THE UNIVERSITY OF CHICAGO

THE EVOLUTIONARY HISTORY OF *HOX* GENE EXPRESSION IN VERTEBRATE APPENDAGES

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF ORGANISMAL BIOLOGY AND ANATOMY

BY

ANDREW REECE GEHRKE

CHICAGO, ILLINOIS

JUNE 2016

To Mary,
for Everything.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

The era of genomics has profoundly aided our understanding of the genetic changes responsible for generating diversity of metazoan form. In contrast, access to a plethora of genomes (and developmental biology in the embryos they control) has further muddled the concept of homology. Morphologically disparate animals utilize the same core genetic toolkit during development, meaning that the same genes are utilized to form vastly different structures among taxa. Did these gene activities arise due to common ancestry, or did they arrive independently? In essence, how do we determine whether a genetic program in two individuals is homologous, or convergent?

This thesis attempts to answer this fundamental question using the fin to limb transition as a model paradigm. The pattern of activity of key genes (specifically *Hox*) that build the pectoral fins of fish and the limbs of tetrapods are extremely similar, yet the morphology of fins and limbs are very different. Did the common ancestor of fish and tetrapods contain this genetic program (homology), or were they invented separately in each group (convergence)? Here, I use comparative regulatory architecture as a foothold to elucidate this question, as homologous expression patterns are likely to utilize homologous regulatory elements.

Through comparative epigenomics, sequence analysis, and functional experiments, I find that fish utilize the same overall enhancer landscapes in the development of fins as the limbs of tetrapods. Thus, the genetic program that builds fins and limbs (*Hox*) were present and functional in the common ancestor of fish and tetrapods, implying concrete evidence for homology between specific segments in fins and limbs. Finally, I provide functional evidence to suggest that tetrapod digits and the fin rays of fish are built using a homologous population of cells, calling into question traditional views of the evolution of the wrist and digits.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Neil Shubin, for making the work presented in this dissertation possible. I am grateful for the scientific environment that he fostered: his excitement for the work was palpable and contagious, he demanded that we aim high, and (most importantly) he almost never said no to an experiment.

My committee was instrumental in the construction of this dissertation. I thank Marcelo Nobrega for pushing me to think above the level of the nucleotide, to Robert Ho for keeping my developmental arguments in check, and Ilya Ruvinsky for teaching me how to (properly) interpret transgenic experiments.

A special thanks to Jose Luis Gomez Skarmeta, and all of the members of the Skarmeta lab who assisted with the work presented here. JL was an unexpected, but ultimately perfect, collaborator. I thank him for all of his extensive help, and look forward to collaborating with him for the rest of my career.

I would like to express my gratitude to my fellow members of the Shubin lab, both past and present. I thank Noritaka Adachi, Justin Lemberg, Joyce Pieretti, and Darcy Ross for their help, friendship, and patience. I would like to extend special heartfelt thanks to Igor Schneider, John Westlund, and Tetsuya Nakamura, whom I had the pleasure of working closely with and benefited immensely from their expertise.

Thank you to all of my classmates, throughout the entire Biological Sciences Division, for their friendship, advice, and in general for pushing me to be better. A special thanks to Tim Sosa, Rob Arthur, Haley Stinnett, Kate Criswell, Mo Siddiq, Carrie Olson-Manning, and Andrew Manning. Finally, thank you Jon Mitchell, for making me a happier person, and for never turning down an invitation for whiskey.

CHAPTER I

INTRODUCTION

1.1 Homology in the era of genomics

Evolutionary developmental biology (evo-devo) is a field concerned with elucidating the genetic changes responsible for generating phenotypic variation, and is grounded in comparing development between and among taxa. Homology, in its most general sense, is defined as the presence of a feature in two individuals that was derived from a single feature of their common ancestor. The concept of homology has remained intimately tied to the field of evo devo, as developmental comparisons between two extant taxa are used to infer the state of their extinct common ancestor. Thus, nearly all work in evo-devo is held together by homology as a linchpin, making definitions and interpretations of homology absolutely crucial for interpreting experiments in evo-devo.

The manner in which biologists consider homology was greatly influenced by the discovery of the "molecular toolkit"– a group of developmental genes, often transcription factors (TFs), that are deeply conserved throughout nearly all metazoans (reviewed in (1)). The discovery that the incredible variety of animal form present in nature shares a common set of genes had a unifying effect on evolutionary biology, where differences in animal body plans were seen as minor variations on common structures (no matter how distantly related or phenotypically distinct). The discovery that morphological disparity in animal body plans persists despite genetic conservation has transformed the field of evolutionary biology. However, interpretations of the molecular toolkit during development have led to a number of controversial theories regarding homology.

The complex nature of homology at the molecular level can be seen in comparisons of appendage development across the animal kingdom. Shortly after the discovery of the toolkit, evolutionary biologists began examining gene expression from animals as distantly related as possible, to understand how common genes build disparate structures. Panganiban and experts from across the tree of life found that a common gene, *Distal-less* (*Dll*) is expressed in developing appendages from structures as diverse as annelid parapodia, onychophoran lobopodia, ascidian ampullae, enchinoderm tube feet, fish fins, and tetrapod limbs (2). Thus, appendages that are phenotypically and functionally distinct, in organisms that are phylogenetially unrelated, show strikingly similar patterns of expression of key genes. Not only are expression patterns conserved over large distances, but function can be as well. Rincon-Limas et al. found that the human ortholog (*Lhx2*) of a *Drosophila* wing-making gene (*apterous*) was able to completely rescue wing formation in flies that were lacking the native *apterous* gene (3). This level of conservation is not restricted to animal appendages, and a similar set of experiments and interpretations have been found when investigating the origin and evolution of eyes (reviewed in (4)).

These experiments point to an undeniable conservation of gene expression and function across the animal kingdom. However, interpretations of these experiments in the context of homology are nuanced and require more than sweeping generalization based on pattern. For instance, the most straightforward interpretation of these results would be that appendages (or eyes) are homologous across the animal kingdom. This makes sense at the level of the gene, in that it seems extraordinarily unlikely that each taxon would have acquired these genes independently (e.g. convergence) to make similar structures. On the other hand, there are profound differences between animal appendages in anatomy, structure, and even

development that cannot be ignored simply in the light of common gene expression. How, then, are we to interpret these seemingly contradictory data (deeply conserved gene networks vs. morphology and evolutionary history) to understand how body plans (and especially novelty) arise? Early qualitative interpretations of common gene expression in diverse phyla are beginning to be replaced by a more conservative approach to the notion of homology (4-6). A key feature of a modern understanding of homology in the age of genomics lies in defining "levels" – a pair of structures in two species may be homologous at the level of the gene regulatory network, yet non-homologous at the level of the overall phenotype or function (e.g. appendages and eyes across protostomes and deutorostomes) (4, 5).

Despite these advances in "homology thinking", it is still difficult to interpret the evolutionary history of gene expression patterns in two individuals, species, or even phyla. The simple question remains that if we observe similar genes being expressed in similar structures in two disparate organisms, did the activity arise by convergence or common ancestry? This thesis attempts to provide a tractable method to assigning homology in gene expression patterns by focusing on the regulatory networks that underlie gene activity. By comparing the regulatory inputs that drive the expression of genes in question, it may be possible to infer if the patterns were generated by convergence (different regulatory inputs) or by common ancestry (the same regulatory inputs). Here, I use the evolution of the tetrapod limb during the fin to limb transition as a system to utilize comparative regulatory logic to understand homology between features in fish fins and tetrapod limbs.

1.2 Homology of fins and limbs and the regulatory logic of limb development

Tetrapod limbs and the pectoral fins of living fishes share remarkably similar patterns of gene activity during their outgrowth and development (Figure 1.1). Despite these conserved developmental modules, the adult tetrapod forelimb and fish pectoral fin are phenotypically distinct, both in morphology and tissue type (Figure 1.1). Mouse forelimbs consist of one long bone (arm; stylopod), two long bones (forearm; zeugopod), a series of small nodular bones (wrist; mesopod), while terminating in a final set of long bones (digits; acropod). The entire limb is composed of a specific type of bone that is preformed in cartilage, called endochondral bone. The pectoral fins of fish are variable, but usually consist of a series of long proximal radials, followed by smaller spherical bones called distal radials, which serve as attachment points for long cylindrical bones called fin rays that make up the vast majority of the total fin (Figure 1.1).

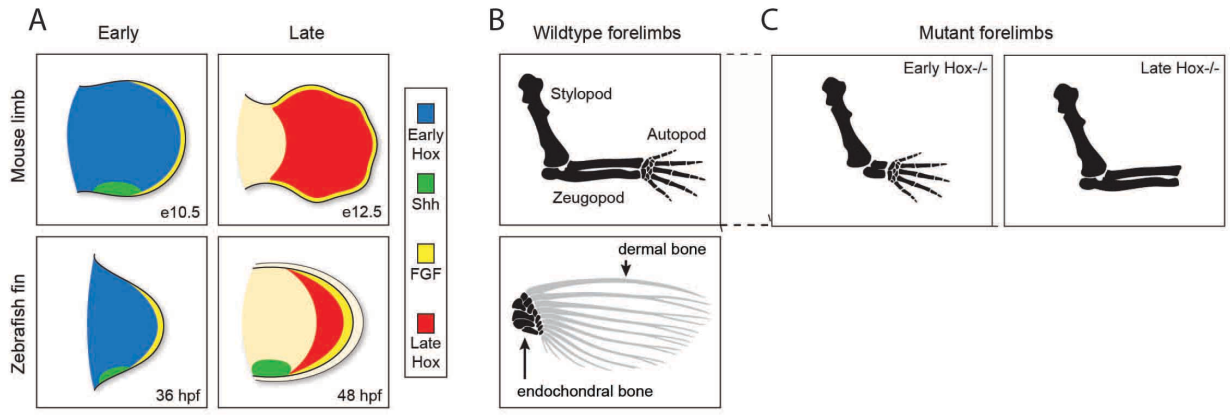


Figure. 1.1. Summary of vertebrate appendage development and mouse mutants highlighting the role of key genes. A) The forelimbs of mice and the pectoral fins of zebrafish utilize similar gene networks during development; *Fgfs* in the AER, *Shh* expression in the posterior, and early and late phases of *Hox* gene expression. Note the early fin fold in the 48 hpf zebrafish fin (shown in white), a structure that does not develop in tetrapods. B) Despite these conserved expression patterns, tetrapod limbs and zebrafish fins are morphologically disparate. The mouse limb is composed entirely of endochondral bone, while the zebrafish fin consists of a base of endochondrol bone followed by a long series of fin rays consisting of dermal bone. C) Mutant mice underscore the key nature of the genes discussed. Loss of early phase *HoxD* and *HoxA* expression leads to a loss of proximal structures of the limb, while loss of late phase expression results in complete deletion of the autopod.

In constrast to tetrapod limbs, fish fins are composed of two separate types of bone, where endochondral bone makes up the base of the fin (proximal and distal radials), while dermal bone (not preformed in cartilage) make up the fin rays.

While it is possible to make comparisons and assign homology between the bones of tetrapod limbs and closely related lobe-finned fossil fish (Sarcopterygians) (7, 8), these comparisons fall apart when considering ancient ray-finned (Actinopterygian) fish. There is no obvious way to assign homology (based on morphology alone) between the bones of tetrapod forelimbs and the fins of non-Sarcopterygian fish (e.g. zebrafish, pufferfish, salmon, etc.). Thus, evolutionary biologists have increasingly turned to developmental biology and genetics to potentially make sense of comparions between tetrapods and extant fish. Can we compare and contrast the genes involved in fin development to understand how bones are related in disparate taxa?

Due to their crucial role in limb development, the evolution of *Hox* gene expression has become a central point of study in deciphering molecular mechanisms behind the fin to limb transition and homology between structures. Tetrapods contain 4 clusters of *Hox* genes (A-D), which are responsible for patterning the primary axis in bilaterians. Two of these clusters, *A* and *D*, have crucial roles in building the secondary axis (limbs) in tetrapods. The limbs of mice are built via two successive "phases" of *HoxD* and *HoxA* expression: an "early" phase that is associated with the arm and forearm, and a "late" phase that orchestrates the construction of the wrist and digits (autopod) (9-11). Patterns of gene expression resembling early and late phase *hox* activity have been described in a number of fish species, including zebrafish, paddlefish, and catshark (12-14). Expression pattern alone cannot be taken as evidence for conservation of underlying gene regulatory mechanisms (15), which has shifted recent focus to the identification of early and late phase enhancers in fish genomes. Enhancers, often called *Cis*-Regulatory

Elements (CREs) are regions of the genome that do no code for proteins, and are instead bound by transcription factors (TFs) that contact the promoter of a target gene to drive expression in proper context in embryonic space and time (16, 17). Expression patterns are likely homologous in two species if the genes utilize the same regulatory architecture (CREs, TFs), the complexity of which makes it unlikely to have evolved by convergence (15).

