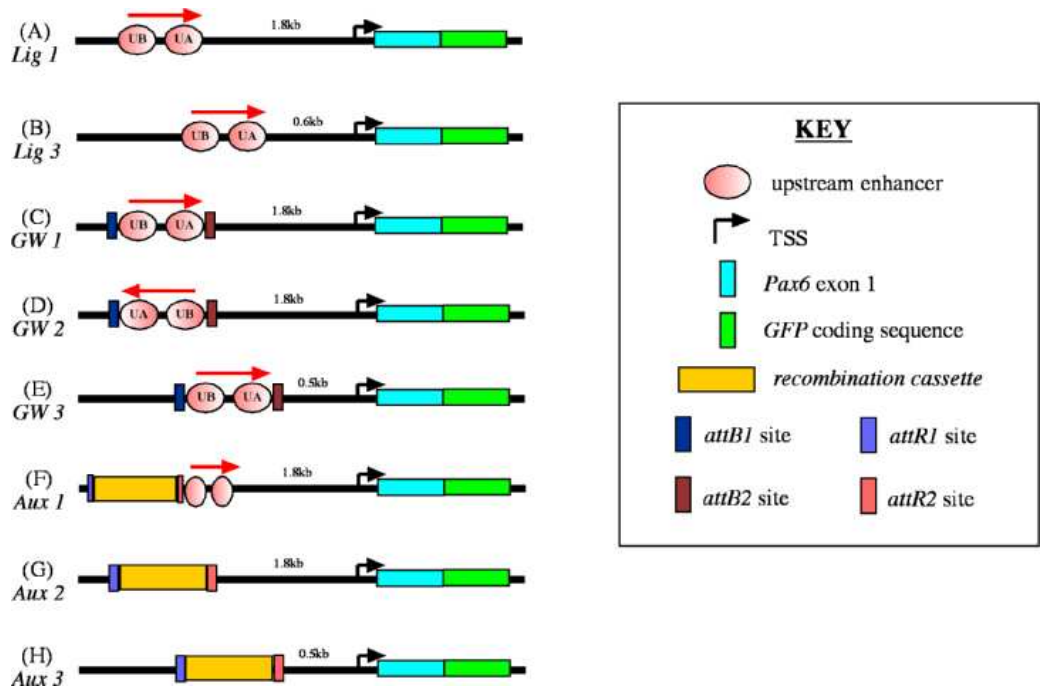


Figure 2. Maps of experimental and intermediate vectors. Relevant sequence elements are shown, beginning with enhancers and ending with *GFP* transgene. Distances represented are not drawn to scale. Arrows above enhancers represent orientation; right is forward and left is reversed. (A) Lig 1 is the positive control for ligation-generated vectors. (B) Lig 3 contains the upstream enhancers at a distance of 0.6 kb from the TSS. (C) GW 1 is the positive control for recombination-generated vectors. (D) GW 2 contains the enhancer cassette in the reversed orientation. (E) GW 3 contains the enhancer cassette at a distance of 0.5 kb from the TSS. (F-H) The Aux vectors were intermediates used to produce the tested vectors. They are described in the Methods section and are therefore included in this figure.

Dechoriation and electroporation were performed 10 minutes after the completion of fertilization. The protocol used was a modified version of a previous protocol used by the Irvine lab (Vierra and Irvine, 2012). Briefly, fertilized eggs were dechorionated using 0.4 mg/ml Pronase E in 1% sodium thioglycolate in 3.3% sodium chloride pH 10.1 for 3 to 5 minutes at 18°C. Dechoriation was quenched by transferring eggs to 1 mM glycine in phosphate buffered saline (PBS). For each electroporation, 250 µl of eggs in filtered seawater (FSW) were added to a solution containing 50 µg of DNA in 100 µl of water and 350 µl of 1 M mannitol. Electroporation was performed using a BTX ECM 830 square wave electroporator. The settings for all electroporations were 32 V, 100 ms. Actual voltage ranged from 20 V to 30 V, with nearly all samples electroporated at 27 V or 30 V; all electroporation times were 100 ms. Following electroporation, embryos were raised in gelatin-coated dishes of FSW containing approximately 15 U/ml penicillin and 15 µg/ml streptomycin. The rearing temperature used was 14°C.

Embryo fixation

All embryos were fixed for analysis at late tailbud stage 1 or 2, around 18 to 22 hours post fertilization when raised at 14°C. Embryos were fixed for 10 min in 2% paraformaldehyde in the dark. After removal of fixative, embryos were incubated in 100 mM glycine in PBS for 30 min to reduce autofluorescence of muscle cells. Glycine was then washed out three times with either PBS or PTw (PBS + Tween 20), and embryos were left in PTw until the time of analysis.

Photography of embryos

In preparation for analysis, embryos were incubated in PBS for 30 min; this PBS served as the mounting medium for microscopy. Slides were prepared by placing two layers of Scotch tape on each side of the slide, to the left and right of where the embryos would be placed. The two layers of tape served the purpose of preventing the embryos from being crushed when a cover slip was placed over them. After the tape was added, the surfaces of the slides and tape were greased with Rain-X except in the middles, where the embryos were to be placed. The slides were then cleaned with Windex. Embryos were mounted in a small drop of PBS between the taped areas of each slide. A cover slip was then laid over the embryos such that the edges of the cover slip were held up by the tape.

Embryos were viewed by epifluorescence illumination microscopy. Each embryo was photographed under a GFP filter using a SPOT Flex digital camera (Diagnostic Instruments). The image setting for all embryos were as follows: fluorescence as the image type, manual exposure, and 251.6 ms as the exposure time. All other settings were left as default. Embryos were photographed within three days after fixation.

Data analysis

All embryos were visually scored for expression or lack of expression of the reporter gene *GFP* using the following definitions. A positive score means that the embryo showed fluorescence in the sensory vesicle above the normal level of autofluorescence, fluorescent nuclei along the nerve cord, or fluorescence

in both locations. A negative score means that no increase in fluorescence was detected in the nerve cord or sensory vesicle. A score of ambiguous means that there appeared to be a slight increase in fluorescence observed in the sensory vesicle or increased fluorescence in the nerve cord without confirmed nuclear localization, but that the observed fluorescence was not definitely attributable to GFP.

Embryos were also measured for green fluorescence using the program ImageJ. For this analysis embryo photographs were first converted to grayscale in Adobe Photoshop. These images were too dim to be visible, so the white input level was reduced from 255 to 30 for each photograph using Photoshop. The adjusted images were then opened in ImageJ. For each photograph four measurements were performed. The first measurement was of the entire embryo or the trunk portion of the embryo, as specified in the results of each analysis. The next three measurements were square sections of the background fluorescence. All measurements included the values of area, mean gray value, and integrated density. Integrated density is the sum of the gray values of all pixels in the measured area; mean gray value is the average intensity per pixel in the measured area.

Whole embryo measurements were corrected for autofluorescence using a set of negative control embryos photographed the same number of days after fixation as each experimental or positive control embryo. The negative control for each trunk measurement was a section of autofluorescence from the trunk of the

same photograph. The mean gray value for the negative control was subtracted from the mean gray value of the embryo being measured to give the corrected density. These corrected density values were averaged for all embryos in each trial. Corrected mean density values are presented as both raw values and normalized values. Vector Lig 3 was normalized against Lig 1, while vectors GW 2 and GW 3 were normalized against GW 1; normalized values are given as percentages of expression relative to Lig 1 or GW 1 from the same set of trials.

