5-3-2021

Research Towards Understanding Hox Regulation by Cdx

Joshua M. Fisher
University of Richmond

Research Towards Understanding *Hox* Regulation by Cdx

by

Joshua M. Fisher

Honors Thesis

Submitted to:

Biology Department
University of Richmond
Richmond, VA

May 3, 2021

Advisor: Dr. Isaac Skromne

UNIVERSITY OF RICHMOND

This thesis has been accepted as part of the
honors requirements in the Department of Biology.

Research Towards Understanding *Hox* Regulation by Cdx

Joshua M. Fisher

Approved:

Isaac Skromne — 05/03/2021

(Advisor Signature) — *(Date)*

05/02/2021

(Reader Signature) — *(Date)*

Fisher, Joshua M.

Research Towards Understanding *Hox* Regulation by Cdx

Abstract of an honors thesis
at the University of Richmond

Thesis supervised by Dr. Isaac Skromne
No. of pages in text (24)

Understanding the regulatory mechanisms that control *hox* gene transcription has drawn intense scrutiny from biologists due to the genes' unique clustered organization in the genome, their pivotal function in specifying cellular identities along the main body axis of all animals, their association with animal body form and evolution, and because their misregulation in humans results in severe congenital malformations and some forms of cancer. While much is known about the key regulators of *hox* gene transcription, how these regulators control when and where each *hox* gene is transcribed remains poorly understood. Unfortunately, understanding *hox* transcriptional control has been challenging to dissect due to the large number of genes (48 in humans, more in other species), all tightly grouped into clusters containing numerous global, regional, and local control elements.

Traditional loss-of-function approaches have greatly aided in mapping *hox* control elements within clusters. By deleting portions of DNA between the genes or by inactivating the transcription factors that bind these regions, these studies have revealed that the transcription factor Cdx, among others, is a master controller of *hox* gene transcription. The targeted loss of Cdx causes numerous defects in *hox* transcription: early in development, loss of Cdx causes delays in *hox* transcriptional initiation; later on, Cdx loss results in mismatches in *hox* expression

domains and cellular identities; finally, loss of Cdx causes a failure in the maintenance of *hox* gene transcription.

Numerous Cdx binding sites have been identified as embedded within the *hox* clusters by sequence analysis and protein binding methodologies. However, these binding sites have not been functionally tested. Therefore, it is unclear whether these Cdx binding sites function to control the local, regional, or global regulation of *hox* genes and whether they regulate the time, place, or levels of *hox* transcription. The previous lack of tools to precisely block individual regulatory elements without affecting surrounding control regions left a significant gap in our understanding of the key molecular mechanisms underlying the specification of axial cell identities. Now, this gap in knowledge can be filled using novel CRISPR gene-editing tools which allows for the very selective blocking of Cdx binding sites; thereby, allowing for the quantification of gene expression due to individual regulatory sequences.

In this project, we aim to functionally test the contribution that individual Cdx binding sites have in *hox* gene regulation. We will achieve this by individually blocking Cdx binding sites using CRISPR/dCas9 in the zebrafish and then analyzing changes in the time, distribution, and levels of *hox* gene transcription.

We will focus our analysis in only one of the seven zebrafish *hox* clusters, the *hox*ca cluster. We are focusing on the *hox*ca gene cluster because it has lost the least number of genes relative to other clusters and is involved in specifying the axial identity of cells in the central nervous system. We hypothesize that deletion of these Cdx binding sites will cause local, regional, and global changes in *hox* gene regulation that would translate to changes in time, place, and levels of *hox* transcription. This information can then be utilized as a roadmap to

understand the regulation of other *hox* clusters, both in zebrafish and in other animal species. A more complete understanding of the *hox* gene regulatory elements will deepen our understanding of the specification of cellular identities along the main body axis, which could ultimately contribute to curing *hox*-associated cancers and malformations.

Acknowledgements

I would sincerely like to thank my mentor Dr. Skromne: you inspired me in lecture, supported me in lab, scholarship opportunities, Departmental Honors endeavors, and helped me whenever I needed it. Additionally, I would like to thank the post-baccalaureates who have helped me so much with my project. Furthermore, I would like to thank all the amazing professors I have had in the Biology department at the University of Richmond. Your dedication and devotion have not gone unnoticed. I would like to thank my secondary reader, Dr. Carpenter, for taking the time to advise my thesis. Lastly, I would like to thank my family and my friends at the University for always being there for me. Thank you.

TABLE OF CONTENTS

LIST OF FIGURES

Section 1: Introduction

Diversity in Vertebrae Morphology

In mammals, the vertebrae can be generalized to five groups: cervical, thoracic, lumbar, sacral, and coccygeal. Cervical vertebrae form and support the neck, the thoracic vertebrae contain ribs providing support to the main trunk, lumbar vertebrae are contained in the lower back, the sacrum connects the spine to the hip bones, and further posterior are the coccygeal vertebrae. We, as humans, have seven cervical vertebrae, twelve thoracic vertebrae, five lumbar vertebrae, five sacral vertebrae, and four coccygeal vertebrae (*Anatomy of the Spine*, 2018). The organization of our vertebral column is quite literally the backbone of our existence, supplying support and protection for our central nervous system, internal organs, and limbs. However, there is enormous diversity among the vertebral column among all animals. For example, pythons have hundreds of thoracic vertebrae while pigeons only have five; furthermore, pigeons have fourteen cervical vertebrae while giraffes have only seven supporting their long necks (Badlangana et al., 2009; De Luliis, 2011; Sood, 1946). The cause of this enormous diversity in animal morphology is due to the expression of the *hox* family of transcriptional regulators.

Hox Overview

Decades of research involving mutants and model organism have increased our understanding of *hox* genes since their discovery. The *hox* genes are a family of homeotic selector genes that act to determine the identities of different bodily segments (Deschamps & van Nes, 2005). These genes contain a conserved homeobox transcription factor domain with a

sixty amino acid DNA-binding domain called the homeodomain (Deschamps et al., 1999). The *hox* family of genes is highly conserved throughout evolution and homologues can be found in almost all animals. Duplication events in certain animal phyla have resulted in copies of these genes, known as paralogous genes, that may exhibit partial redundancy with one another (Young et al., 2009). *Hox* genes reside in a clustered organization within the chromosome, and this clustered organization allows for the collinear temporal and spatial expression of the *hox* genes during development: *hox* genes at the 3' end of the clustered are expressed earlier and more anteriorly in development while *hox* genes at the 5' end of the cluster are expressed later and more posteriorly during development (Deschamps & van Nes, 2005). The primary role of the *hox* genes is conferring identity to the anteroposterior axis during development, however, the genes also provide identities to secondary axes – such as limbs (Young et al., 2009). Gain-of-function mutations result in more anterior segments producing patterns typical of more posterior segments and loss-of-function mutations lead to more anterior patterns being expressed.