The CREs that control the early and late phases of the *HoxD* cluster have been elucidated and examined in detail in mouse (9, 18) (Figure 1.2). Two enhancers residing 3' to the cluster control the early phase of *HoxD* expression, which is followed by a topological shift of chromatin allowing portions of the cluster to contact 5' enhancers that control late phase expression and autopod development.

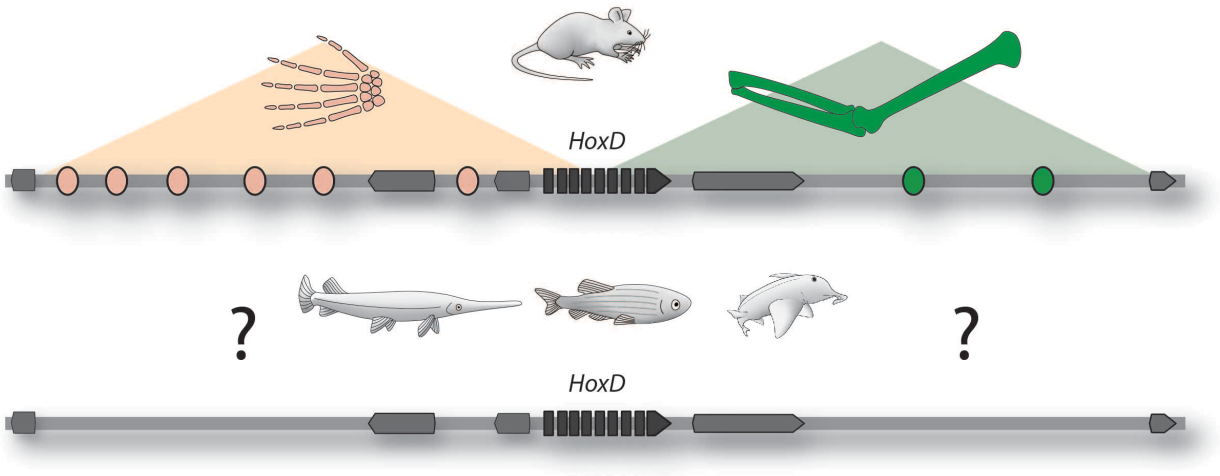


Figure 1.2. Thesis motivation. The regulatory systems that control *Hox* expression mouse limb development (top panel) are well understood. In contrast, the enhancers that dictate *Hox* activity in fish fins (bottom panel) are mostly unknown. Do fish genomes contain appendage enhancers, and if so, how do they function?

The digit "regulatory archipelago" contains as many as seven CREs, though not all may be involved in driving late phase *HoxD* expression (18). *HoxA* activity is controlled by a similar

regulatory topology, where a series of enhancers on one side of the cluster contact 5' *HoxA* genes to drive expression in developing digits (19).

With a detailed understanding of the regulatory logic that underlies *Hox* gene expression in mouse, we can ask the simple question: do fish genomes contain orthologs of these enhancers (Figure 1.2)? If so, how do they function? This thesis represents an attempt to answer these questions by providing a search for limb enhancers in fish genomes on an epigenomic, sequence, and functional level. By comparing and contrasting the regulatory systems in place in fins and limbs, it will be possible to make concrete statements about the evolutionary history of *Hox* gene expression in vertebrate evolution, in turn allowing a more granular understanding of homology between individual bones that make up fins and limbs.

CHAPTER II

THE GENOME OF THE SPOTTED GAR REVEALS AN EARLY PHASE ENHANCER IN FISH GENOMES

PUBLISHED IN *NATURE GENETICS* AS:

The spotted gar genome illumunates vertebrate evolution and and facilitates human-teleost comparisons.

Ingo Braasch, Andrew R. Gehrke, Jeramiah J. Smith, Kazuhiko Kawasaki, Tereza Manousaki, Jeremy Pasquier, Angel Amores, Thomas Desvignes, Peter Batzel, Julian Catchen, Aaron M. Berlin, Michael S. Campbell, Daniel Barrell, Kyle J. Martin, John F. Mulley, Vydianathan Ravi, Alison P. Lee, Tetsuya Nakamura, Domitille Chalopin, Shaohua Fan, Dustin Wcisel, Cristian Cañestro, Jason Sydes, Felix E. G. Beaudry, Yi Sun, Jana Hertel, Michael J. Beam, Federica Di Palma, Mario Fasold, Mikio Ishiyama, Jeremy Johnson, Steffi Kehr, Marcia Lara, John H. Letaw, Gary W. Litman, Ronda T. Litman, Masato Mikami, Tatsuya Ota, Nil Ratan Saha, Louise Williams, Peter F. Stadler, Han Wang, John S. Taylor, Quenton Fontenot, Allyse Ferrara, Stephen M. J. Searle, Bronwen Aken, Mark Yandell, Igor Schneider, Jeffrey A. Yoder, Jean-Nicolas Volff, Axel Meyer, Chris T. Amemiya, Byrappa Venkatesh, Peter W. H. Holland, Yann Guiguen, Julien Bobe, Neil H. Shubin, Jessica Alföldi, Kerstin Lindblad-Toh, & John H. Postlethwait

The work presented in this chapter comprises one section of the manuscript listed above. Experiments were designed by ARG, IS, and NHS. Experiments were performed by ARG with assistance from IS and TN. ARG wrote the text with IS and NHS, and ARG made the figures with image assistance from John Westlund.

2.1 Abstract

Members of the *HoxD* and *HoxA* clusters build mammalian limbs by two phases of gene expression commonly referred to as “early” and “late” phases that pattern proximal and distal limb segments, respectively (9, 18, 20). The regulatory systems responsible for the dual nature of *Hoxd* gene expression in limbs have been extensively studied in mouse, with emphasis on enhancers that drive the late phase of *HoxD* expression (21). Recent work has identified regulatory elements that drive the early phase of *HoxD* activation, in the form of two enhancers that lie on the telomeric side of the cluster deemed CNS39 and CNS65 (9). Early phase *HoxD*

expression in fins of fish and limbs of tetrapods shows several common features that are presumed homologous, and may derive from shared early phase *HoxD* regulatory elements (21). In order to understand early phase *Hox* gene regulation in fish, we performed multiple sequence alignments and functional assays using a variety of model and non-model systems. We find that an early phase enhancer (CNS65) is present in fish genomes, functions nearly identically to its ortholog in mouse, and can only be identified though alignments containing the gar genome. These data suggest that the proximal portions of tetrapod limbs and fish fins are built with at least a portion of a homologous regulatory network present in their common ancestor. Finally, our data are a testament to the utility of the unduplicated genome of the spotted gar in identifying the evolutionary history of vertebrate regulatory elements.

2.2 Introduction

The tetrapod arm is composed of three main segments: the stylopod (upper arm), zeugopod (arm and forearm), and the autopod (wrist and digits). During mouse limb development, genes from the *HoxA* and *HoxD* clusters serve as master regulators to build these segments via two waves of *Hox* expression, often referred to as "early phase", which builds the stylopod and zeugopod, and the "late phase", which builds the wrist and digits (10, 21). The regulatory networks that control these two phases of gene expression represent one of the best-understood systems of transcriptional dynamics in vertebrate development. During early mouse limb development, two enhancers (CNS39 and CNS65) that lie in a gene desert telomeric to the cluster drive early phase expression of 3' *Hoxd* genes (9). At approximately e10.5-e11.5, the *HoxD* cluster undergoes a topological shift, where the 5' domain harboring the early phase enhancers becomes closed, and the 5' region containing digit enhancers becomes open. This

shift results in contact between autopod enhancers in a gene desert centromeric to the *HoxD* cluster coming into contact with 5' *Hoxd* genes, building the wrist and the digits. Do fish genomes contain similar regulatory networks at *Hox* loci?

Previous work has focused mainly on detecting late phase enhancers in fish genomes, as they are responsible for driving gene expression in digits and likely relevant to the evolution of the autopod. In addition, previous work in the evolution of gene regulation in fish has focused mainly on teleost species, which have highly duplicated genomes that may cloud studies of regulatory evolution (Figure 2.1).

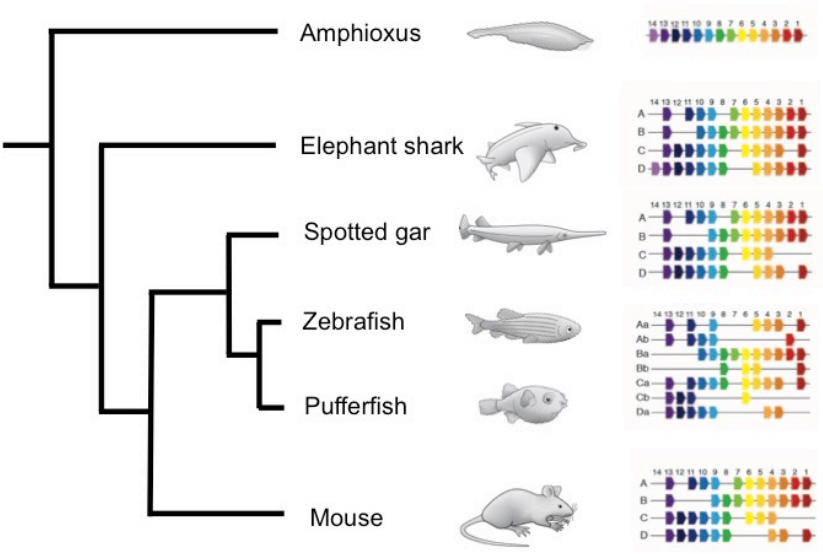


Figure 2.1. Vertebrate phylogeny and the unduplicated nature of the gar genome. Teleosts (e.g. zebrafish, pufferfish) have undergone a lineage-specific genome duplication, potentially clouding studies of regulatory evolution. In contrast, the gar genome contains 4 clusters of *Hox* genes, akin to the genomic complement to tetrapods.

In this study, we have focused on identifying and functionally analyzing potential early phase enhancers in the genomes of fish. Though early phase *HoxD* gene expression patterns are similar in developing limbs and fins, these patterns may have arisen convergently, which would point to independent origins to early phase *Hox* expression in fins and limbs. In order to differentiate between these two scenarios (homology vs. convergence), it is necessary to understand the regulatory control of early phase *Hox* expression in fish fins.

2.3 Results

To identify early phase enhancers in fish genomes, we first performed multiple sequence alignments between tetrapods, two teleost species (zebrafish, pufferfish), and gar. While our analysis failed to detect any sequence conservation of CNS39 with mouse as reference species, we were able to identify a peak of conservation of CNS65 in gar (Figure 2.2, B, middle panel). Interestingly, we were unable to identify the potential CNS65 ortholog when only teleost and mouse genomes were used in the alignment (Figure 2.2, B, left panel). Sequence conservation for this potential enhancer was revealed only when the gar genome was included (Figure 2.2, B, middle panel). Furthermore, the multiple sequence alignment produced peaks of conservation for both zebrafish and pufferfish when gar was used as the baseline (Figure 2.2, B, right panel).

To test these sequences for enhancer activity, we cloned CNS65 from gar and zebrafish into a tol2 vector with a minimal *cfos* promoter driving GFP, and injected the constructs into zebrafish embryos (22). We raised these embryos to sexual maturity to obtain stable lines, and found that both gar and zebrafish CNS65 drove strong, reproducible GFP signal in the developing fin at 38 hours post fertilization (hpf) (Figure 2.2, C and D). Gar CNS65 becomes active in the pectoral fins at 31 hpf, and drives expression throughout the fin until it becomes deactivated around 48 hpf (Figure 2.2, D).