Statistical analysis was performed on all raw data using Student's t-test. For the statistical analysis each experimental trial was compared to the positive control from the same trial set; the Lig 1 and GW 1 trials were also compared. All t-tests were 2-tailed and assumed unequal variance. The assumption of unequal variance was made because there was variability in the observed GFP fluorescence and autofluorescence between trials. The F-test for each statistical comparison was also performed to examine the variance. F-values were calculated offline, then p-values were calculated using the Free p-Value Calculator for an F-Test at (<http://www.danielsoper.com/statcalc3/calc.aspx?id=7>). Calculated p-values for the F-test ranged from 0.019 to 0.495.

For each comparison using the t-test, the p value was calculated using the online T-Test Statistics Calculator at (<http://studentsttest.com/>). Results of the t-test are presented on each graph. The definitions of these statistical results are as follows. A dash (-) represents a p value of 0.05 or greater. One star (*) represents a p value between 0.05 and 0.005. Two stars (**) represents a p value between

0.005 and 0.0005. Three stars (***) represents a p value less than 0.0005. For this investigation a p-value of less than 0.05 is considered statistically different.

RESULTS

Constructs in which the position or orientation of upstream enhancers has been altered show visibly reduced expression of *GFP* in *C. intestinalis* embryos as compared to the positive control constructs. Table 1 shows the number and percentage of embryos from each trial that were scored as positive, ambiguous, and negative. Lig 1 and GW 1 are positive controls for the ligation and recombination methods of vector production, respectively. In two separate trials 100% of embryos containing Lig 1 were positive; GW 1 yielded 93% positive and 7% ambiguous embryos in one trial and 100% positive embryos in another. All positive control embryos showed strong *GFP* expression in the sensory vesicle, and some showed nerve cord expression from either positive control; embryos containing Lig 1, but not GW 1, also showed non-nuclear expression in muscle cells in the tail. The expression in the tail muscle cells is assumed to be ectopic, since all plasmids used in this investigation encode *GFP* with a nuclear localization signal. Figure 3 contains representative photographs of embryos electroporated with each construct. The two ambiguous embryos containing GW 1 show slightly higher fluorescence in the sensory vesicle than the no-DNA controls, and although they cannot be conclusively scored as positive, they are predicted to be expressing *GFP* at a low level. These high percentages of positive embryos confirm that both of the positive control vectors drive expression of *GFP*.

| Construct | n | # (+) | % (+) | # A | % A | # (-) | % (-) |
|-----------|----|-------|-------|-----|-----|-------|-------|
| Lig 1-1 | 11 | 11 | 100 | 0 | 0 | 0 | 0 |
| Lig 1-2 | 3 | 3 | 100 | 0 | 0 | 0 | 0 |
| Lig 3-1 | 23 | 7 | 30 | 6 | 26 | 10 | 43 |
| Lig 3-2 | 5 | 1 | 20 | 1 | 20 | 3 | 60 |
| GW 1-1 | 30 | 28 | 93 | 2 | 7 | 0 | 0 |
| GW 1-2 | 26 | 26 | 100 | 0 | 0 | 0 | 0 |
| GW 2-1 | 26 | 0 | 0 | 3 | 12 | 23 | 88 |
| GW 2-2 | 9 | 2 | 22 | 4 | 44 | 3 | 33 |
| GW 3-2 | 10 | 0 | 0 | 5 | 50 | 5 | 50 |
| (-) C-2 | 12 | 0 | 0 | 0 | 0 | 12 | 100 |

Table 1. Quantification of embryos expressing *GFP*. All embryos were scored as positive [(+)], negative [(-)], or ambiguous (A) for *GFP* expression. Scores for each trial are given as number (#) and percentage (%) of embryos in each category; *n* refers to the total number of embryos in each trial. Trials are labeled with the name of the plasmid used, followed by a hyphen, then the number of the trial set; this trial naming system is also used in certain places in the text. All trials with the same trial set number were performed simultaneously. (-) C-2 is the no-DNA negative control used in trial set 2.

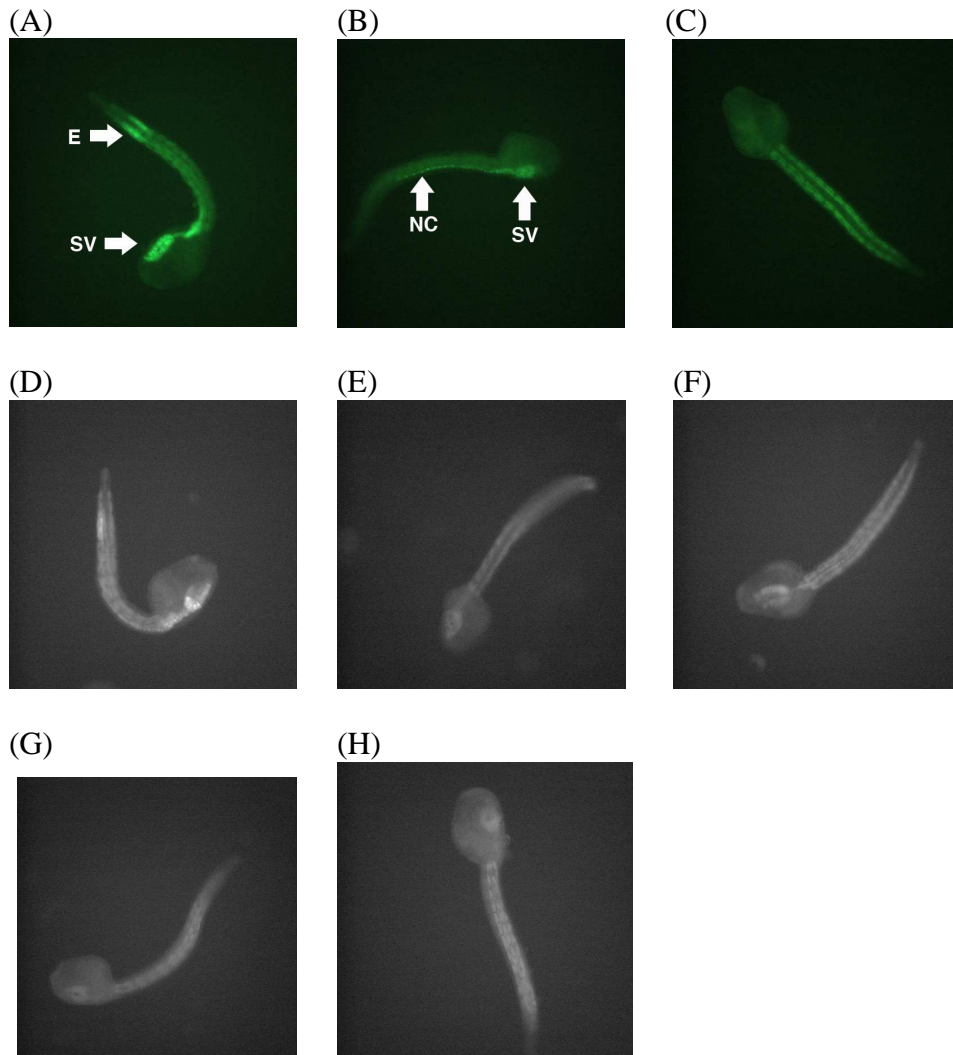


Figure 3. Representative photographs from trials with each plasmid. All images have been scaled to 25% of the original height and width in Photoshop. (A-C) These photographs have been enhanced by decreasing maximum RGB input levels from 255 to 20 to reproduce the appearance under the microscope. Areas of *GFP* expression are indicated by arrows and the following labels: SV, sensory vesicle; NC, nerve cord; E, ectopic expression. (A) Lig 1 embryo; nerve cord expression is occurring but is not visible. (B) GW 1 embryo showing fluorescent nuclei in the nerve cord. (C) Negative control embryo containing no DNA; control for autofluorescence. (D-H) These images have been converted to grayscale, and the white input level has been reduced from 255 to 15. (D) Lig 1 embryo. (E) Lig 3 embryo. (F) GW 1 embryo. (G) GW 2 embryo. (H) GW 3 embryo.