Understanding *Hox* Through Mutants and Model Organism

hox genes play a fundamental role in the development and morphology of the vertebral column. *hox* genes were initially discovered through the use of mutants in the fruit fly *Drosophila melanogaster.* The discovery came in 1978 when Ed Lewis discovered *Drosophila* with two thoraxes (Fig. 1B). Genetic analysis of these mutants led Lewis to identify the genes underlying this remarkable transformation, the *hox* genes, and to conclude that the arrangement of the *hox* genes in the chromosome and arrangement of the drosophila body

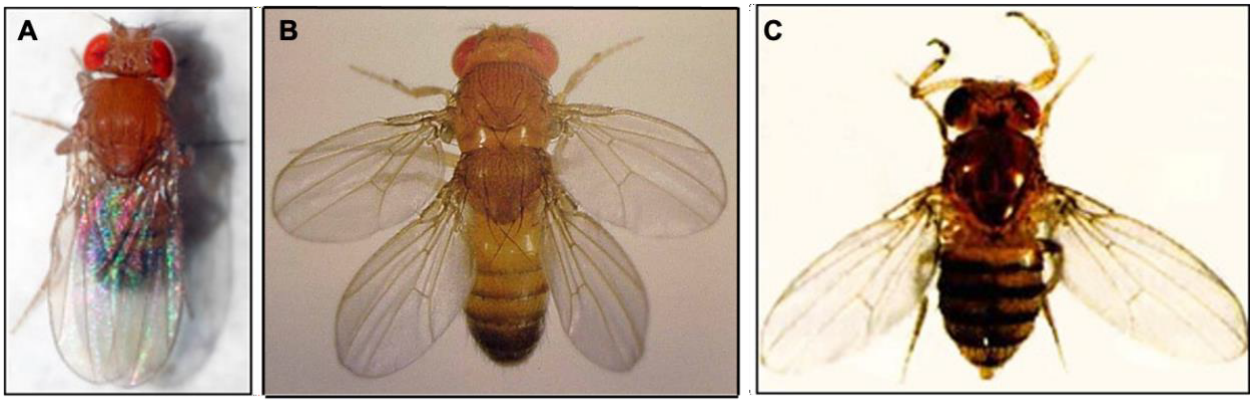

Figure 1. The *Bithorax* and *Antennapedia* Mutants: (A) Wild type *Drosophila,* **(B)** *Bithorax Mutant, and* **(C)** *Antennapedia* gain-of-function mutant.

segments along the anteroposterior axis were the same (Fig. 2) (Lewis, 1978). This property of *hox* genes is termed spatial collinearity; more specifically, *hox* genes located at the 3' end of the cluster will be expressed more anteriorly in development and genes located further 5' will be expressed more posteriorly. The next pivotal discovery in *hox* was in the *Drosophila* mutant *Antennapedia (Antp),* these mutants develop legs in place of their antennae (Fig. 1C). *Drosophila* antennae and legs are homologous structures that develop differently due to the expression of the *hox* Gene *Antennapedia (Antp)* promoting leg identities (Struhl, 1981). Therefore, *Antp* is a gain-of-function mutation, resulting in the expression of a posterior gene in an anterior region of the body that results in

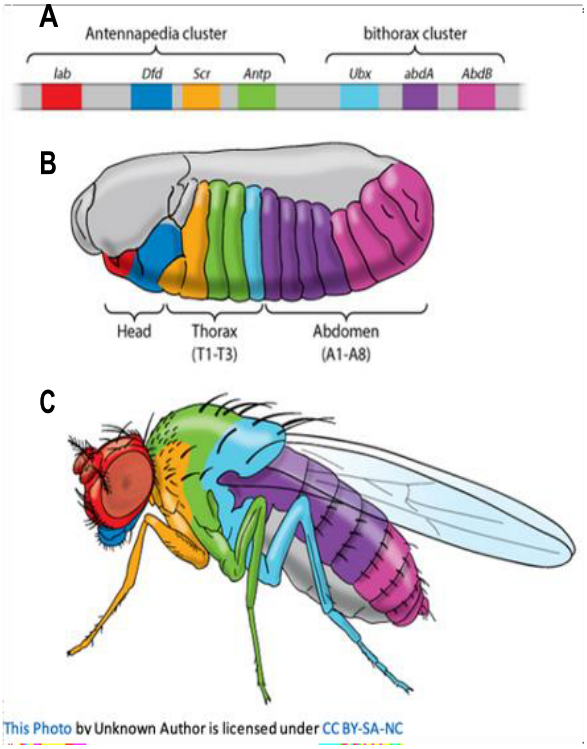

Figure 2. *Hox* in Drosophila: (A) *Drosophila hox* cluster. Genes are indicated with boxes. **(B)** Characteristic *hox* gene expression in embryo and **(C)** in adult.

legs developing instead of antennae (Casares & Mann, 1998; Struhl, 1981) (Fig. 1B). Therefore, *hox* gene mutations are sufficient to induce homeotic transformation in flies.

Since their discovery in *Drosophila*, *hox* genes have been discovered in almost all metazoan phyla (Maeda & Karch, 2009). Further research in *Mus musculus* (mouse) revealed that the disruption of the ordering of the *hox* genes in their cluster affected the timing of initiation of transcription of the *hox* genes (Deschamps & van Nes, 2005). This signifies that not only do the *hox* genes exhibit spatial collinearity of expression but they also exhibit temporal collinearity: *hox* genes at the 3' end of the cluster are expressed earlier in development than genes located at the 5' end of cluster. Therefore, *hox* genes exhibit spatial and temporal collinearity of expression during development (Fig. 3). For example, mouse has four *hox* clusters (Fig. 3A). If one were to follow when and where *hox* genes belonging to group 2 (pink), 4 (green), and 9 (blue) genes are transcribed, one would see that group 2 are transcribed before group 4 and 9 genes (Fig. 3B), and in more anterior regions (Fig. 3C) (Deschamps & van Nes, 2005; Izpisua-Belmonte et al., 1991).

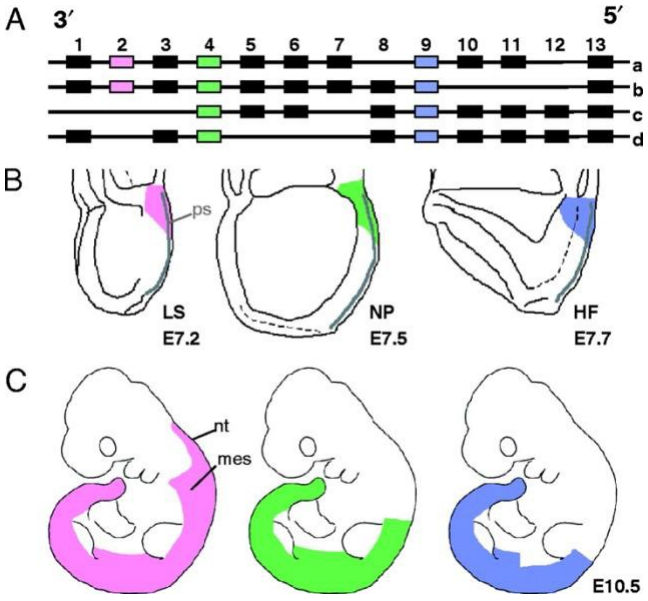

Figure 3. Temporal and Spatial Collinearity of *hox* Expression (A) The four *hox* clusters in mouse, three *hox* genes are shown (2, 4, and 9). Color is used to illustrate collinearity of gene and expression. **(B)** Mouse embryos at different embryonic stages (E) late stage (LS), neural plate (NP), and head fold (HF) stage. **(C)** The anterior limits of *hox* 2 (pink), 4 (green), and 9 (blue). Adapted from Deschamps et al., 2005.

Comparative anatomy of mouse and *Gallus gallus* (chick), Fig. 4A, demonstrate how differences in vertebrae between species can be attributed to differences in expression of *hox* genes. Fig. 4B shows how the boundary of *hox*5 and *hox*6 corresponds with the boundary of cervical and thoracic vertebrae in each species. From this figure, it is evident how the differences in expression of *hox* correlates with the increased number of cervical vertebrae and decreased number of thoracic vertebrae observed in chick compared to mouse.