To further test the potential of these elements to drive appendage expression in a tetrapod, we cloned CNS65 from gar and zebrafish into an Hsp68-LacZ vector and assessed its function in a mouse transgenic assay. We found that gar CNS65 drives strong LacZ expression in the developing forelimb at e10.5 (Figure 2.2, E, left), in a pattern that is indistinguishable from its mouse ortholog (9). The zebrafish ortholog was also able to drive LacZ in the forelimb at the

same time, with weaker activity than the gar enhancer (Figure 2.2, E right). As the activity of mouse CNS65 is absent in the developing autopod at later stages of limb development (9), we

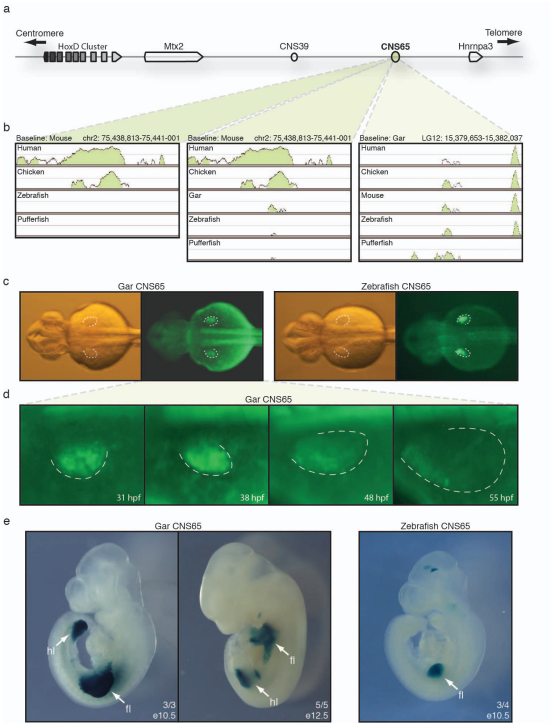


Figure 2.2. Identification and functional analysis of the gar and teleost early phase HoxD enhancer CNS65. a) Schematic representations of the mouse *HoxD* telomeric gene desert, containing two enhancers, CNS39 and CNS65, that drive early phase Hoxd expression in limbs. b) Using mouse as the baseline, Vista alignments of the HoxD gene desert show sequence conservation with human and chicken for CNS65, but not with teleosts (zebrafish, pufferfish) (left). An alignment that includes the gar gene desert, however, revealed a significant peak of conservation in the gar sequence (middle). Using the identified gar CNS65 as the baseline revealed CNS65 orthologs in zebrafish and pufferfish (right). c) Gar (left) and zebrafish (right) CNS65 orthologs are able to drive robust and reproducible GFP expression in zebrafish pectoral fins at 36 hours post fertilization (hpf). d) The pectoral fin activity of gar CNS65 begins at 31 hpf, drives activity throughout the fin, and becomes deactivated around 48 hpf. e) Gar CNS65 drives expression throughout the early mouse fore- and hindlimb (arrows) at stage e10.5 (left). At later stages (e12.5), gar CNS65 activity is restricted to the proximal portion of the limb, and is absent in developing digits (middle). Zebrafish CNS65 is able to drive reporter expression in mouse limbs at e10.5, but signal was detected only in developing forelimbs (right). Dotted lines outline the distal portion of the zebrafish pectoral fin (c,d). Number of LacZ positive embryos that showed limb signal is indicated at the bottom right; fl, forelimb, hl, hindlimb (e).

sought to test if gar CNS65 exhibited a similar spatiotemporal pattern during limb development. We harvested mouse embryos transgenic for gar CNS65 at e12.5 and found staining in proximal, not distal elements, again mimicking the pattern of the murine enhancer (Figure 2.2, E, middle).

2.4 Discussion

In this study, we have used a combination of phylogentic footprinting and functional transgenic assays to identify the presence of an early phase *Hoxd* enhancer, CNS65, in the genomes of osteichthyan (bony) fish. A key strategy in this work involved the inclusion of the spotted gar, which is an actinopterygian (ray-finned) fish that contains an unduplicated genome. Nearly all previously reported actinopterygian model systems used for genomics and development are teleosts (e.g. zebrafish, pufferfish, medaka), a group of fish containing over 96% of extant species and that have undergone a lineage-specific genome duplication (Figure 2.1). Highly duplicated genomes are prone to a significant amount of enhancer divergence, thus obscuring searches for regulatory elements based on sequence conservation with other organisms (23-26).

We were able to uncover sequence conservation for CNS65 only by including the genome of the spotted gar, demonstrating the value of a fish model with an unduplicated genome when attempting to identify vertebrate regulatory elements. Additionally, we were able use gar CNS65 as a baseline to identify CNS65 in zebrafish, showing that genome of the spotted gar can be used as leverage to discover regulatory elements in teleosts that have previously gone undetected.

When tested in transgenic zebrafish, CNS65 from both gar and zebrafish were able to drive reporter expression in pectoral fins markedly similar to the native early phase *Hoxd*

expression, essentially from 31-38 hpf and losing activity at approximately 48 hpf. Mouse embryos transgenic for gar CNS65 exhibited the same spatial and temporal dynamics of reporter expression as the endogenouse CNS65, indicating that the *cis*/*trans* dynamics of this enhancer have been highly conserved throughout vertebrate evolution. Zebrafish CNS65 was also able to drive expression at e10.5, but expression patterns were generally weaker and did not extend through the entire limb (nor hind limb). It is unclear whether the reduced expression from zebrafish CNS65 is due to something intrinstic about the *cis*/*trans* relationship when placed into transgenic mice, or to normal variation in signal due to mosacisim in mouse transgenesis. In either scenario, both gar and zebrafish CNS65 are able to drive expression in the forelimbs of mice akin to their murine orthologs.

Taken together, our data demonstrate that at least a portion of the early phase regulation of *Hoxd* gene expression in both limbs and fins is an ancestral, conserved feature of bony vertebrates (osteichtyhans), and highlight the advantage of utilizing the gar genome over highly derived teleosts to uncover functionally important genomic regulatory elements in vertebrates. Fins and limbs share homologous patterns of *Hoxd* expression in paired appendages, despite their differences in phenotype. Future work should determine whether the potential addition of enhancers (i.e. CNS39) over evolutionary time may have played a role in the transition of proximal bones in the fin to limb transition, or if loci other than *Hox* clusters played a more crucial role in the evolution of the tetrapod forelimb.

2.5 Materials & methods

Genomic alignments: Genomic segments of interest were downloaded from the Ensembl and UCSC genome databases and aligned using the mVista (LAGAN) program with the following

parameters; calc window: 100 bps, Min Cons Width: 100 bps, Cons Identity: 65%. Mouse assembly version 10 (mm10, Dec. 2011) was used in the alignment.

Cloning of gar and zebrafish CNS65: The conserved peaks were amplified by PCR using the following primers: gar- 5'-AAACGATCGCAGTGTTTCAGT-3', 5'-GTCTGGTGGCCTGTGTAAAAA-3' and zebrafish-5'- CCACTTAAACTGCGCATCAA-3', 5'-TGGATGAACCAGGTATTGCAG-3'. Genomic fragments were gel purified using the NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel), and subcloned into PCR8GW/GW/TOPO vector according to the manufacturer's protocols (Invitrogen). The gar and zebrafish CNS65 fragments were shuttled into either the pXIG-cFos-eGFP vector (22) for zebrafish transgenesis, or to the Gateway-Hsp68-LacZ vector (kind gift of Marcelo Nobrega, the University of Chicago) for mouse transgenesis, both using the Gateway system (Invitrogen). Vectors were confirmed by restriction digest and sequencing.

Zebrafish stable line transgenesis: Zebrafish embryos were collected from natural spawning, and staged according to standard conditions(27). Transposase RNA was synthesized using the mMessage mMachine SP6 kit (Ambion), using the pCS2-zT2TP vector that produces RNA that is codon-optimized for zebrafish (28). Solutions for injection were prepared according to ref.(22), and injected into the cytoplasm of 1 or 2 cell wild-type embryos. Embryos were maintained in egg water at 28^{o}C until visualization at the appropriate stage using a Leica M205FA microscope. Embryos displaying bright GFP signal were raised to sexual maturity and outcrossed to WT to identify founders.

Mouse transgenesis: Sequence-confirmed gar and zebrafish CNS65-Hsp68-LacZ vector were delivered to Cyagen Biosciences for mouse transgenesis (Cyagen Biosciences, Santa Clara, CA). Briefly, vectors were linearized with SalI, gel purified, microinjected into fertilized mouse oocytes and transferred to pseudo-pregnant females. Embryos were collected at e10.5 or e12.5, stained for beta-galactosidase activity, and genotyped using DNA extracted from yolk sac. Four embryos were PCR positive for gar CNS65, 3 of which were positive for LacZ activity in the limb. The zebrafish CNS65 injections resulted in 8 PCR positive embryos, 4 of which had LacZ staining in the limb

2.6 Acknowledgements

We thank Mayuri Chandran for technical assistance. This study was supported by National Institutes of Health Grant T32 HD055164 and National Science Foundation Doctoral Dissertation Improvement Grant 1311436 (A.R.G.), the Brinson Foundation and the University of Chicago Biological Sciences Division (N.H.S.)

CHAPTER III

CONSERVATION AND DEVELOPMENTAL SYSTEMS DRIFT OF EARLY PHASE *HOX* ENHANCERS IN THE EVOLUTION OF PAIRED APPENDAGES

TO BE SUBMITTED TO *PROCEEDINGS OF THE ROYAL SOCIETY B*. AS:

Conservation and developmental systems drift of early phase *Hox* enhancers in the evolution of paired appendages

Andrew R. Gehrke*, Igor Schneider*, Juan J. Tena, Elisa de la Calle-Mustienes, José Luis Gómez-Skarmeta and Neil H. Shubin.

*co-1st authors

The work presented in this chapter was designed by ARG, IS, and NHS. Experiments were performed by ARG and IS, except for zebrafish 4C-seq and ATAC-seq which was performed by JJT, ECM, and JLGS. The manuscript was written by ARG, IS, and NHS. Figures were made by ARG with input from IS and NHS.

3.1 Abstract

The proximal bones of fish fins and tetrapods limbs are patterned by similar genetic and regulatory cascades, implying that they are homologous structures. Despite this relationship, the proximal radials of fish fins and the arm and forearm of tetrapod limbs are phenotypically distinct. The proximal segments of mouse limbs are built via an initial ("early") wave of *HoxD* and *HoxA* gene expression. *HoxD* early phase expression in mouse limbs is driven by two enhancers (CNS65 and CNS39), one of which (CNS65) has been identified in teleost fish, suggesting partial conservation of this regulatory system among vertebrates. In order to determine the phylogenetic origin of this regulatory program, we probed the elephant shark (*Callorhinchus milii*) genome using epigenomic profiling and transgenic assays. We find that CNS65 in elephant shark can be identified by sequence conservation, is marked by an area of open chromatin in native developing elephant shark fins, and can drive reporter expression in

transgenic mouse limbs identical to the native murine enhancer. Furthermore, when incorporating elephant shark into multiple sequence alignments, we were able to identify an ortholog of CNS39 in both elephant shark and gar. While elephant shark CNS39 was unable to drive reporter expression in transgenic zebrafish and mice, CNS39 from gar drove robust fin expression in both transgenic hosts. Taken together, our data point to an ancient regulatory system for early phase *Hox* expression shared between tetrapods and fish that consisted of at least two enhancers. The presence of both CNS65 and CNS39 in fish genomes argues against an "enhancer addition" hypothesis to explain the transformation of proximal structures during the fin to limb transition. Finally, we argue that the disparity found in reporter activity in transgenic animals is due to developmental systems drift between donor and host, underscoring the need for testing donor DNA from multiple species when characterizing the evolution of enhancers.

3.2 Introduction

While paired fins and limbs are homologous appendages, the molecular mechanisms behind the fin to limb transition are largely unknown. *Hox* genes have remained a central focus in discussions of the fin to limb transition due to their necessary role in building all three main segments of the tetrapod limb (9, 10). Recent work has focused on the evolution of the autopod (wrist/ankle and digits), yet the proximal bones of fish fins and tetrapods limbs also differ drastically in morphology, and the genetic changes responsible for their transition remains largely unexplored. The development of the stylopod and zeugopod of mouse limbs is orchestrated by 3' *Hox* genes that initiate early in limb bud patterning (9). Loss of these genes via genetic knockout results in the ablation of specific segments of the limbs, e.g. the zeugopod, underscoring their importance to limb development (10). Two enhancers that lie telomeric to the

cluster, CNS65 and CNS39, control this early wave of transcription in the *HoxD* cluster (9). While CNS65 is conserved in bony fish, the extent to which the enhancer and its activity are retained in more distantly related vertebrates is unknown (29). Furthermore, the pattern of gain/loss of CNS39 is unclear, as previous reports failed to detect the enhancer by sequence conservation in bony fish (29).

In order to understand the evolutionary history of early phase *Hox* regulation, we utilized embryos and the sequenced genome of the elephant shark, (*Callorhinchus milii*), a cartilaginous fish that represents a sister group (chondricthyes) to bony fish (osteichtyes) (30). Through a combination of epigenomic profiling, sequence alignment, and interspecies transgenesis, we find that both CNS65 and CNS39 are present in the genome of elephant shark, representing a conserved regulatory network for early phase *Hox* expression in vertebrate appendages. The identification of CNS39 through the use of elephant shark in sequence alignments allowed us to functionally characterize this enhancer in the spotted gar, *Lepisosteus oculatus*. Interestingly, we find that elephant shark enhancers are only able to drive expression in specific transgenic hosts, while those of spotted gar function in both zebrafish and mouse. This diversity of function in transgenic swaps is likely due to developmental systems drift in both the cis architecture of the elephant shark donor DNA and the trans environment of the transgenic host, rather than the inability of the enhancer to function in the native environment. The data presented here suggest that the regulatory elements responsible for patterning the proximal elements of vertebrate appendages are conserved between fish and tetrapods, and that "enhancer addition" did not cause the transformation of proximal radials into the arm and forearm during the fin to limb transition.

3.3 Results

In order to investigate the regulatory landscape of early phase expression in a cartilaginous fish, we performed multiple sequence alignments with the elephant shark genome, concentrating on the gene desert telomeric to the *HoxD* cluster that harbors early phase enhancers in mouse (9). We found a peak of sequence conservation for CNS65 in the elephant shark genome (Figure 3.1). Surprisingly, we also found a conservation peak for CNS39 in elephant shark, which we were unable to detect significantly in previous alignments with teleost species including gar (Figure 3.1).