Construct GW 2 contains the upstream enhancers in the flipped orientation. The results of two trials with this vector were somewhat mixed. The breakdown of embryos from one trial with GW 2 was 12% ambiguous and 88% negative, but a second trial gave 22% positive, 44% ambiguous, and 33% negative embryos. Most GW 2-containing embryos scored as ambiguous showed weak non-nuclear fluorescence in the nerve cord, which usually shows no autofluorescence. However, some control embryos containing no transgene showed similar fluorescence. These results for GW 2 indicate that the flipped enhancers are less effective at promoting *GFP* expression than the same enhancers in the default orientation. A lack of visible expression in most embryos, however, does not rule out the possibility of low levels of GFP that are invisible to the unaided eye.

Constructs Lig 3 and GW 3 have the upstream enhancers moved closer to the TSS. Lig 3 yielded 30% positive and 26% ambiguous embryos in one trial; a second trial yielded 20% positive and 20% ambiguous embryos. Most of the positive embryos were weakly positive; these data suggest that Lig 3 promotes a low level of *GFP* expression that is visible in some embryos but not in others. The data for GW 3 do not match those of Lig 3. One trial was performed with GW 3, in which 50% of embryos were scored as ambiguous and 50% were scored as negative. The ambiguous embryos showed non-nuclear fluorescence in the nerve cord only, which makes them more likely to be negative. GW 3 therefore does not

promote expression of *GFP* to a visible level using the methodology of this investigation.

There is a modest difference in measured *GFP* expression from GW 1 versus from Lig 1. Embryos were analyzed for fluorescence using ImageJ, as described in the Methods section of this thesis. For trial set 2 the negative control was a set of embryos that were fertilized, dechorionated, and fixed alongside these trials, but were not electroporated and did not receive any foreign DNA. A no-DNA control was not available for trial set 1, so Lig 2-1 was used as the negative control for this set of trials. This decision was made because Lig 2-1 gave the lowest corrected mean density value and because nearly all embryos in this trial were scored as negative for visible *GFP* expression.

Figure 4 shows corrected mean density values for all whole embryo trials performed. In trial set 1, Lig 1 gave a mean value of 899 ± 526 while GW 1 gave a mean value of 647 ± 632 . The values from trial set 2 were 1394 ± 126 for Lig 1 and 998 ± 614 for GW 1. This increased fluorescence in trial set 2 relative to trial set 1 was, in general, consistent among individual positive control embryos. Normalized values and statistical analysis for the positive controls are shown in Figure 5. In both trial sets the normalized mean value of GW 1 is approximately 72% that of Lig 1 in spite of the differences in raw mean values between trial sets. This difference was determined to be statistically significant for trial set 2 ($p = 0.012$) but not for trial set 1 ($p = 0.213$).

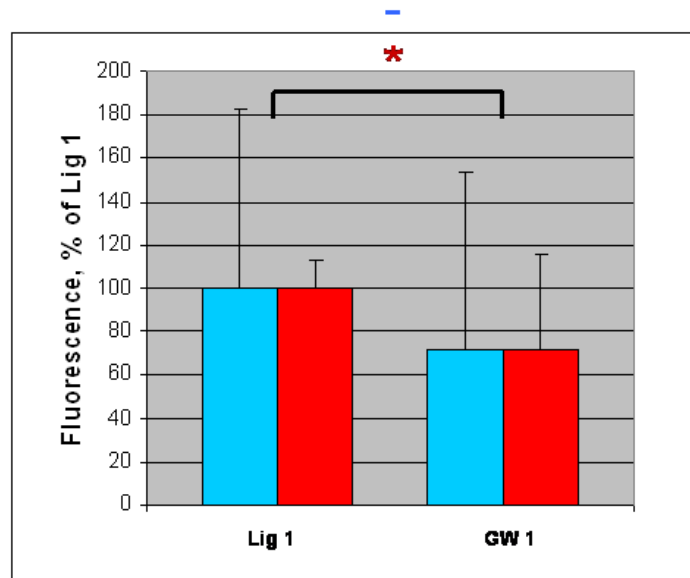


Figure 5. Normalized fluorescence of controls, whole embryo analysis.

Corrected mean density values of the positive control trials from the whole embryo analysis were normalized against Lig 1. For each construct values from trial set 1 are given in blue, and values from trial set 2 are given in red. Results of t-tests follow the same color coding. The statistical results shown in this graph were taken from the raw data, not the normalized values.

Error bars represent a combined standard deviation for the trial and the control against which it was normalized. For the combined standard deviation the standard deviations of the individual trials were treated as random error and propagated as follows. Standard deviations were first converted to percentages of the trial mean. These percent error values were squared, and the two squares were added. The square root of this sum was then taken to get the relative standard deviation as a percentage of the normalized mean. This percentage was then multiplied by the normalized mean to give the combined standard deviation.

All experimental constructs yield lower levels of *GFP* expression than positive control constructs as measured by ImageJ. Figure 4 shows that all experimental constructs show lower corrected mean density values than the two positive controls. Normalized values for the experimental trials are given in Figure 6. The two trials with GW 2 gave mixed results. GW 2-1 gave a mean value of 460 ± 632 , which is approximately 71% of the mean value for GW 1-1. The mean value of GW 2-2 was measured as -546 ± 442 . This mean value was displayed as zero in Figure 4; the standard deviation was not changed. The t-test identified the difference between GW 2 and GW 1 as highly significant for trial set 2 ($p = 1.6 \times 10^{-7}$) but not significant for trial set 1 ($p = 0.274$). These GW 2 measurements do not correlate with the visual observations recorded in Table 1; no positive embryos were seen in GW 2-1, but GW 2-2 gave 22% weakly positive embryos. This difference is most likely due to lower average autofluorescence in GW 2-2. Autofluorescence varied between trials and between embryos within each trial. Most of this autofluorescence was naturally occurring green fluorescence in muscle cells in the tails of the embryos, although the trunks also showed some autofluorescence. In trials where there was very little or no *GFP* expression, some embryos gave negative corrected density values due to low autofluorescence. All nine embryos in GW 2-2 gave negative values, producing a negative mean value for the trial.