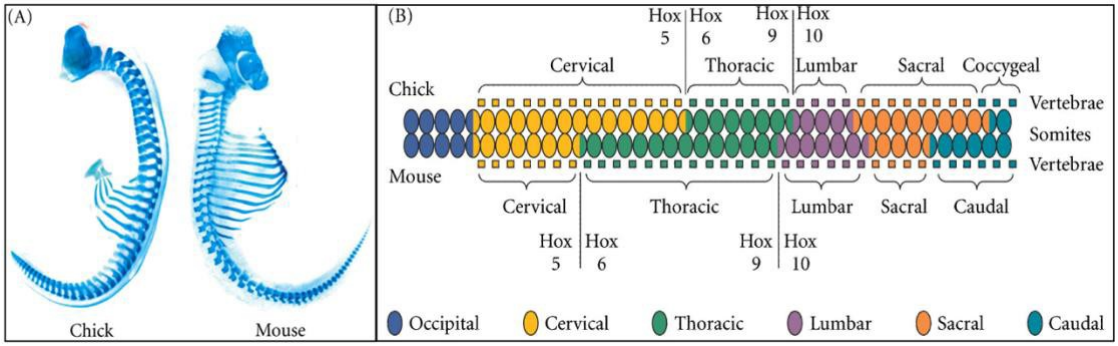

Figure 4. Comparative Anatomy and *hox*: (A) Occipital and vertebral column structure of Chick and Mouse. **(B)** Comparison of somite and vertebrae structural fates in chick and mouse. Color represents structural fate, ovals represent somites, squares represent somites, and vertical lines represent boundary between *hox* genes. Adapted from *Developmental Biology,* Seventh Edition, Figure 11.46

Regulation of *Hox*

Despite many advances in understanding the function of *hox* genes, the mechanisms that regulate *hox* expression are complex and our understanding is incomplete. So far, it known that epigenetic and genetic controls primarily regulate *hox* expression patterns. Epigenetically, long non-coding RNAs, chromatin remodeling factors, and histone writers, erasers, and readers all function in regulating *hox*. In pre-gastrulation mouse embryos, *hox* genes are transcriptionally inactive due to repressive histone modifications, and throughout development, histone modifications that results in active transcription occur from the 3' end of the cluster to the 5' end (Soshnikova & Duboule, 2009; Young & Deschamps, 2009). These

activating histone modification are necessary but not sufficient for proper *hox* gene transcription, indicating that enhancer sequences must also contribute to proper *hox* expression (Soshnikova & Duboule, 2009).

Highly conserved expression profiles and the conservation of clustering indicate that global control regions (GCRs) regulate the transcription of *hox* (Dollé et al., 1989). These GCRs are distantly located from the *hox* cluster and have the potential to regulate the entire cluster (Fig. 5) (Spitz et al., 2003). Tschopps and colleagues demonstrated that targeted mutations disrupting the organization of the cluster affects the temporal collinearity of expression but not the spatial collinearity (Tschopp et al., 2009). Furthermore, the larvacean *Oikopleura* have completely lost clustering of the *hox* genes, and expression of *hox* still occurs in nested anterior-posterior territories similar to the patterns observed in animals with *hox* clustering intact (Seo et al., 2004). Therefore, clustering is not strictly required in order to elicit the spatial collinearity of *hox* genes and the anteroposterior axis. This indicates that there are *cis*-regulatory elements in close proximity either regulating singular *hox* genes – local regulatory elements – or regulating a few *hox* genes – regional regulatory elements – responsible for the spatial collinearity of the *hox* genes (Fig. 5).

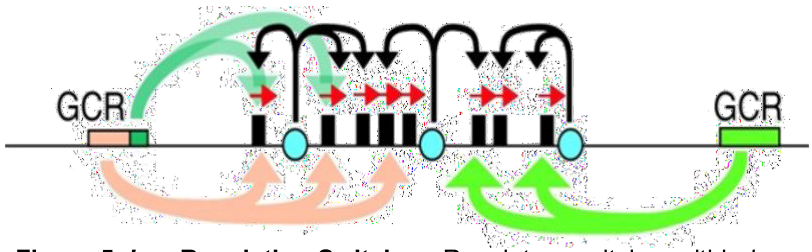

Figure 5. *hox* Regulation Switches: Regulatory switches within *hox* clusters can act as local (red arrows), regional (blue circles), or global control regions (GCRs; green and peach boxes and arrows. Adapted from Deschamps and Duboule, 2017.

Epigenetic and genetic controls result in the transcription of *hox* occurring in three distinct phases: initiation, establishment, and maintenance (Deschamps et al., 1999; Hayward

et al., 2015). During the initiation phase, GCRs direct the temporal collinear activation of *hox* in the gastrulating embryo (Tschopp et al., 2009). During establishment, *hox* posterior expression domains expand anteriorly as the anteroposterior axis expands and rostrally located cells initiate transcription. The anterior expansion of the *hox* expression domain is associated with open chromatin markers accumulating along the cluster (Soshnikova & Duboule, 2009). Finally, during the maintenance phase the anterior boundary of *hox* expression is established by local and/or regional *cis*-regulatory elements and epigenetic modifications (Deschamps et al., 1999; Deschamps & van Nes, 2005; Hayward et al., 2015; Tschopp et al., 2009).

Previous research has shown that the specific *hox* expression patterns and anterior boundaries observed in developing embryos is dependent on the interaction of numerous signaling pathways and transcription factors. Specifically, Wnt, fibroblast growth factor (Fgf), retinoic acid (RA), and the caudal-related genes (Cdx) function to regulate the expression of *hox*. Research in zebrafish and mouse has shown that Wnt signaling is necessary for posterior body formation and mutants lacking Wnt develop posterior body defects. Furthermore, *hox* expression in these mutants was decreased and anteriorly shifted indicating that *hox* expression occurs downstream Wnt (Ikeya & Takada, 2001; Shimizu et al., 2005). Similarly, research in chick has shown that the initiation of *hox* is dependent on RA and Fgf signaling. Bel-Vialar et al. showed that 5' *hox* genes are ectopically activated following Fgf treatment and 3' *hox* genes are ectopically activated following RA treatment (Bel-Vialar et al., 2002). Significantly, these experiments also demonstrated that Fgf, Wnt, and RA signaling are

integrated by the Cdx family of genes into coherent *hox* gene transcription (Fig. 6) (Deschamps & van Nes, 2005). While FGF, Wnt and RA can regulate hox gene transcription directly, they can also do so indirectly through the regulation of Cdx transcription factors.

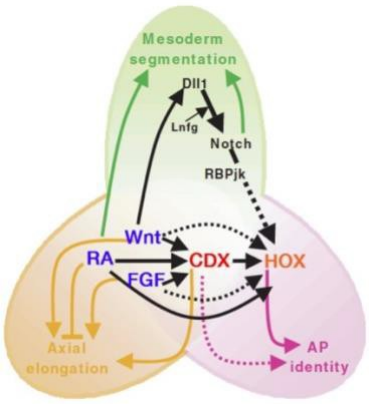

Figure 6. Cdx Integration: Cdx integrates extracellular signal informaton (Wnt, RA, Fgf; in blue) into coherent *hox* expressions. From Deschamps et al., 2005

Cdx Overview

The caudal-related family of genes (Cdx) are highly conserved among animals. They are evolutionarily related to the *hox* family as they are both derived from the *Protohox* gene cluster (Chourrout et al., 2006; Young et al., 2009). As the name suggests, this family of genes confers caudal (posterior) identities to developing embryos. Three paralogous Cdx genes exist in vertebrates with partial redundancy in function. In mouse, loss-of-function mutations in *Cdx2* and *Cdx4* result in embryos with truncated caudal regions, however expression of *Hoxb8* has the ability to rescue the mutant phenotype (Fig. 7) (Young et al., 2009). The ability of 5' *hox* genes to rescue posterior identities indicates that *hox* gene function is downstream of Cdx function. As in mouse, inactivation of *cdx* results in the posterior shift of *hox* gene expression domains in zebrafish (Hayward et al., 2015). Genomic analysis of the *hox* cluster further revealed that many Cdx binding sites are found within the loci (Deschamps & van Nes, 2005; Hayward et al., 2015). However, it is not known how these Cdx binding sites function to regulate *hox* expression locally and/or regionally.