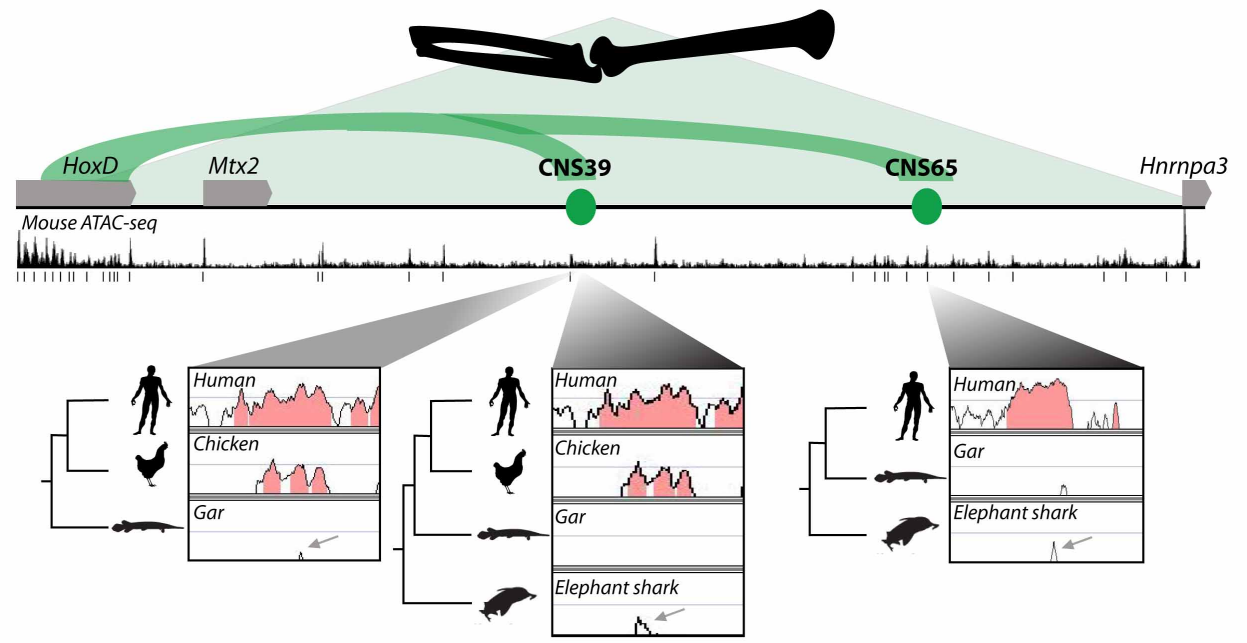


Figure 3.1. Enhancer landscape, epigenetic profile, and vertebrate sequence conservation of early phase *HoxD* regulation in mouse. Top panel depicts two regulatory elements (CNS65 and CNS39) that drive early phase *HoxD* expression in mouse. ATAC-seq in mouse limbs was able to identify a significant peak of open chromatin for CNS65, but not CNS39. Bottom panel shows VISTA plots of sequence conservation utilizing different species as inputs. A peak of sequence conservation for both CNS39 and CNS65 was detected for elephant shark (middle and right panel, denoted by arrows).

Next, we sought to functionally validate these early phase enhancers. We searched our previously reported ATAC-seq (a genome wide assay to detect putative regulatory elements based on chromatin accessibility) data for mouse limbs and whole body zebrafish to determine if these regulatory elements coincide with open chromatin (31, 32). We found a significant peak in our mouse autopod ATAC-seq data for CNS65 but not for CNS39 (Figure 3.1). In zebrafish, CNS65 is also marked by a robust ATAC-seq peak (Figure 3.2, A). In order to validate that CNS65 from zebrafish is contacting 3' *Hox* genes, we performed 4C-seq on whole zebrafish embryos using *hoxd4* as a viewpoint. Zebrafish CNS65 contacts *hoxd4* at 24 hours post fertilization (hpf), consistent with its role as *bona fide* early phase enhancer in fish (Figure 3.2, A).

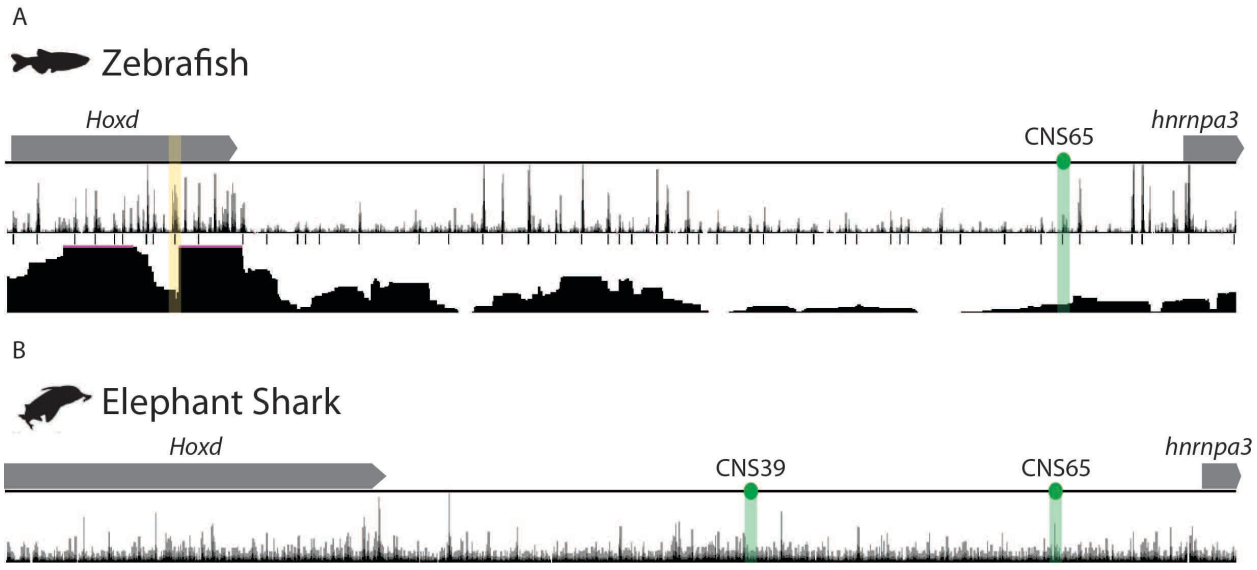


Figure 3.2. Epigenomic profile and enhancer contact of early phase *HoxD* landscape in zebrafish and elephant shark. A) ATAC-seq in zebrafish embryos identifies a significant peak of open chromatin for CNS65. In addition, 4C-seq experiments reveal that this region contacts the promoter of *hoxd4* at 24 hpf, proving an interaction between this enhancer and *Hox* genes involved in the early phase of expression. B) ATAC-seq in embryonic fins of elephant shark at stage ~25 was able to detect open chromatin at the ortholog of CNS65, but not CNS39.

In order to attain functional data for elephant shark, we performed ATAC-seq on elephant shark pectoral fins at ~stage 26 of development (33). We chose (and were limited to) ATAC-seq for epigenetic profiling as elephant shark embryos are extremely rare, and ATAC-seq is a fast and reliable technique that requires a small input of cells (31). Elephant shark CNS65 is significantly marked by ATAC-seq, suggesting that the enhancer is active and lies in an area of open chromatin during cartilaginous fin development (Figure 3.2, B). In contrast, CNS39 from elephant shark was not detected in our ATAC-seq data (Figure 3.2, B). While this could suggest that the enhancer is not active at the developmental time tested, we were also not able to detect CNS39 in developing mouse limbs using ATAC-seq. Thus, ATAC-seq seems unable to detect this enhancer in multiple species tested, even in mouse where it has been validated extensively by both 4C-seq and transgenic assays (9).

Finally, we tested the putative regulatory elements identified by sequence conservation and epigenomics in transgenic assays. CNS65 from elephant shark drove robust reporter gene expression in the forelimbs of transgenic mice at e10.5, in a pattern that recapitulates the endogenous mouse enhancer (9) Figure 3.3, B). Surprisingly, elephant shark CNS65 was unable to drive consistent reporter expression in the pectoral fins of transgenic zebrafish (1/5 transgenic lines with pectoral fin signal). One mouse transgenic for elephant shark CNS39 showed expression in the fore and hind limbs, but the low number of LacZ positive embryos makes this result difficult to interpret (9 embryos genotype positive, 2 embryos with LacZ staining, 1 embryo with limb expression) (Figure 3.3, B). Elephant shark CNS39 was also unable to drive expression in the pectoral fins of zebrafish (0/3 stable lines). We reasoned that the inability of elephant shark CNS39 to function in transgenic assays could be due to either a genuine lack of activity, or due to *cis*/*trans* divergence between donor and host genomes (i.e. both zebrafish and

mouse *trans* environments cannot understand the elephant shark *cis* element). In order to test this hypothesis, we cloned a putative CNS39 from the gar genome, using a small peak of sequence conservation as a guide (Figure 3.1). We found that this putative element from the gar genome was able to drive robust

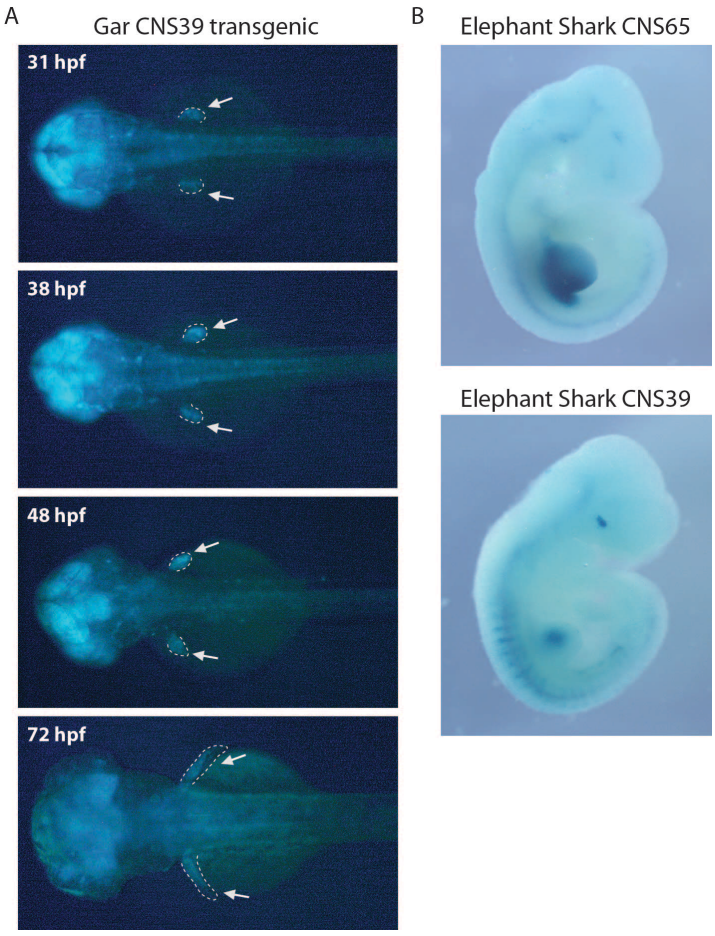


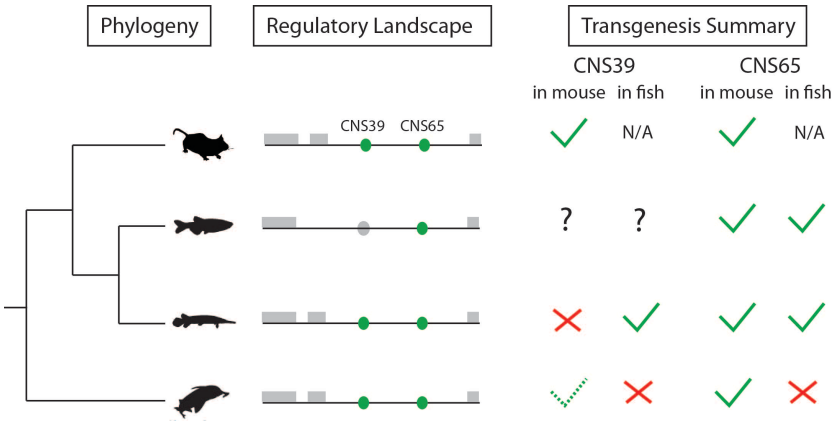
Figure 3.3. Inter-species transgenesis of early phase *HoxD* enhancers. A) Putative CNS39 (see Figure 3.1) from gar was able to drive reporter expression in an early phase like pattern in transgenic zebrafish. Expression in the pectoral fins (white arrows) began at 31 hpf, and was expressed robustly and specifically throughout the fin until ~48 hpf. B) CNS65 from elephant shark was able to drive strong reporter expression in the forelimbs of transgenic mice, nearly identically to its ortholog in teleosts and tetrapods. Elephant shark CNS39 was able to drive expression in mouse forelimbs, but small sample size makes this experiment difficult to interpret (see text).

reporter gene expression in the pectoral fins of transgenic zebrafish at 38 hpf, but was unable to drive limb expression in the limbs of developing transgenic mice (0/3 genotype positive with LacZ staining elsewhere in the embryo) (Figure 3.3, A).