Removal of most of the non-coding sequence from between the enhancers and TSS also caused a loss of measured total fluorescence, although this loss was

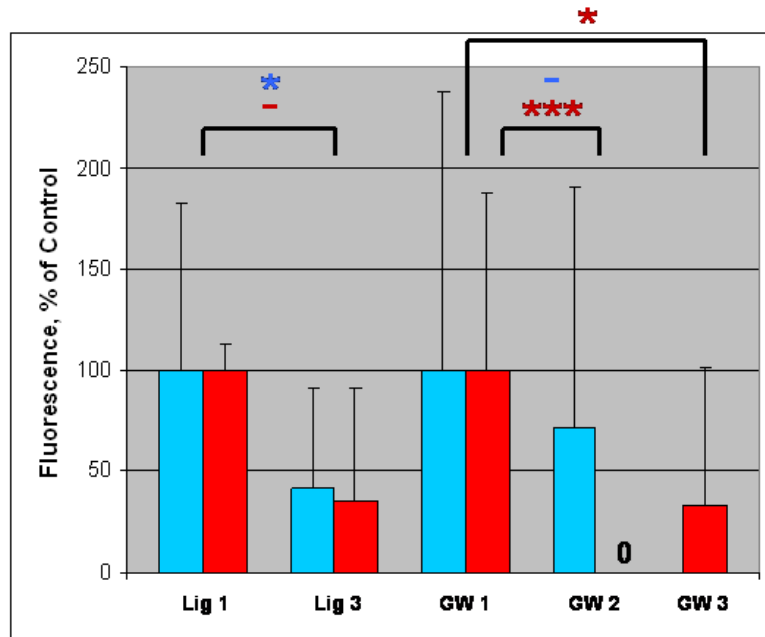


Figure 6. Normalized fluorescence of experimental embryos, whole embryo analysis. Corrected mean density values from Figure 4 were normalized against the corresponding positive control as described in the Methods section. Values are shown as a percentage of Lig 1 or GW 1 from the same trial set. Values from trial set 1 are given in blue, and values from trial set 2 are given in red. Results of t-tests follow the same color coding. The statistical results shown in this graph were taken from the raw data, not the normalized values. Trial GW 2-2 gave a negative value but is represented here as a zero on the chart. Error bars represent a combined standard deviation for the trial and the control against which it was normalized. See figure 5 for an explanation of how the combined standard deviations were calculated.

less dramatic than that caused by inversion of the enhancer cassette. Two trials with Lig 3 yielded raw mean values of 370 ± 389 and 486 ± 784 . These values are 41% of Lig 1-1 and 35% of Lig 1-2, respectively. This loss of expression is statistically significant for the first trial set ($p = 0.010$) but not the second trial set ($p = 0.060$). While these trials suggest a substantial decrease in *GFP* expression from Lig 3, the visual observations of these embryos suggest that the actual decrease in expression is greater. Trial GW 3-2 gave a raw score of 332 ± 650 , which is 33% the level of expression from GW 1-2. This result was statistically significant ($p = 0.013$). The relatively high score from this trial appears to be due mainly to autofluorescence, since no positive embryos were observed in this trial.

All trials showed rather high standard deviations, as shown in Figure 4. In general, embryos within each trial showed substantial differences in corrected mean density, although these values were more or less evenly distributed around the average. Two factors that contributed to this high variability in the data were autofluorescence and background fluorescence, both of which differed from embryo to embryo. Autofluorescence is mainly due to naturally occurring green fluorescence in muscle cells. Background fluorescence occurs in the mounting medium and can be due to certain fluorescent particles or compounds present around the embryo. The combined mean density of autofluorescence and background fluorescence was much larger than the mean density from the true fluorescent signal in all trials. Small differences in either or both of these sources of fluorescence could therefore cause large variations in the corrected mean

density values. Another source of error that affected all trials was the positioning of embryos. Not all embryos were positioned correctly for a single photograph to capture the whole embryo in focus. Parts of the embryo that were out of focus became blurry and gave a diffuse fluorescent signal, which resulted in a lower-than-expected fluorescence measurement. A final source of error that may have affected all embryos except the negative controls was mosaic expression of the transgene. Following electroporation, depending on how the transgenic plasmid is distributed in the egg, it will be partitioned unevenly between cells as the embryo develops. This will cause variation in how much DNA is present in regions that normally express *Pax6*.

Measurements of fluorescence in only the trunk portions of embryos give a clearer confirmation of loss of expression in all experimental embryos.

Most of the GFP fluorescence in these trials is found in the sensory vesicle of the brain (Irvine *et al.*, 2008; Figure 3), while the strongest autofluorescence is seen in the tail (Figure 3C). In order to reduce the impact of autofluorescence on the data collected, the photographs were also analyzed by measuring only the fluorescence in the trunk portions of the embryos. Measurements were taken in the same manner as in the whole embryo analysis except that a different method was used to correct for autofluorescence. In this analysis each embryo measurement was individually corrected for autofluorescence by measuring a portion of the trunk excluding the brain and using this mean density value as the control for autofluorescence. The corrected mean density scores from the trunk

based analysis are given in Figure 7. Normalized scores from this analysis are given for the positive controls in Figure 8 and for the experimental trials in Figure 9.

Using this method the corrected mean density values for Lig 1 were 1293 ± 520 and 2044 ± 474 in trial sets 1 and 2, respectively. The higher values seen in the trunk based analysis versus the whole embryo analysis are due to a combination of more concentrated GFP signal in the trunk and lower autofluorescence. The two trials with the control GW 1 gave corrected mean values of 837 ± 385 for trial 1 and 1229 ± 393 for trial 2. These values are 65% and 60% of the corresponding Lig 1 values, respectively. The difference in expression between the controls was statistically significant for trial 1 ($p = 0.019$) but not for trial 2 ($p = 0.086$).

As in the whole embryo analysis, the trunk-based analysis revealed lower fluorescence in all experimental constructs than in the two positive controls. GW 2 gave values of 104 ± 521 in trial set 1 and 657 ± 190 in trial set 2. These values are 12% and 53% of GW 1, respectively. This loss of expression is statistically highly significant for both trial set 1 ($p = 4.2 \times 10^{-7}$) and trial set 2 ($p = 3.3 \times 10^{-6}$). These results do not match up with the whole embryo results, for which trial 1 was measured at 71% of the control and trial 2 gave a negative value. The visual observations of positive GW 2 embryos agree with the results of the trunk-based analysis, not the whole embryo analysis. Lig 3 gave values of 517 ± 312 in trial 1 and 829 ± 272 in trial 2. The normalized values for this construct are 40% and

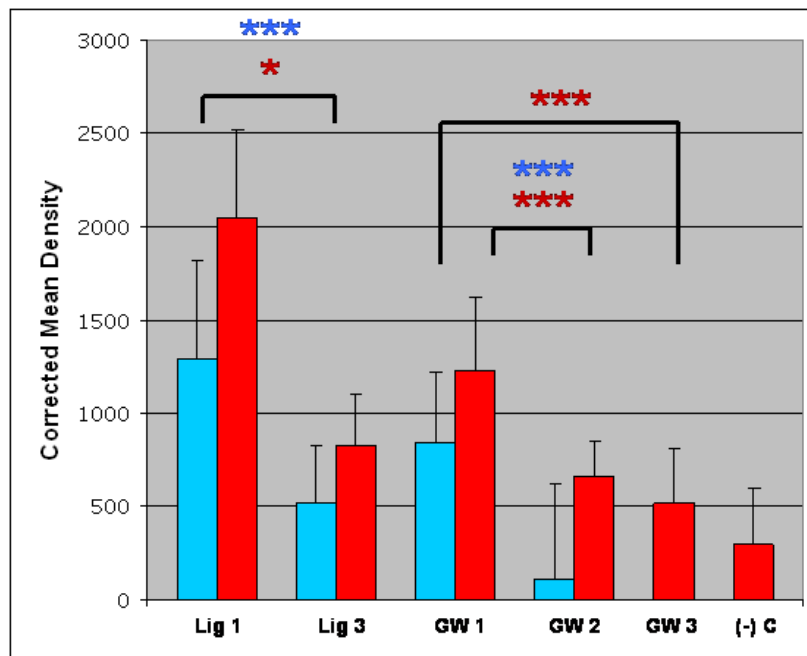


Figure 7. Raw fluorescence values from trunk based analysis. Raw corrected mean density values for all trials performed, determined for trunk portions of embryos. For each construct values from trial set 1 are given in blue, and values from trial set 2 are given in red. Results of t-tests follow the same color coding. The negative control from trial set 2 was included in the trunk based analysis and is labeled (-) C. Statistical analysis is not shown for the negative control in order to maintain readability. Error bars represent the standard deviation of each trial.