Section 2: The Project and Methodologies

Project Summary

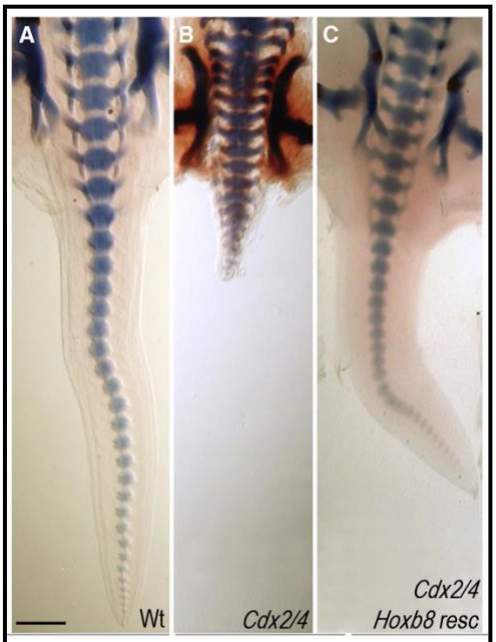

Figure 7. Truncation in Cdx Mutants in Mouse: Skeletal preparations of the caudal regions of (**A**) wildtype (WT), **(B)** Cdx2/4 loss of function mutants, and **(C)** Cdx2/4 mutants carrying the *hox*b8 transgene. Adapted from Young et al., 2009.

Understanding the regulatory mechanisms that control *hox* gene transcription has intrigued biologists due to the genes' clustered organization in the genome (Fig. 8a), their function in specifying cellular identities along the anteroposterior axis of all animals (Fig. 8b), their association with animal morphology and evolution, and because their misregulation in humans is associated with severe congenital malformations and some cancers. (Bhatlekar et al., 2014; Deschamps & van Nes, 2005; Gaunt, 2018). While several key regulators of *hox* gene transcription have been identified genetically (Deschamps & van Nes, 2005), how they regulate the transcription of each *hox* gene in unique spatial domains remains poorly understood molecularly. The Cdx family of transcription activators has emerged as critical *hox* regulators, integrating several signaling inputs into coherent *hox* gene outputs (Fig. 6). However, how Cdx activates *hox* transcription is not understood: It is unknown which of the many Cdx binding sites embedded within the *hox* clusters act as bona fide molecular switches (Fig. 8a), and whether they activate *hox* transcription locally, regionally, or globally, as demonstrated for other regulators (Fig. 5). The lack of tools to exclusively block individual Cdx

sites without altering other control sequences has left a significant gap in our understanding of the mechanisms regulating *hox* transcription.

With the purpose of understanding *hox* regulation and axial cell fate specification, we aim to functionally characterize the activity of evolutionarily conserved Cdx binding sites in a single *hox* cluster of the zebrafish, the *hoxca* cluster (Fig. 8a). We hypothesize that some but not necessarily all Cdx binding sites in the *hox*ca cluster will regulate the transcription of one,

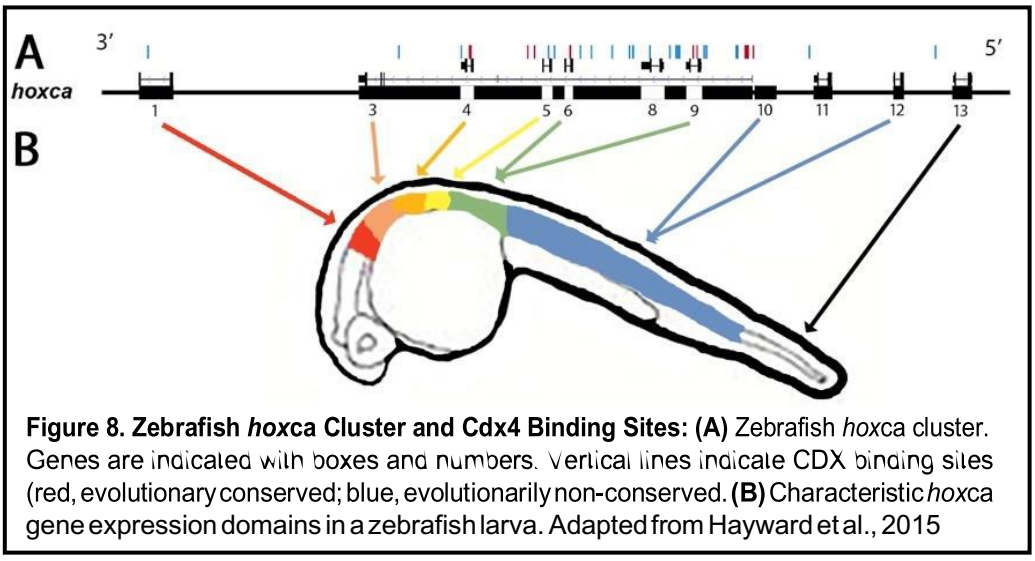

Figure 8. Zebrafish *hoxca* Cluster and Cdx4 Binding Sites: (A) Zebrafish *hox*ca cluster. Genes are indicated with boxes and numbers. Vertical lines indicate CDX binding sites (red, evolutionary conserved; blue, evolutionarily non-conserved. **(B)** Characteristic *hox*ca gene expression domains in a zebrafish larva. Adapted from Hayward et al., 2015

several, or most *hoxca* genes. We will achieve this goal by systematically blocking Cdx binding sites (CRISPR/dCas9 gene editing), and then analyzing the cellular distribution and levels of *hox* gene transcript (in situ hybridization and quantitative PCR, respectively). Results of this analysis will identify which Cdx sites are functionally responsible for regulating the expression of one, several, or many *hox* genes (interpreted as local, regional, and global control regions). This work is significant because it expands our understanding of how molecular switches regulate the expression of complex genetic loci, and how the vertebrate body axis is specified and can be altered during abnormal development and evolution.

Project Justification

Because *hox* genes are evolutionally conserved throughout the animal kingdom, it is rather unsurprising that they are also found in humans. Unfortunately, their misregulation results in severe congenital malformations and some forms of cancer. Both synpolydactyly and Hand-Foot-Genital syndrome are caused by mutations in *hox* genes. *HOX*A is reported to be overexpressed in breast and ovarian cancers, *HOX*B in color cancer, *HOX*C in prostate and lung, and *HOX*D in color and breast cancers (Bhatlekar et al., 2014) . The importance of *hox* in specifying the anteroposterior axis in humans and its association with certain maladies makes it imperative that we understand the regulatory mechanism directing its expression.

Zebrafish as a Model Organism

Zebrafish is the ideal model organism to study the regulation of *hox* gene expression by Cdx. Unlike the development of chicken or mouse embryos, zebrafish development occurs externally allowing for direct visual observation. Furthermore, chicken and mice are relatively much more expensive to maintain (Veldman & Shou, 2008). Zebrafish development occurs quickly with the onset of gastrulation occurring only after five hours, segmentation after ten hours, and hatching starting at two days post fertilization (Kimmel et al., 1995). Zebrafish can be bred year-round and can be housed in high density aquariums. Females are capable of laying hundreds of eggs per breeding event. Furthermore, the zebrafish genome has been fully sequenced and is readily available. Lastly, there are numerous genetic tools available to explore and manipulate the zebrafish genome including *in situ* hybridization, morpholino antisense oligos, and CRISPR (Veldman & Shou, 2008).