3.4 Discussion

Here we present a phylogentic perspective on the evolution of early phase *Hox* expression, providing a more comprehensive scenario of which regulatory elements were present and active in the ancestor of gnathostomes. We find that the elephant shark genome harbors orthologs of both murine early phase *Hox* enhancers, and that CNS65 has a deeply conserved function throughout all of jawed fish.

Our transgenic results show instances of both deep conservation and functional drift of enhancers, even in regulatory elements that act upon the same locus in the same species (Figure 3.4). For instance, elephant shark CNS65 is able to drive strong expression in the forelimbs of transgenic mice, but is unable to drive activity in the pectoral fins of transgenic zebrafish. CNS39 from elephant shark has no appendage activity in both transgenic mice and zebrafish. This variety of activation potential is likely not due to properties intrinsic to the enhancer, but due to developmental systems drift (DSD) in donor and host species (34, 35).



For instance, the fact that CNS65 from elephant shark functions in mouse and not zebrafish

Figure 3.4. Conservation and drift in the evolution of early phase *HoxD* enhancers. Summary of transgenic experiments shows the diversity of function of early phase enhancers depending upon transgenic host. CNS65 from gar functions in zebrafish and mouse transgenic hosts, while CNS39 is able to drive expression only in transgenic fish. CNS65 from elephant shark can drive expression in mouse limbs, but not in transgenic fish fins. CNS39 from elephant shark is unable to drive robust expression in either mouse or zebrafish hosts.

likely indicates that elephant shark and mouse share the ancestral *cis*/*trans* relationship, while zebrafish has undergone divergence in its *trans* regulators. CNS39 from gar appears to exhibit the opposite scenario: it is able to drive fin expression in transgenic zebrafish, but not in the forelimbs of transgenic mice, this time pointing to divergence in the mouse *trans* environment (Figure 3.4). Elephant shark CNS39 is a more extreme example of this divergence, as the *trans* systems of both zebrafish and mouse cannot decode the *cis* information of the elephant shark enhancer (Figure 3.4). Still, it is possible that elephant shark does not contain a *bona fide* CNS39 enhancer, which is difficult to ascertain experimentally as transgenesis using elephant shark as a host is currently impossible. However, given the presence of the enhancer by sequence conservation in elephant shark and the functional activity of the gar ortholog in zebrafish, we propose that CNS39 is an ancient feature of gnathostomes, and that DSD between the donor enhancer and its transgenic hosts is a more parsimonious explanation for lack of activity.

Taken together, our data point to a deeply conserved regulatory system controlling early phase *Hox* expression in gnathostomes. These data have implications for understanding homology between the proximal bones of osteicthyan and chrondricthyan fins as well as the mechanisms behind the fin to limb transition in general. The shared regulatory control of early phase *HoxD* expression in bony and cartilaginous fish indicates that these expression patterns are homologous in gnathostomes. Thus, the proximal bones of sharks and rays, bony fish, and tetrapods are all built according to a common developmental plan, despite later divergence in morphology according to other downstream molecular pathways. Finally, the conserved landscape of early phase regulation indicates that sequential addition of *HoxD* regulatory elements likely did not play a causative role in transformation of proximal fin to proximal limb.

The two enhancers found to control early phase regulation in mouse limbs are likely to represent the total complement of early *HoxD* enhancers, as their discovery was based on a comprehensive analysis including 4C-seq, extensive transgenesis, and functional studies using genomic inversion (9). Thus, the entire early phase regulatory landscape for the *HoxD* cluster was likely present in the common ancestor of gnathostomes. It is possible that alterations in the number or activity potential of *HoxA* enhancers throughout evolution has a causative role in shaping proximal bones, but the regulatory elements driving early *HoxA* expression have yet to be discovered. While the addition of *HoxD* enhancers throughout the fin to limb transition may have played a role in the evolution of the autopod (36), it will be important for future studies to consider loci other than *Hox* in the evolution of the stylopod and zeugopod.

3.5 Materials & methods

Genome alignments: Human, mouse, chicken, and elephant shark genomic segments (from *Hoxd13* to *Hnrnpa3*) of interest were downloaded from the UCSC genome database, and the gar genome from the ENSEMBL database. Alignments were made using the mVista (LAGAN) program with the following parameters; calc window: 100 bps, Min Cons Width: 100 bps, Cons Identity: 70%.

Elephant shark ATAC-Seq: ATAC-seq experiments were performed as previously described (31, 32). Approximately 75,000 cells were isolated from the pectoral fin of an anesthetized elephant shark embryo at ~stage 26 (33) and used for ATAC-seq. The resulting library was sequenced in a Hiseq 2000 pair end lane, and reads were aligned to elephant shark (calMil1 assembly Dec

2013). Statistically significant peaks were determined by extending ATAC-seq reads to 100bp and using MACS software (37) with default parameters.

Circular chromosome conformation capture: Circular chromosome conformation capture-seq (4C-seq) assays were performed as previously reported (38-41). Five hundred zebrafish embryos at 48 hpf were dechorionated using pronase and disrupted in 1 ml of Ginzburg Fish Ringers (55 mM NaCl, 1.8 mM KCl, 1.25 mM NaHCO3). Isolated cells were lysed (lysis buffer: 10 mM Tris-HCl pH 8, 10 mM NaCl, 0.3% IGEPAL CA-630 (Sigma), 1X protease inhibitor cocktail (cOmplete, Roche) and the DNA digested with DpnII (NEB) and Csp6I (Fermentas, Thermo Scientific) as primary and secondary enzymes, respectively. T4 DNA ligase (Promega) was used for both ligation steps. Specific primers were designed at gene promoters, and Illumina adaptors were included in primer sequence. 8 PCRs were performed with Expand Long Template PCR System (Roche) and pooled together. This library was purified with a High Pure PCR Product Purification Kit (Roche), its concentration measured using the Quanti-iTTM PicoGreen dsDNA Assay Kit (Invitrogen, P11496) in a Qubit machine and sent for deep sequencing in an Illumina Hiseq 2000 machine multiplexing 10 samples per lane. 4C-seq data were analysed as previously described (40). Briefly, raw sequencing data were demultiplexed using the primer sequences as barcodes and aligned using zebrafish July 2010 assembly (danRer7) as the reference genome. Reads located in fragments flanked by two restriction sites of the same enzyme, or in fragments smaller than 40 bp were filtered out. Mapped reads were then converted to reads-per-first-enzyme-fragment-end units, and smoothed using a 30 fragment mean running window algorithm. Embryos from three different stages (80% epiboly, 24 hpf, and 48 hpf) were used for the *Hoxd4* promoter, where similar contact patterns were found when comparing

stages.

Enhancer cloning: Elephant shark CNS39, CNS65, and gar CNS39 were ordered as gBlocks fragments from Integrated DNA Technologies (IDT), sequences can be found in Table S3.1. Blunt oligos from IDT were A-tailed by incubating the oligo with 0.2 mM dATP (Thermo Fisher Scientific), 1.5 mM MgCl2 (Promega), in 1X GoTaq Reaction buffer (Promega) with 5 units of GoTaq Flexi DNA polymerase (Promega) in a total volume of 10 uL for 30 minutes at 70^{o}C. Two uL were subsequently used for TA cloning with the pCR8/GW/TOPO vector according to the manufacturers protocols (Thermo Fisher Scientific). Fragments were then transferred via Gateway system LR reaction (Invitrogen) into either the pXIG-cFos-eGFP vector for zebrafish transgenesis or to a Gateway-Hsp68-LacZ vector for mouse transgenesis. Final destination vectors were confirmed by restriction digest and sequencing.

Zebrafish and mouse transgenesis: All zebrafish work was performed according to standard protocols approved by The University of Chicago (ACUP #72074). *AB zebrafish embryos were collected from natural spawning and injected according to (22). Transposase RNA was synthesized from the pCS2-zT2TP vector using the mMessage mMachine SP6 kit (Ambion) (26). Approximately 120 embryos were injected at the one to two cell stage with ~2 nL of injection solution, resulting in ~30 F0 adults. Potential founders were outcrossed to *AB fish to identify transgenic founders by visualization of GFP or RFP. We produced a minimum of 3 founders per construct tested. Mouse transgenesis was performed by Cyagen Biosciences (Cyagen Biosciences, Santa Clara, CA) using an enhancer-Hsp68-LacZ vector. Vectors were linearized with SalI, gel purified, microinjected into fertilized mouse oocytes, and transferred to

pseudopregnant females. Embryos were collected at e10.5, stained for β-galactosidase activity, and genotyped with LacZ primers using DNA extracted from yolk sacs. Transgenic embryos from fish and mouse were visualized using a Leica M205FA microscope

3.6 Acknowledgements

We thank John Westlund for assistance in figure preparation, and Noritaka Adachi for assistance with ATAC-seq. Elephant shark embryos were graciously provided by Dr. Peter Currie, Alysha Heimberg, and Steve Mcleod, department of Australian Regenerative Medicine Institute (ARMI), Monash University, Clayton, Victoria, Australia. This study was supported by National Institutes of Health Grant T32 HD055164 and National Science Foundation Doctoral Dissertation Improvement Grant 1311436 (A.R.G.), the Brinson Foundation and the University of Chicago Biological Sciences Division (N.H.S.)

CHAPTER IV

DEEP CONSERVATION OF WRIST AND DIGIT ENHANCERS IN FISH

PUBLISHED IN *THE PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES* AS:

Deep conservation of wrist and digit enhancers in fish (2014).

Gehrke AR, Schneider I, de la Calle-Mustienes E, Tena JJ, Gomez-Marin C, Chandran M, Nakamura T, Braasch I, Postlethwait JH, Gomez-Skarmeta JL and Shubin NH

The work in this chapter was designed by ARG, IS, and NHS. Zebrafish transgensis experiments were performed by ARG, IS, MC, and TN. Mouse and zebrafish fin ATAC-seq was performed by ARG. Whole body zebrafish ATAC-seq and 4C-seq was performed by ECM, JJT, CGM, and JLGS. Gar genome sequence was provided by IB and JHP. The manuscript was written by ARG with input from IS and NHS. Figures were designed by ARG with image assistance from John Westlund.

4.1 Abstract

There is no obvious morphological counterpart of the autopod (wrist/ankle and digits) in living fishes. Comparative molecular data may provide insight into understanding both the homology of elements and the evolutionary developmental mechanisms behind the fin to limb transition. In mouse limbs, the autopod is built by a "late" phase of *Hoxd* and *Hoxa* gene expression, orchestrated by a set of enhancers located at the 5' end of each cluster. Despite a detailed mechanistic understanding of mouse limb development, interpretation of *Hox* expression patterns and their regulation in fish have spawned multiple hypotheses as to the origin and function of "autopod" enhancers throughout evolution. Using phylogenetic footprinting, epigenetic profiling, and transgenic reporters, we have identified and functionally characterized *hoxD* and *hoxA* enhancers in the genomes of zebrafish and the spotted gar, *Lepisosteus oculatus*, a fish lacking the whole genome duplication of teleosts. Gar and zebrafish "autopod" enhancers drive expression in the distal portion of developing zebrafish

pectoral fins, and respond to the same functional cues as their murine orthologs. Moreover, gar enhancers drive reporter gene expression in both the wrist and digits of mouse in patterns that are nearly indistinguishable from their murine counterparts. These functional genomic data support the hypothesis that the distal radials of bony fish are homologous to the wrist and/or digits of tetrapods.

4.2 Introduction

The origin of novel features is a key question in evolutionary biology, and the autopod–wrists, fingers, ankles and toes–is a hallmark example (6). While paleontological data, such as that from the Devonian lobe fin *Tiktaalik roseae,* reveal a sequence of changes in the elaboration of the bony elements of fins into limbs (7), in living taxa there is a lack of obvious homology between the wrist and digits of tetrapod limbs and the pectoral fin skeleton of extant fish (21). Tetrapod forelimbs are generally composed of a series of long bones (upper arm and forearm), followed by small nodular bones (wrist), and ending in another group of long bones (digits). Ray-finned (Actinopterygian) pectoral fins are diverse, but are usually composed of a series of long proximal radials, followed by a set of smaller distal radials. The open question remains: do extant fish have the equivalent of wrists or digits?

The molecular mechanisms governing the development of mammalian limbs have been approached in mouse models through multiple levels of analysis, from chromatin dynamics, to enhancer sequence, to gene expression patterns (42). Murine limbs display two successive phases of gene expression of the *HoxD* and *HoxA* gene clusters. The initial, or "early" phase of expression begins with members at the 3' end of the clusters being expressed broadly, and members at the 5' end of the cluster being expressed in an increasingly restricted number of cells

(10). This "early" phase of *Hox* expression is associated with the development of the upper arm (stylopod) and forearm (zeugopod). The initial wave of *Hox* expression is followed by a temporally distinct second activation of *Hox* genes, this time beginning with members of the 5' end of the cluster being expressed most broadly and in the presumptive digits. This second, "late" phase of expression is necessary for autopod formation, as evidenced by a loss of this domain resulting in deletion of the wrist and digits (10). The genomic regulatory elements and chromatin dynamics responsible for enacting these two phases have been studied in detail in the *HoxD* cluster, where the "early" and "late" phases are governed by enhancers that lie on opposite sides of the *HoxD* cluster–3' and 5', respectively–and activated in turn by shifting domains of open chromatin (9, 18). In addition, recent work has identified a series of enhancers that drive late phase expression of the *HoxA* cluster in the developing mouse autopod in a fashion similar to that of *HoxD* (19).