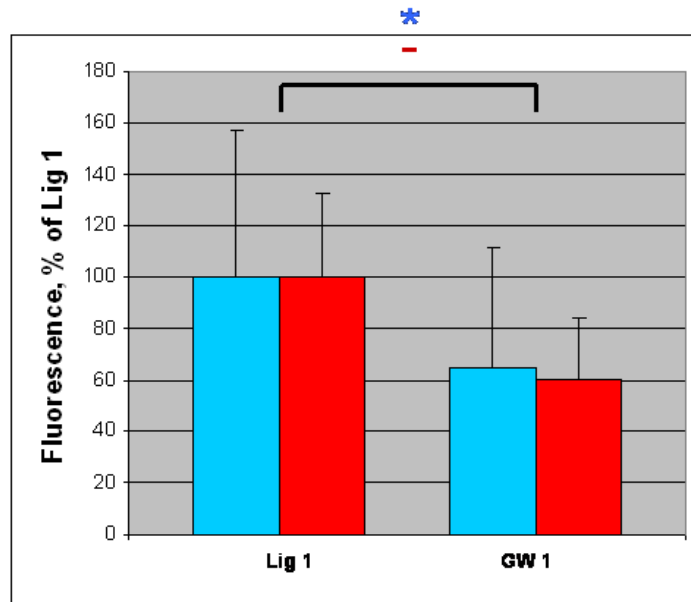


Figure 8. Normalized fluorescence of controls, trunk based analysis.

Corrected mean density values of both positive controls from the trunk-based analysis were normalized against Lig 1. Values from trial set 1 are given in blue, and values from trial set 2 are given in red. Results of t-tests follow the same color coding. The statistical results shown in this graph were taken from the raw data, not the normalized values.

Error bars represent a combined standard deviation for the trial and the control against which it was normalized. See figure 5 for an explanation of how the combined standard deviations were calculated.

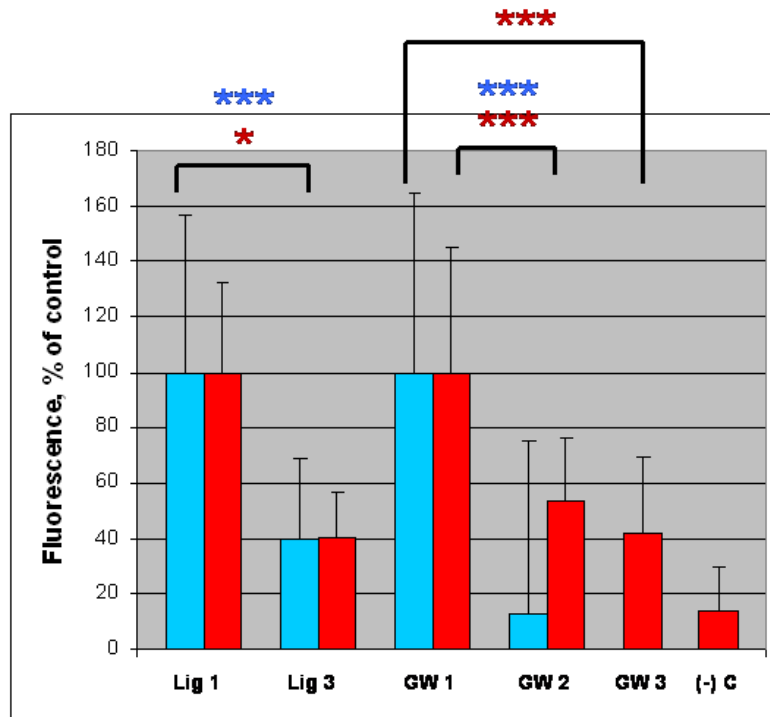


Figure 9. Normalized fluorescence of experimental embryos, trunk based analysis. Corrected mean density values for trunk based analysis were normalized against the corresponding positive control as described in the Methods section. Values are shown as a percentage of Lig 1 or GW 1 from the same trial set. Values from trial set 1 are given in blue, and values from trial set 2 are given in red. Results of t-tests follow the same color coding. The statistical results shown in this graph were taken from the raw data, not the normalized values. The negative control from trial set 2 is included and is labeled as (-) C. Statistical analysis is not shown for the negative control in order to maintain readability. Error bars represent a combined standard deviation for the trial and the control against which it was normalized. See figure 5 for an explanation of how the combined standard deviations were calculated.

41%, respectively. This loss of expression is highly significant for trial set 1 ($p = 4.8 \times 10^{-4}$) and significant for trial set 2 ($p = 0.030$). The trial with GW 3 gave a value of 515 ± 296 , which is 42% of GW 1-2. This result was statistically highly significant ($p = 6.6 \times 10^{-6}$). All of these values for Lig 3 and GW 3 match up fairly closely with the whole embryo results, further supporting the conclusion that moving the *Pax6* upstream enhancers closer to the TSS reduces expression to about 30 to 40% of the normal value.

The negative control set of embryos from trial set 2 was also analyzed by the trunk-based method; these results appear in Figure 7 and Figure 9. The raw score for this negative control set was 290 ± 304 . This value comes out to 14% of Lig 1-2 or 24% of GW 1-2; the measurement shown in Figure 9 is normalized against Lig 1-2. This result was statistically very significant based on the t-test result ($p = 3.3 \times 10^{-9}$). For a no-DNA control this value is somewhat high. The reason for this high value is that the embryos show higher autofluorescence in the brain than in the rest of the trunk, and the brain is intentionally excluded from the autofluorescence measurements. This means that the true GFP fluorescence values are likely to be lower than the measured values for all embryos.

DISCUSSION

***CiPax6* upstream enhancers show sensitivity to changes in their position and orientation.** Based upon the results of this study, changes in either the position of the U_B-U_A enhancer cassette or its orientation severely reduce or eliminate expression of *GFP*. The effect of altering the orientation of these enhancers is a substantial loss of expression; two trials using GW 2 gave normalized fluorescence values of 12% to 53% of the corresponding controls using the more reliable trunk based measurements. Repositioning the enhancer cassette 500 to 600 bp upstream of the TSS also severely reduces expression of *GFP*, but there is still visible expression in some embryos transformed with Lig 3. All three trials using either Lig 3 or GW 3 gave normalized values within the range of 33% to 42% of the appropriate positive control based on both whole embryo measurements and trunk based measurements.

The inability to avoid autofluorescence during the data analysis was a limiting factor in this study, particularly in the whole embryo analysis. The ideal solution to this problem would be to analyze only the sensory vesicle and nerve cord, thereby avoiding nearly all autofluorescence. However, these regions generally were not distinguishable unless there was very strong *GFP* expression. For this reason analysis of only the nerve cord and sensory vesicle would have required additional labeling of these regions in experimental and negative control embryos. This technique was not used in this investigation but could be applied to future analyses.