CRISPR and dCas9

Clustered regularly interspaced palindromic repeats (CRISPRs) were first discovered in *Escherichia coli* but have since been observed in numerous bacteria and archaea. CRISPR is DNA conferring adaptive immunity to bacteria through the use of CRISPR associated (Cas) proteins. This adaptive immunity features two main stages: 1) bacterium store small segments of DNA from invading viruses as spacer sequences in the CRISPR array, 2) CRISPR RNAs (crRNAs) direct Cas proteins to foreign nucleic acids which cleave them upon complementary base pairing (Fig. 9) (Doudna & Charpentier, 2014; Hsu et al., 2014; Jiang & Doudna, 2017; Loureiro & Da Silva, 2019; Ran et al., 2013).

CRISPR/Cas9 is a relatively new-genetic editing technique that is incredibly powerful. Cas9 is a dual-RNA-guided DNA endonuclease that has the ability to cleave genomic DNA at highly specific sites using guide RNA (gRNA), which is a duplex of trans-activating crRNA (tracrRNA) and crRNA. The tracrRNA functions to hold the gRNA to the Cas9 protein while the crRNA is what guides the Cas9 protein to a specific site in the model organism's genome. The twenty-nucleotide crRNA sequence can be easily designed to target any genomic loci as long as the target sequence contains a promoter adjacent motif (PAM) sequence directly 3' the twenty base pair target sequence (Ran et al., 2013). The PAM sequence is crucial for initial complementary base pairing and the absences of the PAM will result in Cas9 not binding the target DNA (Doudna & Charpentier, 2014; Sternberg et al., 2014). Cas9 will introduce a double stranded break – usually a few base pairs upstream the PAM sequence - in the DNA at the target sequence upon binding (Doudna & Charpentier, 2014). Protective genomic processes will

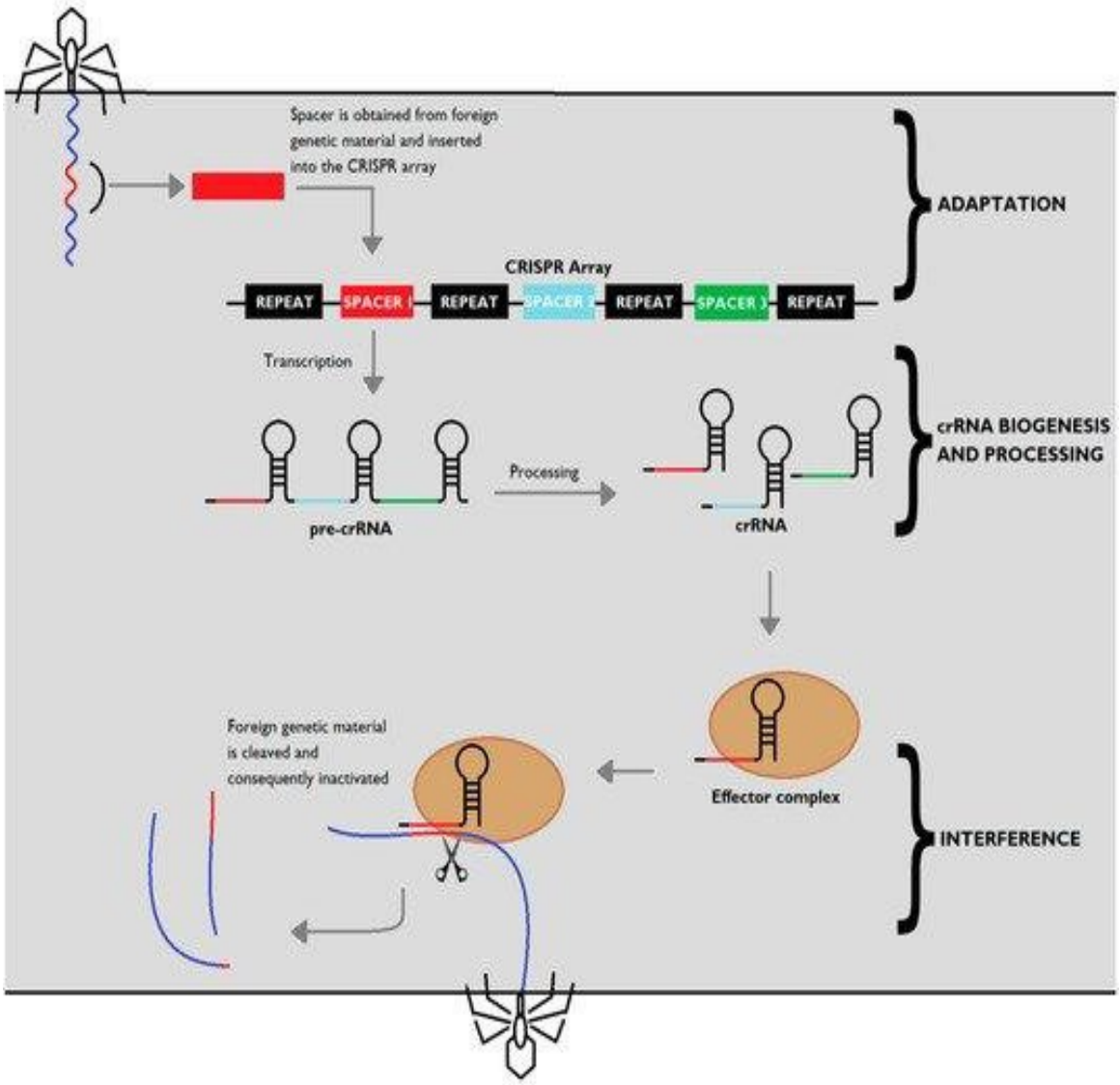

Figure 9. CRISPR-Cas Adaptive Immunity in Bacteria: Adaptation: Foreign genetic material is inserted into CRISPR array. **crRNA Biogenesis and Processing:** Spacer sequences and transcribed and processed into crRNA. **Interference:** crRNA directs Cas proteins to cleave foreign genetic material. From Loureiro et al., 2019.

repair this damage in one of two ways: non-homologous end joining (NHEJ) or homology-directed repair (HDR). NHEJ is quick and error prone and introduces insertion/deletion mutation in the locus as a result of re-ligation of non-homologous ends. HDR occurs much less frequently than NHEJ and requires template DNA (single or double stranded) to repair the locus in a mutation free manner. Scientists can take advantage of both pathways by using the NHEJ

pathway to knockout genes or by providing template DNA in the HDR pathway to introduce novel genetic material/genes into a locus (Fig. 10) (Doudna & Charpentier, 2014; Hsu et al., 2014; Jiang & Doudna, 2017; Loureiro & Da Silva, 2019; Ran et al., 2013).

CRISPR/Cas9 editing is cost-effective and easy to use method to introduce specific genomic edits. Unfortunately, this system features some limitations: specifically, off target effects and the PAM requirement. However, these limitations are not severe as off target effects can be easily mitigated/eliminated by ensuring that no unintentional pairwise sequence alignments occur between the crRNA sequence and the model organism's genome. Furthermore, PAM sequences can typically be found every eight to twelve base pairs (in humans). While these generic limitations of the CRISPR/Cas9 system are easily mitigated there is one additional limitation that specifically affects this research.