To what extent are the regulatory mechanisms that drive autopodial development present in fish fins and, if they are present, what is their developmental role? Previous work has shown that at least one of the "autopod" enhancers (CsB) is present and active in the common ancestor of gnathostomes (43). Additionally, recent work in zebrafish has shown that the early and late topological chromatin domains are indeed observed in bony fish (44). However, teleost fish enhancer domains were unable to drive reporter gene expression in the developing digits of transgenic mice, suggesting that while bony fish do contain a version of the autopod regulatory apparatus, these enhancers are not responsive to the regulatory program present in murine digits (44). Thus, the number, extent, and function of "limb" enhancers in fish remain to be fully explored, especially in fish species outside those of traditional model systems that might resemble ancestral characters more closely than teleosts.

To address these issues, we used a combination of epigenetic profiling in zebrafish and phylogenetic footprinting utilizing the genome assembly of the recently sequenced spotted gar (*Lepisosteus oculatus*) (23) to investigate the enhancers that control *hoxd* and *hoxa* expression in bony fish. The phylogenetic position of gar is crucial to our investigation, in that gar represents a lineage that diverged from teleost fishes before the teleost genome duplication (TGD), an event that may cloud studies of regulatory evolution (Figure 2.1 and Figure 4.1) (24-26). The data presented here reveal an unprecedented and previously undescribed level of deep conservation of the vertebrate autopod regulatory apparatus, suggesting homology between the distal radials of bony fish and the autopod of tetrapods.

4.3 Results

To identify orthologs of murine *Hox* limb enhancers in bony fish, we performed a multiple sequence alignment of the genomic region upstream of the *HoxD* and *HoxA* clusters from a number of tetrapod and teleost species (Figure. 4.1A, and Figure 4.2). Our initial survey revealed a dearth of sequence conservation for late phase enhancers in teleost species, aside from previously studied CsB, and an additional limb enhancer previously called Island III (18, 43). As we broadened our taxonomic input by including the genome of the spotted gar, we found peaks of conserved non-coding elements (CNEs) for the *HoxD* late phase digit enhancer Island I (Figure. 4.1A) and the *HoxA* late phase enhancers e16, e13, and e10 (Figure 4.2) (18, 19). We reasoned that the unduplicated nature of the gar genome makes its *Hox* clusters a better representative of the common ancestor of bony fish (Osteichthyes), revealing sequences that have diverged beyond recognition in derived, duplicated teleost genomes (24).

To identify potential orthologs in the zebrafish genome that are not revealed by sequence

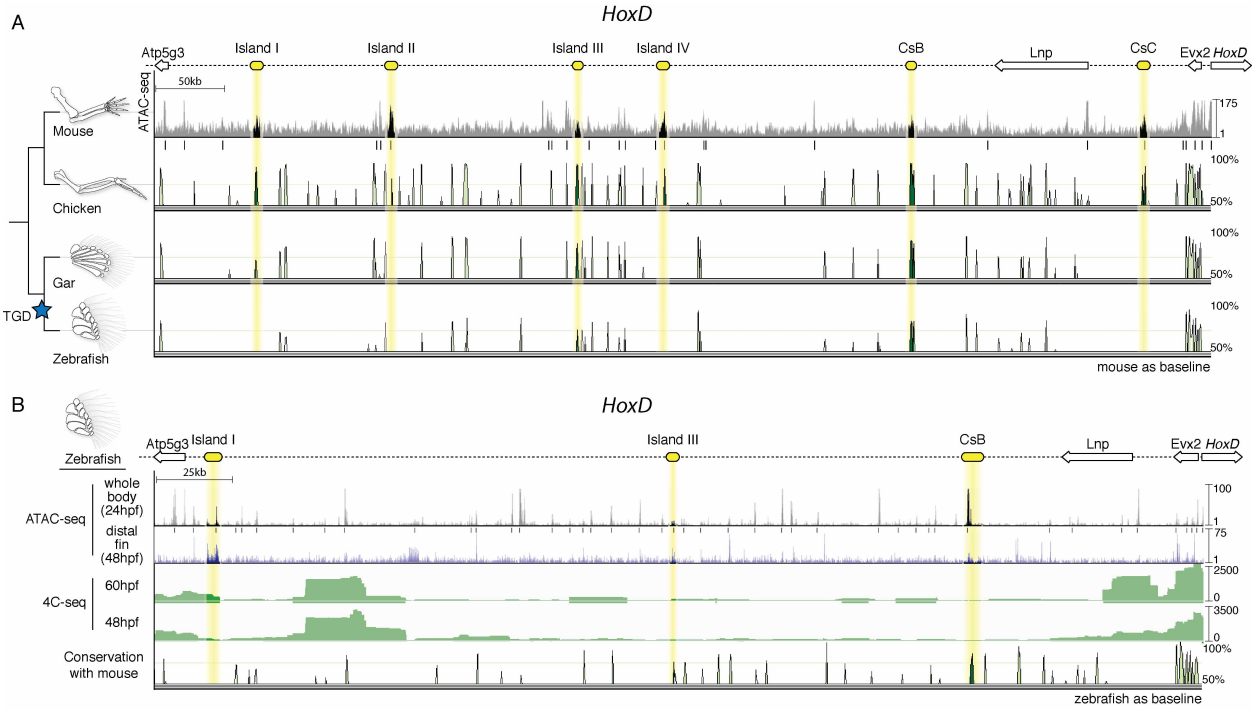


Figure 4.1. Chromatin state and sequence conservation of the *HoxD* autopod "regulatory archipelago" gene desert among select vertebrates. (A) (Top) Schematic representation of the HoxD centromeric gene desert, with cis-regulatory "islands" active in mouse denoted in yellow. ATAC-seq data are shown for mouse autopods at e12.5, providing a view of open chromatin. Statistically significant peaks are denoted by black bars. Sequence conservation is shown below for chicken, gar, and zebrafish. Note that sequence conservation for Island I is only found in gar, a nonteleost actinopterygian, and not in the teleost zebrafish. A blue star marks the teleost-specific genome duplication (TGD). (B) The zebrafish hoxD regulatory archipelago, with candidate "autopod" enhancers shown in yellow. ATAC-seq results for 24 hpf whole-body and 48 hpf distal fin are shown, identifying areas of open chromatin. 4C-seq results on whole-body 48 and 60 hpf embryos using hoxd13a as the target are shown in green. The putative teleost ortholog of Island I shows significant interaction with the hoxd13a promoter at 60 hpf. Vista plot with zebrafish as the baseline shows no sequence conservation with mouse for autopod enhancers other than Island III and CsB.

conservation, we performed the "Assay for Transposable Accessible Chromatin" (ATAC-Seq) on 24 hour post-fertilization (hpf) zebrafish embryos, as well as e12.5 mouse autopods as a control for the assay (45, 46). ATAC-seq on mouse autopods significantly identified the majority of validated *Hox*D and *HoxA* enhancers with autopod activity (Figure 4.1 and Figure 4.2). While not all previously defined *Hox* autopod enhancers were identified (i.e. Island I, CsB) in the mouse sample, we found substantial overlap of our ATAC-seq peaks with validated limb enhancers from other studies (Table S4.1), making us confident that the assay would be able to

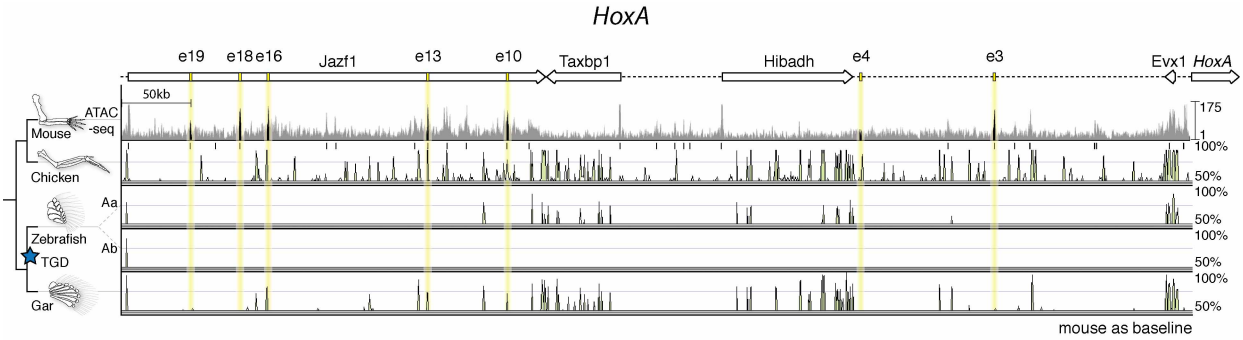


Figure 4.2. Chromatin state and sequence conservation of HoxA autopod enhancers among select vertebrates. A schematic of the HoxA **"**autopod**"** enhancer region is shown at the top, highlighting autopod-specific enhancers, with ATAC-seq data from the mouse autopod directly below. Sequence conservation for chicken, zebrafish, and gar are provided. The gar genome uniquely reveals sequence conservation of three hoxA enhancers (e16, e13, e10) identified previously in mouse.

identify enhancers in zebrafish. We performed ATAC-seq on zebrafish embryos and noticed an area of open chromatin in the genome near the *atp5g3a* gene that roughly matched the genomic coordinates of the late phase enhancer Island I (Fig. 4.1B). To determine if this region interacts with the zebrafish *hoxd13a* promoter, we performed "Circular Chromatin Conformation Capture" (4C-seq) to detect contact with enhancers up to ~1 MB away, on 48 and 60 hpf zebrafish embryos. We found that the area containing the putative ortholog of Island I in zebrafish significantly interacts with the *hoxd13a* locus at 60 hpf (Figure 4.1B). As the assay was performed on whole embryos, it is possible that these interactions may not be specific to the developing fin. We did not observe significant contact between Island I and *hoxd13a* prior to 60 hpf, possibly due to limitations on the numbers of fin cells available when using whole body embryos for 4C-seq. While the ATAC and 4C-seq data revealed genomic areas in zebrafish that could be orthologous to *HoxA* enhancers (Figure. S4.1), these areas contained multiple candidates, and thus we sought to characterize only the gar orthologs of *HoxA* where sequence orthology was clear.

Having identified potential enhancers according to chromatin structure and sequence

conservation, we sought to characterize their activity by cloning the sequences from gar and zebrafish into reporter vectors and injecting them into zebrafish embryos to assay for domains of expression (22). We performed transient injections of gar Island III, as well as the genomic "areas" from the gar genome that could potentially contain cryptic autopod enhancers (potential orthologs of mouse Islands II and IV not found by sequence alignment), and detected no fin signal for these regions (Table S4.2). In contrast, we were able to detect GFP in the pectoral fin of zebrafish embryos injected with constructs containing either gar Island I, CsB, or e16, which we raised to sexual maturity to obtain stable transgenic lines. The late phase enhancer gar Island I exhibited little activity at 31 hpf, but increased and became distally restricted by 38 hpf (Figure 4.3A and Figure S4.2). At 48 hpf, gar Island I drove a pattern of expression that was restricted to the anterior portion of the distal pectoral fin (Fig. 4.3A and Figure S4.2). Gar CsB drove a pattern of activity that was similar to that of Island I with a peak of expression at 48 hpf, but with a posterior-distal restriction to its expression. Of the three *hoxA* autopod enhancers that were identified in the gar genome by sequence conservation (Figure 4.2), e16 from mouse has been shown to drive the most robust native pattern of expression throughout the autopod of transgenic mice (19). As a result, we chose to make stable transgenic zebrafish lines of the gar e16 regulatory element, and found that it drove a distally-restricted pattern of gene expression at 48 hpf, much like the late phase *hoxD* enhancers from gar (Figure 4.3A and Figure S4.2).