One issue that should be addressed from these results is the observation that there was a consistent difference in *GFP* expression between the positive controls Lig 1 and GW 1. GW 1 gave fluorescence values that were approximately 70% of Lig 1 fluorescence based on whole embryo measurements or 65% based on trunk measurements. The only sequence differences between these two vectors are a single base deletion in Lig 1 and differences at the 5' and 3' ends of the enhancer cassette. Differences at the 5' and 3' ends include remnants of the recombination cassette used in the Gateway method, including an *attB1* site at the 5' end and an *attB2* site at the 3' end; as well as a 12 bp stretch of DNA found in Lig 1 that is missing from GW 1. The 5' difference does not affect the region previously identified as enhancer sequence. In addition, Lig 1 gives the highest level of expression of all vectors tested, so the single base deletion found in this vector cannot be responsible for the loss of expression seen in the other constructs. Therefore the difference at the 3' end of the enhancer cassette appears to be the only explanation for the reduction in expression from GW 1. The locations of the two enhancers were previously determined by a series of PCR-based deletions from the 5' end of the upstream sequence (Irvine *et al.*, 2008). The study did not make use of 3' deletions to determine the TSS-proximal end of the enhancer cassette; instead this 3' boundary was identified from the minimum sequence whose deletion eliminated detectable expression of *GFP*. This leaves the possibility that the real 3' edge of the U_A element is actually downstream of the predicted position, and that this element has consequently been interrupted in

GW 1 as well as all experimental constructs. The Irvine lab is currently working on the production of new experimental constructs that retain additional sequence at the 3' end of the enhancer cassette. Another possibility, discussed more in the next section of this discussion, is that the sequences flanking the enhancer cassette somehow modify the activity of the enhancers (Elion and Warner, 1986). While this may explain the modest loss of expression in GW 1, there appears to be an additional and much greater loss of expression when the position or orientation of the enhancers is also altered.

Other examples of position or orientation sensitivity have been documented but have often been overlooked. Many studies, including those referenced in this thesis, have found some loss of enhancer activity when the position or orientation of an enhancer element is altered. One clear example of orientation sensitivity already mentioned in the introduction was the study of two *cis*-regulatory elements of the *Peg3* and *Usp29* bidirectional promoter (Kim *et al.* 2008). The enhancer CSE2 consists of a series of binding sites for the TF YY1; the CSE1 element is a silencer for both *Peg3* and *Usp29*. Both of these elements are found within the first intron of *Peg3*. In this study the expression of a reporter gene was measured at 1.7 fold when both the CSE1 and CSE2 elements were present. The expression level was defined as 1.0 fold for a construct containing the silencer CSE1 but not the enhancer CSE2. Inversion of the entire CSE2 element gave a 1.1-fold level of expression; a similar value was observed when all YY1 sites were abolished by mutation. This was interpreted by the authors of the

article as orientation dependence of this element. The CSE1 element was also found to be orientation sensitive. A construct containing CSE2 but not CSE1 gave a 2.7-fold level of expression due to the loss of silencer function of CSE1 in the reverse orientation. Inversion of CSE1 (with no alteration to CSE2) gave an expression level of 2.5 fold, suggesting that its silencer activity was nearly completely lost in the reverse orientation. The CSE1 element was also tested in a different location, 3' of the *Peg3*-differentially methylated region (DMR) where the enhancers are normally found. The expression with CSE1 in this new position was about 1.4 fold with CSE1 in the forward orientation and 1.8 fold with CSE1 in the reverse orientation, which showed that this silencer was fully functional, perhaps slightly more functional, in the altered position.

The study by Small *et al.* (1993) uncovered an example of extreme position sensitivity of two enhancers in the *eve* gene in *Drosophila*, as mentioned in the introduction. The *eve* gene controls the proper patterning of segmentation stripes in the *Drosophila* embryo. Stripes 2 and 3 are under the control of separate upstream enhancers; these two enhancers are separated by a 1.7 kb spacer of non-coding DNA. It was discovered that this 1.7 kb fragment could be removed and replaced with a different spacer without altering the stripe pattern produced. The spacer fragment used could even be of a different size than the natural 1.7 kb fragment; one fragment that preserved the stripe patterning was only 160 bp in size. When the two enhancers were linked, however, abnormal expression patterns resulted. Similar effects were seen when the positions of the stripe 2

enhancer and the stripe 3 enhancer were switched with or without a spacer, although the abnormal pattern seen with the linked enhancers was different depending on which enhancer was positioned first. This indicated that the normal stripe pattern was dependent on the separation of the two stripe enhancers by some spacer. The position sensitivity of these enhancers differs from other examples of position or orientation sensitivity in that the proper function of these enhancers is dependent on their distance from each other, not from the TSS. Nonetheless, this example illustrates the idea that enhancer position can reflect the need for the proper genomic context.

Many studies of *cis*-regulatory elements from the 1980s also uncovered some apparent sensitivity to changes in position or orientation. At this time, however, the definition of *cis*-regulatory elements as flexible in their position and orientation was newly established as a criterion for defining such an element as an enhancer. To satisfy this criterion, it was enough for a newly discovered enhancer to show that the enhancer in the altered setup still increased expression of the target gene to a significant extent. Many cases of substantial reduction in enhancer activity in these altered constructs were therefore ignored at the time. For example, the study by Jaynes *et al.* (1988) found that a muscle-specific enhancer located 1050 bp upstream of the muscle creatine kinase gene retained 100% of its enhancer activity when moved to a position 80bp upstream of the TSS, but expression dropped to around 20% when the enhancer was either moved downstream of the reporter gene chloramphenicol acetyltransferase (CAT) in

either orientation, or inverted in orientation at the -80 position. It should be noted that the constructs that gave this strong decrease in expression had deletions of all upstream sequence excluding the enhancer element to position -80. A set of constructs with the enhancers in the downstream position (in either orientation) and with upstream deletions only to position -776 gave between 42 and 74% expression. The study by Shaul *et al.* (1985) found that an enhancer for the hepatitis B virus core antigen (HBcAg) promoter gave levels of CAT activity ranging from 50 to 300 in the forward orientation, but the enhancer in the reverse orientation gave CAT activity ranging from 25 to 50. These numbers are normalized to a comparable construct using an enhancer from simian virus 40 (SV40), which was defined as an activity level of 100. Similarly, Laimins *et al.* (1984) found that an enhancer from the long terminal repeat of Moloney murine sarcoma virus gave nearly equal CAT expression in the forward and reverse orientations but was only 30 to 40% active when moved downstream of the CAT gene in either orientation.

Elion and Warner (1986) discovered an example of position and orientation sensitivity in a yeast ribosomal DNA (rDNA) enhancer. The position sensitivity was slight; expression was just under 60% of the default position expression when the upstream enhancer was moved closer to the TSS or downstream of the target gene. The result of inversion of the enhancer was initially found to be a complete loss of enhancer function. However, this effect was further examined in the study by inverting a larger fragment consisting of the

correctly oriented enhancer flanked by about 750 bp of non-enhancer DNA on either side. The resulting vector restored enhancer activity to about 50% in spite of the inverted enhancer orientation in this construct. When this same larger inversion was performed on the test vector with the minimal enhancer already inverted, the resulting vector showed virtually no enhancer activity even though the enhancer ended up in the forward orientation. It was concluded that the observed loss of enhancer activity was due to a silencing effect of the fusion of the 3' end of the inverted enhancer with pBR322 vector sequence normally found at the 5' end of the enhancer.

Taken together, these examples demonstrate that many enhancers show some sensitivity to changes in position and orientation. The degree of sensitivity varies from one enhancer to the next and may also depend on the nature of the change. Some enhancers are sensitive to changes in position but not in orientation, or *vice versa*; some enhancers also appear to show different degrees of activity when different combinations of position and orientation changes are introduced (Kim *et al.*, 2008; Elion and Warner, 1986). In some cases (Small *et al.*, 1993; Jaynes *et al.*, 1988; Elion and Warner, 1986) the activity of an enhancer was found to be influenced by other nearby sequence elements, which may point to a universal explanation of why certain enhancers are only fully functional in the naturally occurring position or orientation.