Unfortunately, the primary repair pathway of double stranded base repair is the NHEJ pathway which typically results in a locus that differs slightly in size from the original locus due to insertion/deletion mutations. Because the temporal collinearity of expression of *hox* is dependent on the distance of the *hox* genes from the GCRs, altering the length of the *hox* locus will affect *hox* expression pattern making interpretation of the results difficult. Fortunately, there are alternatives to the Cas9 protein that will allow for the same precise, easy, and affordable genome editing that CRISPR/Cas9 affords but will also not alter the length of the *hox* locus. The first alternative to Cas9 is dead-Cas9 (dCas9), dCas9 is the nuclease-deactivated variant of Cas9. This protein can then be used to interfere with transcription via steric blockages of RNA polymerases binding or transcription elongation (Brocken et al., 2018; Qi et al., 2013). Furthermore, dCas9 can be genetically fused to repressor domains (KRAB) to decrease

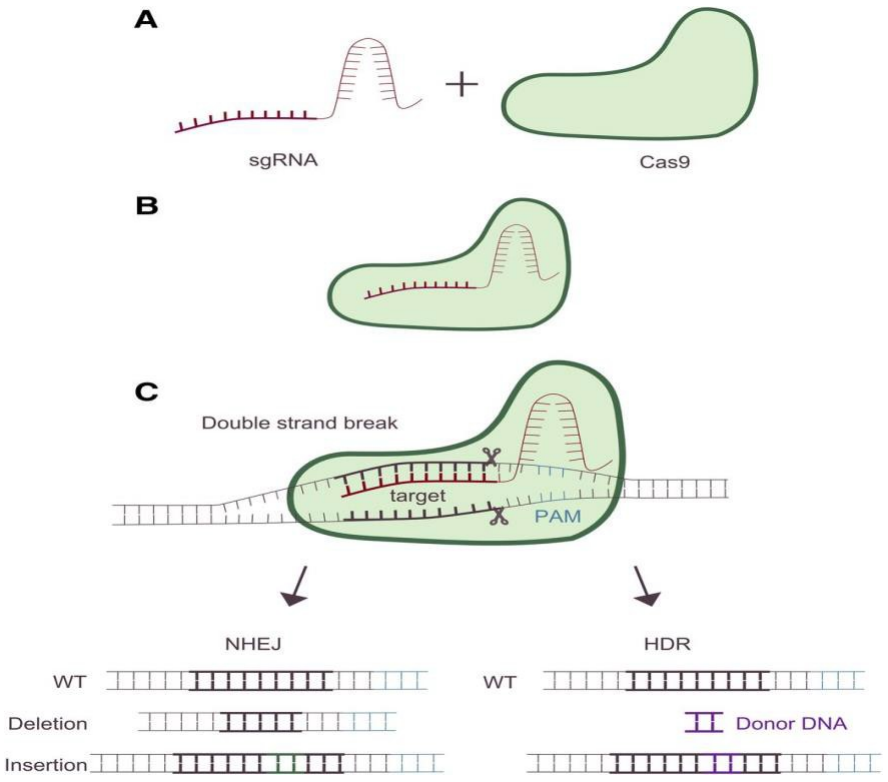

Figure 10. The CRISPR/Cas9 System: (A) A single guide RNA duplex composed of tracrRNA and crRNA and the Cas9 protein. **(B)** sgRNA conjugates to the Cas9 protein via the tracrRNA portion of the sgRNA. **(C)** Cas9 protein matches crRNA portion of sgRNA to genomic DNA and introduces double strand break which can be repaired through the error prone non-homologous end joining pathway (NHEJ) or the accurate but less frequent homology-directed repair (HDR) pathway.

targetgene expression (Brocken et al., 2018). Therefore, this research utilized a dCas9 which targeted evolutionarily conserved Cdx4 binding sites within the *hox*ca cluster in zebrafish.

In-Situ Hybridizations

In-situ hybridization (ISH) is a technique to detect the localization of nucleic acids within tissues. In ISH, nucleic acid sequences that complementary bind to the nucleic acid sequence of interest is attached to a reporter molecule with radio-, fluorescent-, or antigen-labeled bases (Jensen, 2014). Therefore, localization of the nucleic acid sequence of interest can be observed.

This research plans to use ISH to analyze the expression patterns and cellular distribution of *hox* gene transcripts after blocking Cdx4 binding sites within the *hoxca* cluster with dCas9.

RT-qPCR

Quantitative Reverse Transcriptase Polymerase Chain Reaction (RT-qPCR) is a form of PCR that uses RNA as the starting material and provides quantitative information about the starting RNA. In RT-qPCR, RNA is first converted into complementary DNA by reverse transcriptase. Then during each cycle of PCR, the quantity of DNA in the sample can then be measured by the amount of fluorescent signals given off by double-stranded DNA binding dyes. RT-qPCR will be used in this research to quantify *hox* gene expression levels before and after blocking Cdx4 binding sites within the *hox*ca cluster (*RT-qPCR - Quantitative Reverse Transcriptase PCR*, n.d.).

Project Strategy

Zebrafish contain seven distinct *hox* gene clusters as a result of duplication events that occurred within the Cypriniformes order of teleost fish (Stellwag, 1999). Of the seven *hox* clusters, *hoxca* is the most complete making it the best candidate for experimental manipulation and investigation. All seven clusters contain numerous binding sites for Cdx transcription factors (Paik et al., 2013). There are two paralogous Cdx genes in zebrafish: *cdx1a* and *cdx4; cdx4* is redundant with *cdx1a* in expansion of posterior identities such that knock out of *cdx1a* alone does not result in posterior truncation, knock out of *cdx1a* and *cdx4* does result in posterior truncation, and knock out of *cdx4* alone results in posterior truncation (Fig. 11)

(Skromne et al., 2007). Therefore, it is because of this redundancy that this project focuses on *cdx4.*

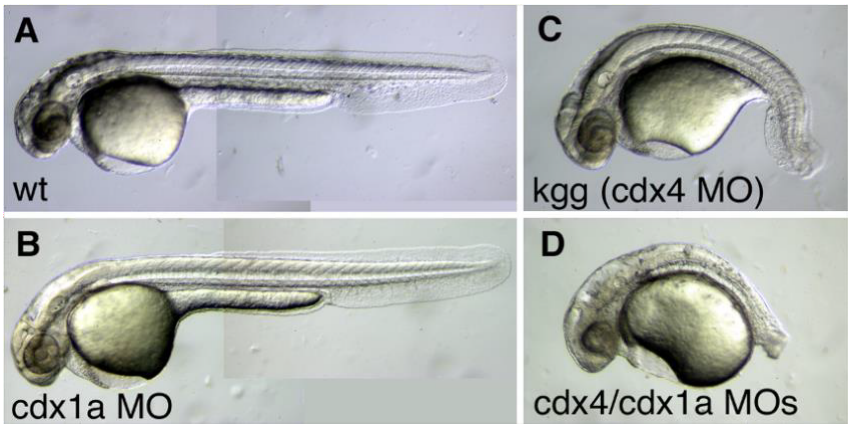

Figure 11. Cdx Mutants in Zebrafish: (A) Wildtype Zebrafish, **(B)** Zebrafish injected with *cdx1a* morpholino, **(C)** Zebrafish injected with Cdx4 morpholino, and **(D)** Zebrafish injected with cdx4 and cdx1a morpholinos. From Skromne et al., 2007.

The strategy to functionally characterize Cdx4 binding sites relies heavily on a set of stringent rules. These rules include (1) restricting the analysis to the most complete cluster, *hoxca* (Fig. 8a); (2) selecting sites that have been shown biochemically to be bound by Cdx4 (it is unknown if this binding leads to gene activation); (3) of these, selecting sites that are evolutionarily conserved across vertebrates; and (4) restricting analysis to a single developmental time point when *hoxca* gene transcription has stabilized but is still sensitive to loss of Cdx4 activity (20 hours post fertilization). Using these rules, we have identified the specific Cdx binding sites to block.

We have designed an experimental pipeline to create, identify, and analyze blocked Cdx binding sites (Fig. 12). To block Cdx, embryos will be injected with specific guide RNAs and commercial dCas9 enzyme. In parallel, a control guide will be injected to target a gene whose inactivation causes known phenotypic defects. Only those experiments in which over 70% of control injected embryos display a mutant phenotype will be further processed. This control

experiment is also significant in determining the optimum concentrations of components in the CRISPR/dCas9 injection mix. DNA and RNA from individual control and experimental embryos at 20 hours post fertilization will be extracted using commercially available kits. RNA from the samples will then be processed for *hox* transcript levels and quantified by RT-qPCR analysis. To determine changes in *hox* expression patterns, whole embryos will be processed for in situ hybridization, photographed, and then analyzed for genetic changes by PCR. All these methodologies are standard in our laboratory.