Using the data from ATAC-seq, 4C-seq, and gar conservation as a guide, we cloned the putative region of Island I from zebrafish and tested its potential activity by cloning it into our reporter vector and producing stable zebrafish transgenic lines. We found that zebrafish Island I drives strong expression of GFP in the distal pectoral fin at 48 hpf, in a pattern similar to that of its gar ortholog (Figure 4.3A and Figure S4.2). To confirm that the activity of zebrafish Island I

was present in the distal fin (and not simply active elsewhere in the body), we grew multiple F1 fish from one founder, and collected a pool of distal fin cells from embryos at 48 hpf using

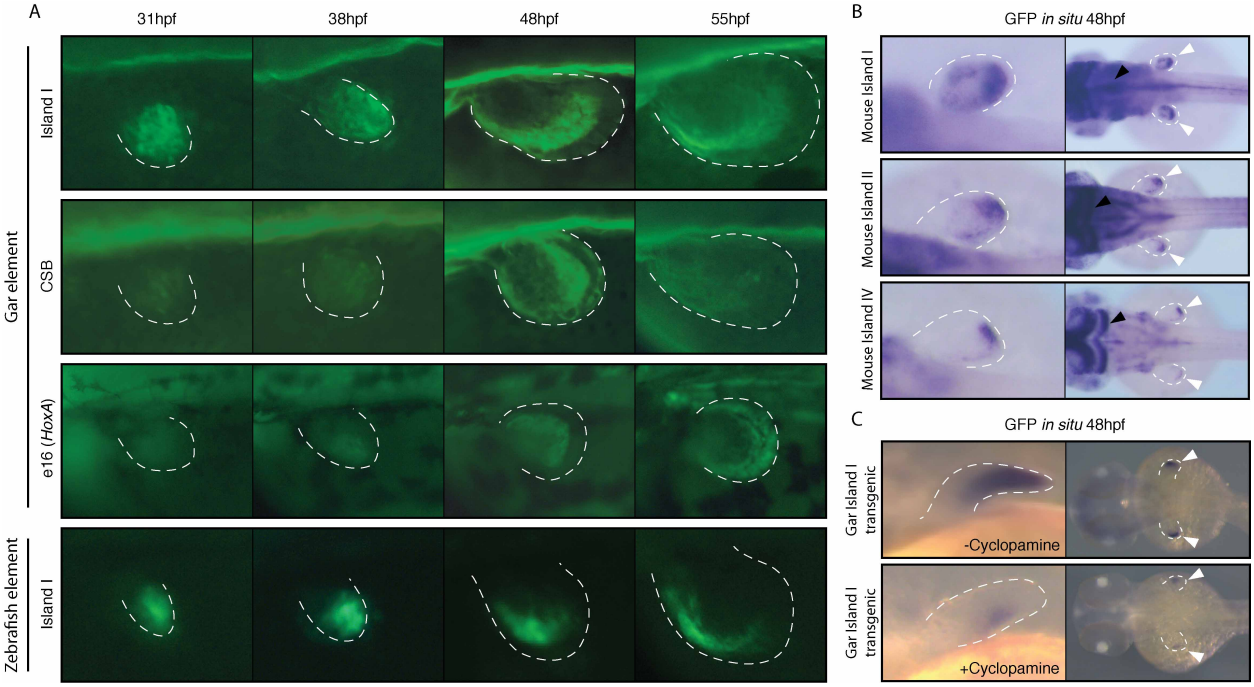


Figure 4.3. **Transgenic zebrafish reveal the expression dynamics of multiple "autopod" regulatory elements present in fish genomes**. (**A**) Left pectoral fins of stable zebrafish lines transgenic for putative late phase enhancers. The gar Island I, CsB, and e16 enhancers all drove reporter expression at 48 hpf in a strip of expression at the distal edge of the fin. Island I from zebrafish, which was identified through ATAC-seq and 4C-seq, drove a pattern of expression nearly identical to its ortholog in gar. All views are dorsal, with anterior to the left, posterior to the right, except for e16 which shows a lateral view. (**B**) The expression patterns of mouse autopod enhancers Island I, Island II, and Island IV in transgenic zebrafish. (**Left**) Dorsal views of right pectoral fins; (**Right**) dorsal views of the embryo. Island I from mouse drove a pattern of expression that was strongest at 48 hpf, with most expression at the distal compartment of the pectoral fin. Similarly, mouse Islands II and IV were active distally in the pectoral fin. Expression in the brain (black arrows) is due to a strong hindbrain enhancer present in the vector that serves as a positive control for transgenesis. (**C**) Like their murine counterparts, fish late phase enhancers depend on Shh signaling. (**Upper**) Lateral view of an in situ hybridization for the GFP transgene at 48 hpf on an embryo transgenic for the gar Island I enhancer. (**Lower**) Reduced GFP expression in a transgenic embryo that was treated with the Shh inhibitor cyclopamine from 31 to 48 hpf. Distal edge of the fin is marked by a dotted white line in all pictures.

fluorescence-activated cell sorting (FACS). We subjected these distal fin cells to ATAC-seq and found that open chromatin signal for Island I was enriched in this sample (Figure 4.1B). Combined with ATAC-seq and 4C-seq data, our transgenic zebrafish analysis indicates that zebrafish Island I functions endogenously as a distal fin enhancer acting upon the *hoxd13a* gene.

Because Island I from gar seemed to mark a distal compartment of the zebrafish pectoral fin, we sought to investigate whether Island I from mouse would elicit a similar pattern in zebrafish. To test this hypothesis, we cloned Island I from mouse into a zebrafish enhancer detection vector, and again created independent stable lines of transgenic zebrafish. Mouse Island I drove a pattern of robust expression in the distal portion of the endochondral disc at 48 hpf, marking a distal segment of the fin much like its gar ortholog (Fig. 4.3B). To further assess the activity of the other murine *HoxD* "autopod" enhancers when injected into zebrafish, we cloned mouse Islands II and IV and created stable zebrafish lines (at least three independent lines/construct). Both of these mouse enhancers drove GFP in the distal compartment of the zebrafish pectoral fin (Fig. 4.3B). These experiments demonstrate that the mouse autopod enhancer Island I, when transgenic in zebrafish, drives expression in the same pattern and area (the distal endochondral compartment of the fin) as the fish ortholog of this enhancer. Furthermore, mouse autopod enhancers that may not be present in fish genomes (Islands II and IV) are active in this same distal portion of the zebrafish pectoral fin.

The late phase of *Hox* expression in the mouse autopod and zebrafish distal pectoral fin depend on sonic hedgehog (Shh) (12, 47, 48). To test if the expression driven by the gar late phase enhancers also depends on Shh signaling, we inhibited Shh function by adding its antagonist cyclopamine (12) to transgenic zebrafish embryos from 31-48 hpf. Embryos transgenic for the gar late phase enhancer Island I that were treated with cyclopamine had markedly decreased expression of the reporter at 48 hpf, while the overall fin morphology remained normal (Figure 4.3C). We repeated these experiments on five independent transgenic lines of gar Island I, and found that nearly all embryos for all lines exhibited a loss of late phase GFP expression when treated with cyclopamine (n=21/25, 5 embryos/line). These findings show

that both mouse and gar late phase enhancers depend on *Shh* signaling, and suggest that *Shh* control of distal *Hoxd* expression is an ancestral characteristic of bony fish.

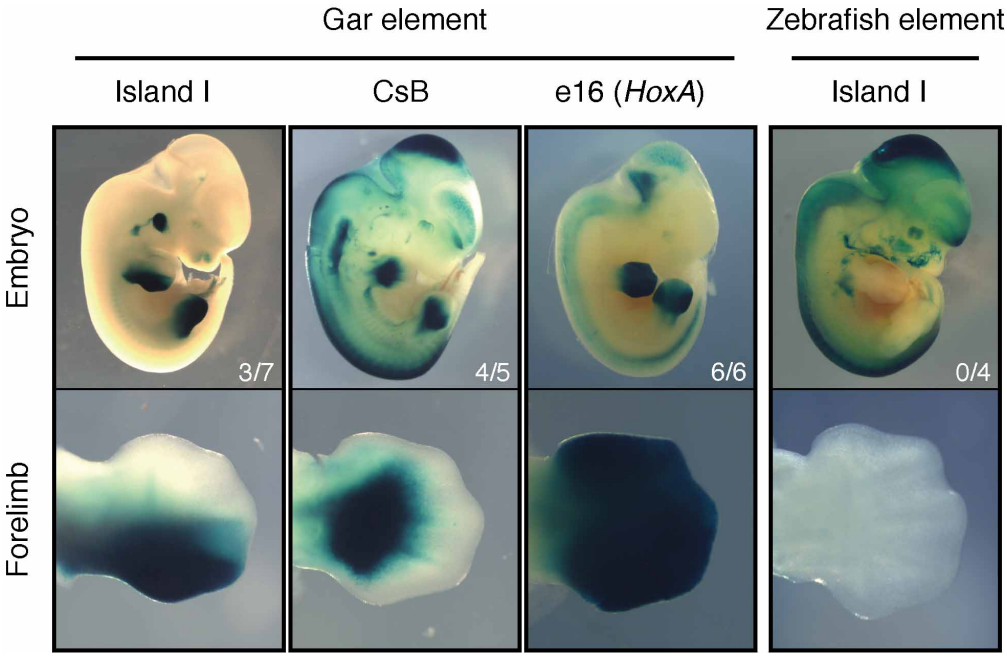


Figure. 4.4. Late phase enhancers from the nonteleost gar drive expression throughout the autopod in transgenic mice. Island I from gar drove expression in a posterior strip throughout the autopod and to the distal tip of the developing digits. Gar CsB drove strong expression in the autopod and in the presumptive digits. The hoxA late phase enhancer e16 from gar drove robust expression in the entire autopod, with a sharp boundary at the zeugopod. In accordance with previous reports, Island I from the teleost zebrafish had no activity in the autopod of transgenic mice.

Finally, to test whether the fish enhancers could respond to the trans-acting environment of developing tetrapod digits, we cloned Island I from zebrafish, and Island I, Island III, CsB, or e16 from gar, into an Hsp68-LacZ reporter vector for mouse transgenesis. In line with previous findings using teleost enhancers (43, 44), zebrafish Island I did not elicit activity in the developing mouse digits at e12.5 (Figure 4.4 and Figure S4.3). The gar Island III construct, which failed to produce fin reporter expression in zebrafish fins, also failed to elicit reporter expression in transgenic mouse limbs (Figure S4.3). However, the gar late phase enhancer

Island I drove a pattern of expression that was restricted to the posterior half of the limb, extending throughout the autopod and into the digits (Figure 4.4). In addition, the late phase element CsB from gar drove robust expression in the developing autopod and in the neural tube. Finally, we found that the e16 (*hoxA*) enhancer from gar drove strong expression in the entire autopod of transgenic mice at embryonic day 12.5, with a sharp boundary at the zeugopod (Figure 4.4). In sum, gar Island I, CsB, and e16 evoked late phase patterns of activity in developing mouse digits that were nearly indistinguishable from those driven by the orthologous elements from the mouse genome (Figure 4.4 and Figure 4.5) (18, 19, 49).

4.4 Discussion

Using a combination of model and non-model systems, functional genomics, and transgenesis, we have identified "autopod" building enhancers of the *hoxD* and *hoxA* cluster in the genomes of bony fish. We find that these enhancers drive distal expression in fish fins and respond to the same functional cue as their murine counterparts. There are likely to be three classes of "autopod" enhancers that emerge in comparisons between fish and tetrapods; those that are fish specific, tetrapod specific, and those that are shared between the two groups. Here, we have defined a potential minimum complement of enhancers that are shared between fish and tetrapods and were present in their common ancestor. Our investigations suggest that a set of enhancers are tetrapod-specific (18, 19) and may have been assembled during the fin to limb transition to elaborate the autopod (36). However, a complete investigation of the total number of late phase *Hox* enhancers in fish is necessary to determine the pattern of regulatory elements gained or lost over evolutionary time. Overall, our results provide regulatory support for an ancient origin of the "late" phase of *Hox* expression that is responsible for building the autopod.