Aside from a misidentification of the minimal enhancer sequence, there are several possible ways that an enhancer could be dependent on other sequence

elements. Two such possibilities are discussed in the following paragraphs. One simple explanation is that there may be other TF binding sites outside of the defined minimal enhancer, and TFs bound to these external sites interact with enhancer-bound TFs to bring about 100% transcriptional activation. These external sites could be deleted or interrupted in some of the altered constructs, preventing the TFs from binding these sites. Conversely, if there are multiple TF binding DNA elements that must interact with the basal promoter through bound TFs, inverting or repositioning one such element, such as the enhancer, may alter the formation of the necessary loop structure and thereby prevent transcription.

An alternate explanation for sensitivity to genomic context is that sequences flanking an enhancer may play a critical role in TF binding. In higher organisms, DNA exists as part of the material chromatin, a complex of the DNA and various bound proteins referred to as histones. Most of the DNA in chromatin is found in nucleosomes, which are composed of stretches of approximately 147 bp of DNA wrapped around a core of eight histones (Segal and Widom, 2009). Certain DNA sequences are favorable for nucleosome formation, while other sites tend to remain nucleosome-free or have lower nucleosome density because the sequence is less favorable for nucleosome occupancy (reviewed in Segal and Widom, 2009). TFs cannot bind their target sites when nucleosomes occur at these sites, meaning that TF binding sites must either occur in nucleosome-free stretches of DNA or be subject to some mechanism of nucleosome displacement. Some DNA sequences that are unfavorable for nucleosome formation tend to

prevent nucleosomes from developing in flanking sequences as well, which means that enhancers located adjacent to nucleosome-free regions may be more accessible to their TFs (Segal and Widom, 2009). Altering the position or orientation of such an enhancer could therefore reduce TF binding. This effect will often be visible using transient plasmids, in which the DNA appears to form nucleosomes in a normal fashion, although higher order chromatin structures may differ from those seen in the correct genomic context (Hebbar and Archer, 2008).

Changes in enhancer position or orientation may affect processes other than target gene expression. There have been a few studies that have linked changes in enhancer position or orientation to deficiencies in processes not related to gene expression. One such study was performed by Bachl *et al.* (1998). The focus of this study was the intronic enhancer for the heavy (H) chain of the immunoglobulin (Ig) gene, specifically the role of this enhancer in hypermutation in the variable (V) region of the H chain. For this study the Ig basal promoter was replaced with a thymidine kinase (*tk*) promoter fused to the SV40 enhancer; this chimeric promoter had previously been found to work in hypermutation experiments. The rest of the Ig gene was left intact. Deletion of the intronic enhancer caused a roughly 100-fold decrease in the mutation rate, showing that this enhancer is necessary for the natural hypervariability of the V region. Similarly, a construct with the intronic enhancer in the reverse orientation showed a 10-fold drop in the mutation rate, and there was a 7-fold drop in mutation frequency when the enhancer was repositioned 3' of the entire V region. These

decreases in hypermutation were not accompanied by changes in H chain expression, since the levels of mRNA produced from these different constructs appeared to be close to equal on a gel.

Another study by Chandrasekharappa and Subramanian (1987) focused on the connection between the 72-bp-repeat enhancer in SV40 and the process of DNA replication. One significant finding of this study was that DNA replication at the SV40 core replication origin was dependent upon the distance between this core origin and the 72-bp-repeat enhancer. The core origin and enhancer are separated by less than 100 bp in SV40. Compared to a vector with the SV40 core origin but without the enhancer, vectors containing both elements separated by 8 or 9 bp gave an approximately tenfold increase in replication efficiency, while those containing the two elements separated by 99 bp or more gave a replication efficiency equal to or lower than the enhancer-less vector. The increase in DNA replication was not sensitive to changes in the enhancer orientation; however, it was found that activation of DNA replication occurred only when the end of the core origin containing the 17 bp A+T rich element was facing the enhancer. This was true regardless of whether the core origin was found in its default position or moved to the opposite end of the enhancer. This study also noted that the 72-bp-repeat enhancer is not position or orientation sensitive as a transcriptional activator.

Conservation of enhancer position and orientation is not a definitive indicator of sensitivity of that enhancer to position or orientation changes.

Genetic sequence conservation is a useful tool for locating genes as well as functional non-coding DNA elements, since preservation of sequence between distantly related species tends to represent selective pressure against any mutation in that element. Similarly, one may infer that if the position or orientation of a *cis*-regulatory element is evolutionarily conserved, then the observed position or orientation of that element is critical for its function. This appears to be the case for the *C. intestinalis Pax6* upstream enhancers, which show position and orientation conservation with *Pax6* in *C. savignyi* (Irvine *et al.*, 2008) and were found in this study to be sensitive to changes in both position and orientation. While *C. intestinalis* and *C. savignyi* are rather closely related species for such a determination of sequence conservation, *Pax6* homologs such as mouse *Pax6* and the *Drosophila* gene *eyeless* have slightly different sets of enhancers (reviewed in Irvine *et al.*, 2008), making such a determination of position or orientation conservation impossible for these larger evolutionary distances. Other examples of enhancers with conserved position and orientation have shown more flexibility than *Pax6*. The study by Kim *et al.* (2008) noted that the position and orientation of the CSE1 element were evolutionarily conserved relative to the TSS in mammals. The orientation of the entire CSE2 element, and the orientation of all individual YY1 binding sites within this element, were also conserved. The functional analysis found that both elements were orientation sensitive, but the CSE1 element was clearly flexible in its position relative to the TSS. In the *Drosophila eve* gene, the promoter arrangement is conserved between *D.*

melanogaster and the rather distant relative *D. grimshawi* (Small *et al.*, 1993).

This conservation includes the position of the stripe 3 enhancer upstream of the stripe 2 enhancer; yet the positions of these two enhancers can be switched without a visible effect on the stripe pattern that develops.

Examples such as these do support the logical hypothesis that enhancers that are sensitive to changes in position or orientation will show the corresponding evolutionary conservation, but not all examples of conservation point to inflexibility in the genomic environment. The inconsistency of this relationship between position or orientation conservation and position or orientation sensitivity creates more questions than answers. One possibility is that some of these conserved enhancers have other roles outside of transcriptional activation, and that there is some type of sensitivity to genomic context in this alternate role. The conserved position or orientation in these cases would be a reflection of the non-transcriptional function of the enhancer. This explanation is likely to apply to a few specific cases but is not likely to provide a general reconciliation of contextual conservation with position or orientation flexibility. Another possibility is that the flexibility seen using transgenic plasmids does not reflect the true genomic context of these enhancers. This could be due to differences in chromatin structure in the plasmid versus the genome (Hebbar and Archer, 2008), or it could be due to interaction of TFs that bind the enhancer with TFs bound to a different sequence element that was not included in the experimental plasmids.

The results of this study and others create new questions about the nature of enhancers and other *cis*-regulatory DNA modules. One of the defining criteria of an enhancer is that it is capable of increasing expression of its target gene in a position- and orientation-independent manner. This idea continues to be well supported even by the findings of this study, although there may be exceptions to this definition (Kim *et al.*, 2008). However, this definition does not mandate that enhancers in an altered position or orientation function as effectively as in their default setup. This study, along with other studies discussed earlier, indicate that many enhancers lose much of their activity when they are repositioned or inverted. This raises the questions of how widespread this trend is among enhancers and whether there is a pattern to which enhancers show such sensitivity.