Current Status

So far, we have identified Cdx4 binding sites within the *hoxca* cluster that are evolutionarily conserved among many species (Fig. 8, Fig. 13). We have developed gRNAs to use with CRISPR/dCas9 in order to block Cdx4 transcription factors from binding (Fig. 12). We have designed primers for qPCR that flank the Cdx4 binding sites and have designed primers to quantify *hox* transcript using RT-qPCR (Fig. 13). We are in the process of using the primers against the Cdx4 binding sites to test the function of the guides. Additionally, we are working on systematically blocking Cdx4 sites using CRISPR. After those steps are complete, DNA and mRNA extraction of embryos will need to be performed in order to quantify *hox* expression levels using RT-qPCR. Furthermore, analysis of *hox* expression patterns in-situ should also be performed (Fig. 12).

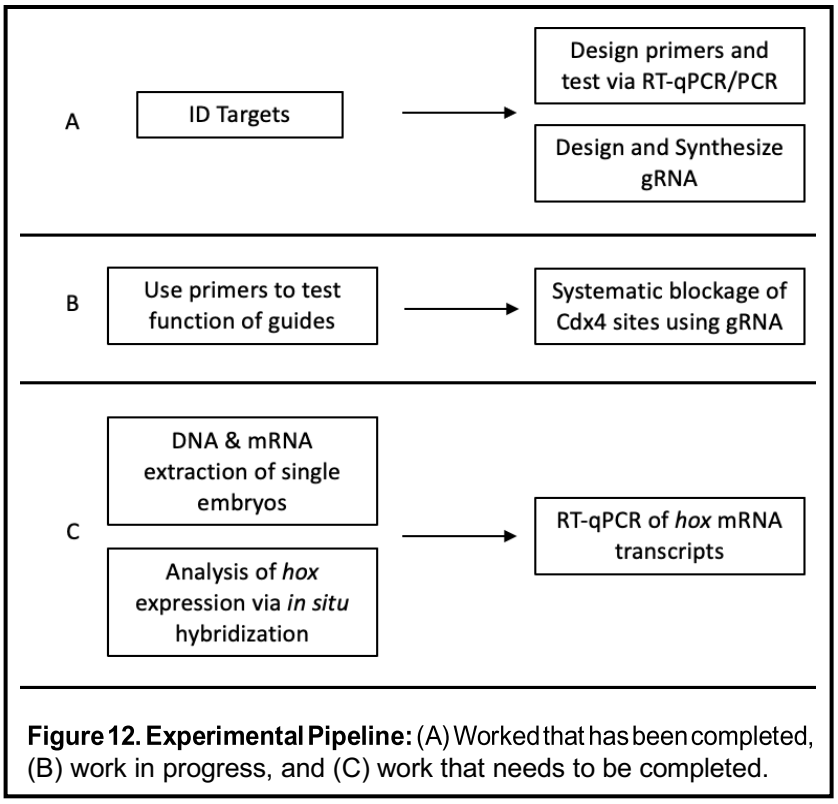

Figure 12. Experimental Pipeline: (A) Worked that has been completed, (B) work in progress, and (C) work that needs to be completed.

Cdx Binding Site	Cdx Binding: Squence	Cdx Binding: position	gRNA: Sequence	gRNA: position	PCR Left Primer	PCR Right Primer
A	AATGGCTTTGCAGAACACTTTAA GGTCCACCAACAACGTCCATTAA ACCATAAAGTAATTTAGCCCAGA CGTCTAGCCAGTTAGCCAGGTTG GCTTTGT	chr23:36098867-36098965	GGTTTAATGGACGTTGTTGG	chr23:36098917-36098936	TTACAGCCCTGTTGGAGCTT	TGTTGGAGGATTGTTGTTGC
B	ACAAACAACACCAAGCTCCAACA GGGCTGTAAAACTGGCATTTTAG TTCTAGGGCGGAAACGTGGTCG ATCAGGCTTTGAAATCACCGCAT CCTGTTCG	chr23:36098753-36098851	GGGCGGAAACGTGGTCGATC	chr23:36098781-36098800	TCCACCAACAACGTCCATTA	AGGATGCGGTGATTTCAAAG
C	TACAACCCAATCCTATGGCGTTT GTCATCAATGGATAGCCCTAAAG CAGACTTGAGTCGGGATGAAAG AGTATAGAACTTGGTTTTACGAC GGACAACA	chr23:36095982-36096080	GATGACAAACGCCATAGGAT	chr23:36096052-36096071	GTTTATTGCGTTCGCTTGGT	ACGTTCTTTGCCAGACTTGG
D	AGGATATTGCTGCTATATCCGCA TGCTAGGCGGCAGTGTCCTGTG GGGCTATAACGCGTGCTGTAAAC TCTTTGTTGGTGACGCATTCCAG AGGAGAAC	chr23:36089763-36089861	GGACACTGCCGCCTAGCATG	chr23:36089821-36089840	GAATGCGTCACCAACAAAGA	ATAACACCCCATGCACCACT

Figure 13. Subset of Cdx4 Binding Site Data: Example of the Cdx4 binding sites, gRNA sequences that target binding sites, and PCR primers flanking Cdx4 binding sites generated during research.

Potential Outcomes, Predictions, and Future Research

The blockage of individual Cdx binding sites will impact *hox* gene transcription in one out of four different ways. An elimination could affect transcription of (1) the closest *hox* gene

only, (2) several nearby *hox* genes, (3) most *hox* genes within the locus, or (4) have no effect on transcription. For the first three cases, we would interpret the results to suggest that an individual Cdx binding site has local, regional, or global regulatory functions, respectively. For the last case, the result would suggest that the Cdx site either does not control *hox* transcription or it functions redundantly with other Cdx sites. The information obtained from this project can then be utilized as a roadmap to understand the coordinated regulation of genes that are grouped in clusters (*hox* and others), both in zebrafish and in other animal species. A more complete understanding of the *hox* gene regulatory elements will impact our understanding of the specification of cellular identities along the main body axis, their impact on vertebrate body evolution, and their function underlying congenital malformations and cancers.

REFERENCES

Anatomy of the Spine. (2018). Mayfield Clinic: Brain and Spine. https://doi.org/10.2106/00004623-196547010-00018

Badlangana, N. L., Adams, J. W., & Manger, P. R. (2009). The giraffe (Giraffa camelopardalis) cervical vertebral column: A heuristic example in understanding evolutionary processes? *Zoological Journal of the Linnean Society*, *155*(3), 736–757. https://doi.org/10.1111/j.1096-3642.2008.00458.x

Bel-Vialar, S., Itasaki, N., & Krumlauf, R. (2002). Initiating Hox gene expression: in the early chick neural tube differential sensitivity to FGF and RA signaling subdivides the HoxBgenes in two distinct groups. *Development*, *129*(22), 5103–5115.

Bhatlekar, S., Fields, J. Z., & Boman, B. M. (2014). HOX genes and their role in the development of human cancers. *Journal of Molecular Medicine*, *92*(8), 811–823. https://doi.org/10.1007/s00109-014-1181-y

Brocken, D. J. W., Tark-Dame, M., & Dame, R. T. (2018). dCas9: A versatile tool for epigenome editing. *Current Issues in Molecular Biology*, *26*, 15–32. https://doi.org/10.21775/cimb.026.01

Casares, F., & Mann, R. S. (1998). Control of antennal versus leg development in Drosophila. *Nature*, *392*(6677), 723–726. https://doi.org/10.1038/33706

Chourrout, D., Delsuc, F., Chourrout, P., Edvardsen, R. B., Rentzsch, F., Renfer, E., Jensen, M. F., Zhu, B., De Jong, P., Steele, R. E., & Technau, U. (2006). Minimal ProtoHox cluster inferred from bilaterian and cnidarian Hox complements. *Nature*, *442*(7103), 684–687. https://doi.org/10.1038/nature04863

De Luliis, G. (2011). The Pigeon. In *The Dissection of Vertebrates* (2nd ed., pp. 287–310). Academic Press Inc.