Investigation of these new questions could begin with genomic screening for identified enhancer elements that show conservation in their position and orientation across species. Elements that show such conservation relative to their target genes are most likely to be sensitive to inversion or repositioning. Functional studies could then be employed to identify which of these enhancers are position or orientation sensitive. Once these enhancers have been identified, the next step would be to look for patterns that may explain why these enhancers are sensitive to position or orientation changes while others are not. It may be that enhancers for certain types of genes are less tolerant of changes in position or orientation, or sensitivity to position or orientation may reflect a dependence on

genomic context that is common to all of these enhancers. This type of position or orientation sensitivity could also have a substantial biomedical impact in cases where proper gene function is sensitive to the level of expression. In these cases inversion or repositioning of an enhancer could affect expression levels enough to alter the phenotype of an organism or, in the case of *Pax6* or other developmental regulators, produce a developmental disorder. This may also explain the observed phenomenon of evolutionary constraint of position and orientation in certain enhancers. The results of these future studies will build upon our current understanding of how enhancers function.

BIBLIOGRAPHY

- Adachi, Y., Hauck, B., Clements, J., Kawauchi, H., Kurusu, M., Totani, Y., Kang, Y.Y., Eggert, T., Walldorf, U., Furukubo-Tokunaga, K., Callaerts, P., 2003. Conserved *cis*-regulatory modules mediate complex neural expression patterns of the *eyeless* gene in the *Drosophila* brain. *Mechanisms of Development* 120, 1113-1126.
- Atchison, M.L., 1988. Enhancers: mechanisms of action and cell specificity. *Annual Review of Cell Biology* 4, 127-153.
- Bachl, J., Olsson, C., Chitkara, N., Wabl, M., 1998. The Ig mutator is dependent on the presence, position, and orientation of the large intron enhancer. *Proceedings of the National Academy of Sciences of the United States of America* 95, 2396-2399.
- Chandrasekharappa, S.C., Subramanian, K.N., 1987. Effects of position and orientation of the 72-base-pair-repeat transcriptional enhancer on replication from the simian virus 40 core origin. *Journal of Virology* 61, 2973-2980.
- Corbo, J.C., Levine, M., Zeller, R.W., 1997. Characterization of a notochord-specific enhancer from the *Brachyury* promoter region of the ascidian, *Ciona intestinalis*. *Development* 124, 589-602.
- Dehal, P., *et al.*, 2002. The draft genome of *Ciona intestinalis*: Insights into chordate and vertebrate origins. *Science* 298, 2157-2167.
- Elion, E.A., Warner, J.R., 1986. An RNA polymerase I enhancer in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology* 6, 2089-2097.
- Goode, D.K., Snell, P., Smith, S.F., Cooke, J.E., Elgar, G., 2005. Highly conserved regulatory elements around the *SHH* gene may contribute to the maintenance of conserved synteny across human chromosome 7q36.3. *Genomics* 86, 172-181.
- Griffin, C., Kleinjan, D.A., Doe, B., van Heyningen, V., 2002. New 3' elements control *Pax6* expression in the developing pretectum, neural retina and olfactory region. *Mechanisms of Development* 112, 89-100.
- Hauck, B., Gehring, W.J., Walldorf, U., 1999. Functional analysis of an eye specific enhancer of the *eyeless* gene in *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America* 96, 564-569.

- Hebbar, P.B., Archer, T.K., 2008. Altered histone H1 stoichiometry and an absence of nucleosome positioning on transfected DNA. *The Journal of Biological Chemistry* 283, 4595-4601.
- Irvine, S.Q., Fonseca, V.C., Antony, R., Zompa, M.A., 2008. *Cis*-regulation of the *Pax6* gene in the ascidian *Ciona intestinalis*. *Developmental Biology* 317, 649–659.
- Jaynes, J.B., Johnson, J.E., Buskin, J.N., Gartside, C.L., Hauschka, S.D., 1988. The muscle creatine kinase gene is regulated by multiple upstream elements, including a muscle-specific enhancer. *Molecular and Cellular Biology* 8, 62-70.
- Johnson, D.S., Davidson, B., Brown, C.D., Smith, W.C., Sidow, A., 2004. Noncoding regulatory sequences of *Ciona* exhibit strong correspondence between evolutionary constraint and functional importance. *Genome Research* 14, 2448–2456.
- Kammandel, B., Chowdhury, K., Stoykova, A., Aparicio, S., Brenner, S., Gruss, P., 1999. Distinct *cis*-essential modules direct the time–space pattern of the *Pax6* gene activity. *Developmental Biology* 205, 79–97.
- Kim, J.D., Yu, S., Choo, J.H., Kim, J., 2008. Two evolutionarily conserved sequence elements for *Peg3/Usp29* transcription. *BMC Molecular Biology* 9, 108-118.
- Kleinjan, D., Seawright, A., Childs, A., van Heyningen, V., 2004. Conserved elements in *Pax6* intron 7 involved in (auto)regulation and alternative transcription. *Developmental Biology* 265, 462–477.
- Kleinjan, D., Seawright, A., Mella, S., Carr, C., Tyas, D., Simpson, T., Mason, J., Price, D., van Heyningen, V., 2006. Long-range downstream enhancers are essential for *Pax6* expression. *Developmental Biology* 299, 563–581.
- Laimins, L.A., Gruss, P., Pozzatti, R., Khoury, G., 1984. Characterization of enhancer elements in the long terminal repeat of Moloney murine sarcoma virus. *Molecular and Cellular Biology* 4, 183-189.
- Lee, H., Khan, R., O’Keefe, M., 2008. Aniridia: current pathology and management. *Acta Ophthalmologica* 86, 708-715.
- Plaza, S., Saule, S., Dozier, C., 1999. High conservation of *cis*-regulatory elements between quail and human for the *Pax-6* gene. *Development Genes and Evolution* 209, 165–173.

- Ptashne, M., 1986. Gene regulation by proteins acting nearby and at a distance. *Nature* 322, 697-701.
- Ptashne, M., 1988. How eukaryotic transcriptional activators work. *Nature* 335, 683-689.
- Satoh, N., Satou, Y., Davidson, B., Levine, M., 2003. *Ciona intestinalis*: an emerging model for whole-genome analyses. *TRENDS in Genetics* 19, 376-381.
- Segal, E., Widom, J., 2009. What controls nucleosome positions? *TRENDS in Genetics* 25, 335-343.
- Shaul, Y., Rutter, W.J., Laub, O., 1985. A human hepatitis B viral enhancer element. *The EMBO Journal* 4, 427-430.
- Simpson, T.I., Price, D.J., 2002. *Pax6*; a pleiotropic player in development. *BioEssays* 24, 1041-1051.
- Small, S., Arnosti, D.N., Levine, M., 1993. Spacing ensures autonomous expression of different stripe enhancers in the even-skipped promoter. *Development* 119, 767-772.
- Thompson, J.A., Ziman, M., 2011. *Pax* genes during neural development and their potential role in neuroregeneration. *Progress in Neurobiology* 95, 334-351.
- Vierra, D.A., Irvine, S.Q., 2012. Optimized conditions for transgenesis of the ascidian *Ciona* using square wave electroporation. *Development Genes and Evolution* 222, 55-61.
- Xu, P.X., Zhang, X., Heaney, S., Yoon, A., Michelson, A.M., Maas, R.L., 1999. Regulation of *Pax6* expression is conserved between mice and flies. *Development* 126, 383-39.