Deschamps, J., Van Den Akker, E., Forlani, S., De Graaff, W., Oosterveen, T., Roelen, B., & Roelfsema, J. (1999). Initiation, establishment and maintenance of Hox gene expression patterns in the mouse. *International Journal of Developmental Biology*, *43*(7), 635–650. https://doi.org/10.1387/ijdb.10668974

Deschamps, J., & van Nes, J. (2005). Developmental regulation of the Hox genes during axial morphogenesis in the mouse. In *Development* (Vol. 132, Issue 13, pp. 2931–2942). https://doi.org/10.1242/dev.01897

Dollé, P., Izpisúa-Belmonte, J. C., Falkenstein, H., Renucci, A., & Duboule, D. (1989). Coordinate expression of the murine Hox-5 complex homoeobox-containing genes during limb pattern formation. *Nature*, *342*(6251), 767–772. https://doi.org/10.1038/342767a0

Doudna, J. A., & Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. *Science*, *346*(6213). https://doi.org/10.1126/science.1258096

Gaunt, S. J. (2018). Hox cluster genes and collinearities throughout the tree of animal life. *International Journal of Developmental Biology*, *62*(11–12), 673–683. https://doi.org/10.1387/ijdb.180162sg

Hayward, A. G., Joshi, P., & Skromne, I. (2015). Spatiotemporal analysis of zebrafish hox gene regulation by Cdx4. *Developmental Dynamics*, *244*(12), 1564–1573. https://doi.org/10.1002/dvdy.24343

Hsu, P. D., Lander, E. S., & Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, *157*(6), 1262–1278. https://doi.org/10.1016/j.cell.2014.05.010

Ikeya, M., & Takada, S. (2001). Wnt-3a is required for somite specification along the anteroposterior axis of the mouse embryo and for regulation of cdx-1 expression. *Mechanisms of Development*, *103*(1–2), 27–33. https://doi.org/10.1016/S0925-4773(01)00338-0

Izpisua-Belmonte, J. C., Falkenstein, H., Dolle, P., Renucci, A., & Duboule, D. (1991). Murine genes related to the Drosophila AbdB homeotic gene are sequentially expressed during development of the posterior part of the body. *EMBO Journal*, *10*(8), 2279–2289. https://doi.org/10.1002/j.1460-2075.1991.tb07764.x

Jensen, E. (2014). Technical review: In situ hybridization. *Anatomical Record*, *297*(8), 1349–1353. https://doi.org/10.1002/ar.22944

Jiang, F., & Doudna, J. A. (2017). *CRISPR – Cas9 Structures and Mechanisms*. 505–531.

Kimmel, C., Ballard, W., Kimmel, S., Ullmann, B., & Schilling, T. (1995). Stages of embryonic development of the zebrafish. *Developmental Dynamics*, *203*(3), 253–310.

Lewis, E. (1978). A Gene Complex Controlling Segmentation in Drosophila. *Nature*.

Loureiro, A., & Da Silva, G. J. (2019). Crispr-cas: Converting a bacterial defence mechanism into a state-of-the-art genetic manipulation tool. *Antibiotics*, *8*(1). https://doi.org/10.3390/antibiotics8010018

Maeda, R. K., & Karch, F. (2009). Chapter 1 The Bithorax Complex of Drosophila. An Exceptional Hox Cluster. *Current Topics in Developmental Biology*, *88*(09), 1–33. https://doi.org/10.1016/S0070-2153(09)88001-0

Paik, E. J., Mahony, S., White, R. M., Price, E. N., Dibiase, A., Dorjsuren, B., Mosimann, C., Davidson, A. J., Gifford, D., & Zon, L. I. (2013). A Cdx4-sall4 regulatory module controls the transition from mesoderm formation to embryonic hematopoiesis. *Stem Cell Reports*, *1*(5), 425–436. https://doi.org/10.1016/j.stemcr.2013.10.001

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., & Lim, W. A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, *184*(3), 844. https://doi.org/10.1016/j.cell.2021.01.019

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., & Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nature Protocols*, *8*(11), 2281–2308. https://doi.org/10.1038/nprot.2013.143

RT-qPCR - Quantitative Reverse Transcriptase PCR. (n.d.). Sigma Aldrich.

Seo, H. C., Edvardsen, R. B., Maeland, A. D., Bjordal, M., Jensen, M. F., Hansen, A., Flaat, M., Weissenbach, J., Lehrach, H., Reinhardt, R., & Chourrout, D. (2004). Hox cluster disintegration with persistent anteroposterior order of expression in oikopleura dioica. *Nature*, *431*(7004), 67–71. https://doi.org/10.1038/nature02709

Shimizu, T., Bae, Y. K., Muraoka, O., & Hibi, M. (2005). Interaction of Wnt and caudal-related genes in zebrafish posterior body formation. *Developmental Biology*, *279*(1), 125–141. https://doi.org/10.1016/j.ydbio.2004.12.007

Skromne, I., Thorsen, D., Hale, M., Prince, V. E., & Ho, R. K. (2007). Repression of the hindbrain developmental program by Cdx factors is required for the specification of the vertebrate spinal cord. *Development*, *134*(11), 2147–2158. https://doi.org/10.1242/dev.002980

Sood, M. (1946). *The Anatomy of the Vertebral Column in Serpentes*. https://doi.org/10.1007/bf00416096

Soshnikova, N., & Duboule, D. (2009). Epigenetic temporal control of mouse hox genes in vivo. *Science*, *324*(5932), 1321–1323. https://doi.org/10.1126/science.1171468

Spitz, F., Gonzalez, F., & Duboule, D. (2003). A global control region defines a chromosomal regulatory landscape containing the HoxD cluster. *Cell*, *113*(3), 405–417. https://doi.org/10.1016/S0092-8674(03)00310-6

Stellwag, E. J. (1999). Hox gene duplication in fish. *Seminars in Cell and Developmental Biology*, *10*(5), 531–540. https://doi.org/10.1006/scdb.1999.0334

Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C., & Doudna, J. A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, *507*(7490), 62–67. https://doi.org/10.1038/nature13011

Struhl, G. (1981). A Homeotic Mutation Transforming Leg to Antenna in Drosophila. *Nature*.

Tschopp, P., Tarchini, B., Spitz, F., Zakany, J., & Duboule, D. (2009). Uncoupling time and space in the collinear regulation of Hox genes. *PLoS Genetics*, *5*(3), 1–12. https://doi.org/10.1371/journal.pgen.1000398

Veldman, M., & Shou. (2008). Zebrafish as a Developmental Model Organism for Pediatric Research. *Pediatric Research*, *64*(5), 470–476.

Young, T., & Deschamps, J. (2009). Chapter 8 Hox, Cdx, and Anteroposterior Patterning in the Mouse Embryo. *Current Topics in Developmental Biology*, *88*(09), 235–255. https://doi.org/10.1016/S0070-2153(09)88008-3

Young, T., Rowland, J. E., van de Ven, C., Bialecka, M., Novoa, A., Carapuco, M., van Nes, J., de Graaff, W., Duluc, I., Freund, J. N., Beck, F., Mallo, M., & Deschamps, J. (2009). Cdx and Hox Genes Differentially Regulate Posterior Axial Growth in Mammalian Embryos. *Developmental Cell*, *17*(4), 516–526. https://doi.org/10.1016/j.devcel.2009.08.010