With a largely conserved regulatory system, the question remains as to how (or if) changes to *HoxD* and *HoxA* enhancers have contributed to appendage evolution. It has been suggested that while late phase *Hox* expression is present in fish pectoral fins, a type of genetic "rewiring" of the regulatory landscape may have occurred in the transition from fins into limbs

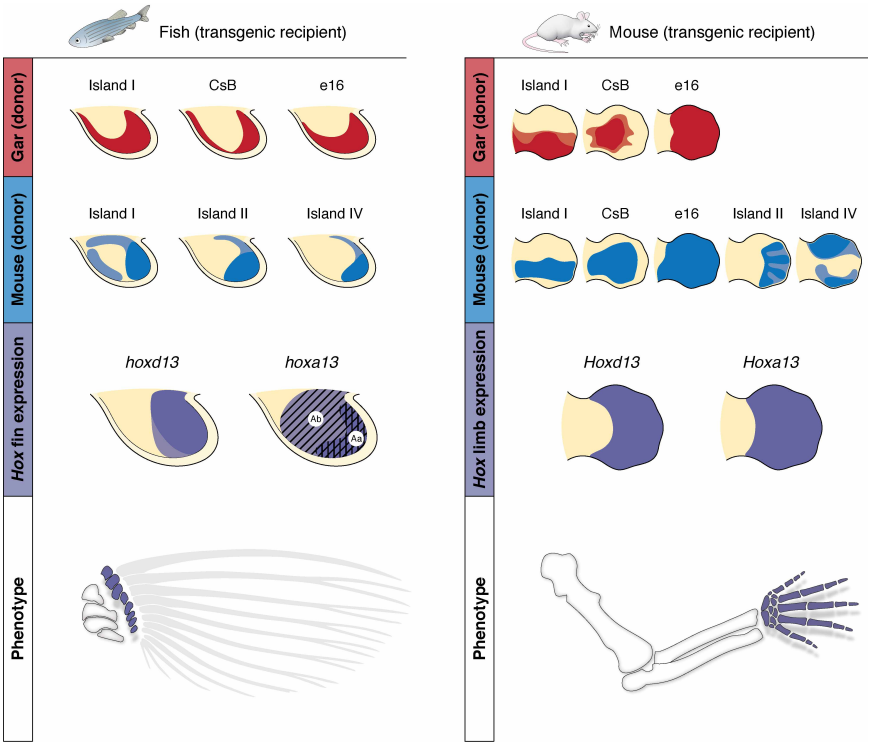


Figure 4.5. Summary and comparison of transgenic animals and their implications for the origin and evolution of the autopod. Transgenics and expression dynamics for fish (Left) and mouse (Right). Patterns driven by gar enhancers are drawn in red; mouse enhancer activity is drawn in blue. Late phase enhancers from gar drive distal fin expression in transgenic zebrafish, whereas late phase enhancers from mouse drive various patterns of autopod expression in mouse (row 1). When introduced into zebrafish, autopod enhancers from mouse (row 2, Left) drive reporter expression in the same distal compartment of the pectoral fin as the endogenous fish enhancers (compare rows 1 and 2, Left). Transgenic mice carrying late phase enhancers from gar (row 1, Right) show patterns of reporter expression throughout the autopod that are nearly identical to the endogenous activity from mouse (compare rows 1 and 2, Right). The comparative regulatory data shown here define the late phase Hox expression from fish and mouse (row 3) as homologous. Row 4 depicts a hypothesis of homology between the area of future distal radials in fish fins and the autopod of tetrapods (shown in purple), supported by the expression and chromatin conformation data reported here.

to produce digits (44). Our data using zebrafish enhancers, together with previous reports, support the hypothesis that late phase enhancers are present in teleosts and are active during fin

development, but are unable to drive expression in the developing digits of transgenic mice (Figure 4.4, Figure 4.5) (43, 44). Our results show that in contrast to teleosts, gar late phase enhancers Island I, CsB, and e16 are able to drive expression throughout the mouse autopod (including digits) in a pattern that is markedly similar to those of the mouse enhancers (Figure 4.4 and Figure 4.5) (18, 19). Taken together, these data suggest that at least a subset of regulatory elements that late phase expression of *Hox* genes was present in the last common ancestor of bony vertebrates (Osteichthyes) and have remained functionally conserved throughout evolution. Teleosts, perhaps due to genome duplication followed by subfunctionalization (50), represent a derived state that has lost the ability in some taxa to respond to a distal program in the developing limb.

Regulatory data presented here define the late phases of *Hoxd* and *Hoxa* expression in the pectoral appendages of ray-finned fish and tetrapods as homologous domains. Thus, we suggest that the distal compartment marked by the late phase in fish, which will contribute to the distal radials, is equivalent to the autopod of tetrapods (Figure 4.5). This relationship raises the intriguing possibility that digits and/or mesopodials are an ancient feature of fish fins represented by distal radials, and makes a specific prediction testable by future genomic manipulations. Our data point to a generally conserved regulatory network governing *Hox* genes in fins, emphasizing the need to study subtle modifications to *Hox* regulation as well as additional genomic regions that may influence fin morphology and that emerged during the fin to limb transition.

4.5 Materials & methods

VISTA alignments: Genomic segments of interest were downloaded from the Ensembl and UCSC genome databases and aligned using the mVista LAGAN program. The following parameters were used: calc window: 100 bps, Min Cons Width: 100 bps, Cons Identity: 70%.

Enhancer cloning:

Zebrafish and gar: Putative enhancers were amplified by PCR using primers found in supplementary table (Table S4.3), purified genomic DNA, and Platinum Taq DNA polymerase High Fidelity (Invitrogen). PCR fragments (~3 kb in length to accommodate flanking sequence) were gel purified using NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel), and subcloned into PCR8GW/GW/TOPO vector (Invitrogen) according to the manufacturers protocols. Fragments were then moved via the Gateway system (Invitrogen) into either the pXIG-cFos-eGFP vector (22) for zebrafish transgenesis, or to a Gateway-Hsp68-LacZ vector (kind gift of Marcelo Nobrega) for mouse transgenesis. Final destination vectors were confirmed by restriction digest and sequencing.

Mouse: Mouse islands were cloned in the PCR8GW/GW/TOPO vector as stated above. These enhancers were shuttled via Gateway into an enhancer detection vector composed of a *gata2* minimal promoter, an enhanced GFP reporter gene, and a strong midbrain enhancer (z48) that works as an internal control for transgenesis. All these elements were previously reported (51) and were assembled with a tol2 transposon (52).

Zebrafish Transgenesis: Zebrafish embryos (strain *AB) were collected from natural spawning. Transposase RNA was synthesized from the pCS2-zT2TP vector using the mMessage mMachine SP6 kit (Ambion) (53). Injections solutions were made following (22), and approximately 2

nanoliters were injected into the cytoplasm of embryos at the 1 or 2 cell stage. Injected embryos were raised to sexual maturity according to standard protocols (22, 53) and outcrossed to *AB fish to identify transgenic founders by visualization of GFP. Transgenic embryos from founders were visualized using a Leica M205FA microscope.

Mouse Transgenesis: Mouse transgenesis was performed by Cyagen Biosciences (Cyagen Biosciences, Santa Clara, CA). Briefly, enhancer-Hsp68-LacZ vectors were linearized with SalI, gel purified, microinjected into fertilized mouse oocytes and transferred to pseudo-pregnant females. Embryos were collected at e12.5, stained for beta-galactosidase activity, and genotyped with LacZ primers using DNA extracted from yolk sac. A summary of primers used for mouse constructs as well as injection results appear in Figure S4.3 and Table S4.3.

Cyclopamine treatment and RNA in situ *hybridization:* Cyclopamine was resuspended in 100% ethanol to make a stock of 10 mM and stored protected from light. Zebrafish embryos were dechorionated manually and placed into embryo medium containing 100 uM cyclopamine (LC Laboratories) in the dark from 31 hpf until fixation at 48 hpf. RNA *in situ* hybridization experiments targeting the GFP transcript were performed after 2 days of fixation in 4% PFA, according to standard protocols (54).

ATAC-Seq:

Mouse: ATAC-seq experiments were performed as previously described (31). Mouse autopods at e12.5 were isolated by cutting the limb at the zeugopod/autopod boundary. Three autopods were placed in PBS supplemented with 0.125 % collagenase (Sigma) and left shaking

at 300 rpms for 20 minutes. The solution was then pipetted to make a homogeneous solution, and filtered through a 0.45 μm cell strainer. 50,000 cells were then subsequently used for transposition (31).

Zebrafish: Briefly, 10 zebrafish embryos were manually dechorionated , anaesthetized with tricaine (Sigma), and disrupted in 500 μl of Ginzburg Fish Ringers (55 mM NaCl, 1.8 mM KCl, 1.25 mM NaHCO3). The cells in the resulting solution were counted using a neubauer chamber, then pelleted at 500 *g* in a tabletop centrifuge and washed with cold PBS. 75,000 cells were lysed (lysis buffer: 10 μM Tris pH7.4, 10 μM NaCl, 3 μM MgCl2, 0.1% IGEPAL) and incubated for 30 min at 37ºC with the TDE1 enzyme. The sample was then purified with Qiagen Minelute kit, and a PCR was performed with 13 cycles using Ad1F and Ad2.1R primers (45) with KAPA HiFi hotstart enzyme (Kapa Biosystems). The resulting library was sequenced in a Hiseq 2000 pair end lane producing 180M of 49 bp reads per end. Reads were aligned using zebrafish July 2010 assembly (danRer7) as the reference genome. Duplicated pairs or those ones separated by more than 2Kb were removed. The enzyme cleavage site was determined as the position -4 (minus strand) or +5 (plus strand) from each read start, and this position was extended 5 bp in both directions. ATAC-seq on FACs sorted cells were performed and analyzed in the same manner with the following modifications: approximately 400 embryos from the gar Island I transgenic line were grown to 48 hpf, anaesthetized with tricaine (Sigma), and placed in PBS supplemented with 0.125 % collagenase (Sigma). Embryos were incubated at 37^{o}C for 1 hour shaking at 300 rpm, with gentle pipetting every 20 minutes to disrupt embryos. The solution was filtered twice with a 0.45 μm cell strainer to obtain single a cell suspension, and FACs sorted for GFP using an Avalon 2-4 machine at the University of Chicago. The resulting solution contained the 40,000 brightest GFP+ cells, and was processed immediately for ATAC-

seq as described above. Statistically significant peaks were determined by extending ATAC-seq reads to 100bp and using MACS software (37) with default parameters.

4C-Seq: 4C-seq assays were performed as previously reported (38-41). 500 zebrafish embryos at 48 or 60 hpf were dechorionated using pronase and disrupted in 1 ml of Ginzburg Fish Ringers (55 mM NaCl, 1.8 mM KCl, 1.25 mM NaHCO3). Isolated cells were lysed (lysis buffer: 10 mM Tris-HCl pH 8, 10 mM NaCl, 0.3% IGEPAL CA-630 (Sigma), 1X protease inhibitor cocktail (cOmplete, Roche) and the DNA digested with DpnII (NEB) and Csp6I (Fermentas, Thermo Scientific) as primary and secondary enzymes, respectively. T4 DNA ligase (Promega) was used for both ligation steps. Specific primers were designed at gene promoters, and Illumina adaptors were included in primer sequence. 8 PCRs were performed with Expand Long Template PCR System (Roche) and pooled together. This library was purified with a High Pure PCR Product Purification Kit (Roche), its concentration measured using the Quanti-iTTM PicoGreen dsDNA Assay Kit (Invitrogen, P11496) in a Qubit machine and sent for deep sequencing in an Illumina Hiseq 2000 machine multiplexing 10 samples per lane. 4C-seq data were analysed as previously described (40). Briefly, raw sequencing data were demultiplexed using the primer sequences as barcodes and aligned using zebrafish July 2010 assembly (danRer7) as the reference genome. Reads located in fragments flanked by two restriction sites of the same enzyme, or in fragments smaller than 40 bp were filtered out. Mapped reads were then converted to reads-per-first-enzyme-fragment-end units, and smoothed using a 30 fragment mean running window algorithm. Embryos from three different stages (80% epiboly, 24 hpf, and 48 hpf) were used for each *Hox* promoter (*hoxd13a, hoxa13a, or hoxa13b*), where similar contact patterns were found when comparing stages.

4.6 Acknowledgements

We thank John Westlund for excellent illustration and assistance assembling the figures, and Marcelo Nobrega, Mike Coates, Ilya Ruvinsky, and Tom Stewart for critical reading of the manuscript. We thank Gokhan Dalgin, Noritaka Adachi, Joyce Pieretti, Darcy Ross, and Justin Lemberg for helpful discussions. We are grateful to the Prince and Ho labs for reagents and help with zebrafish. This study was supported by The Brinson Foundation to N.H.S., NSF DDIG 1311436 to A.R.G., the Brazilian National Council for Scientific and Technological Development (CNPq) grants 402754/2012-3 and 477658/2012-1 to I.S., NIH grant R01RR020833 (R01OD011116) to J.H.P: "Initiative Evolutionary Biology" from the Volkswagen Foundation, Germany and Alexander von Humboldt-Foundation to I.B., and the Spanish and Andalusian Governments grant BFU2010-14839, BFU2013-41322-P and Proyecto de Excelencia BIO-396 to J.L.G-S.

CHAPTER V

DIGITS AND FIN RAYS ARE BUILT BY HOMOLOGOUS DEVELOPMENTAL MECHANISMS

TO BE SUBMITTED TO *NATURE* AS:

Digits and fin rays are built by homologous developmental mechanisms

Tetsuya Nakamura*, Andrew R. Gehrke*, Andrew Latimer, Rebecca Thomason, Julie Szymaszek, Jonathan D. Gitlin and Neil H. Shubin†

*co-first authors
†corresponding author

The work in the chapter was designed by ARG, TN, and NHS. TN performed all CRISPR/CAS9 experiments. ARG performed all lineage-tracing experiments. AL, RT, and JDG assisted with zebrafish husbandry. JL performed *in situ* hybridization in zebrafish. The manuscript as presented in this thesis was written by ARG and NHS.

5.1 Abstract

The progressive loss of dermal fin rays and the expansion of endochondral bone is an established theory in studies of the fin to limb transition, supported by both the fossil record and developmental genetics. In this theory, the wrist and digits (autopod) of tetrapods are homologous to the distal endochondral bones of lobe-finned fish, while the dermal fin rays are composed of a separate developmental module that was completely lost over time. In mouse limbs, the autopod is built by a "late" phase of *Hox* expression: deletion of *Hoxd13* and *Hoxa13* results in the specific ablation of the autopod (55), and lineage tracing data shows that *Hoxa13* expressing cells make up the bone of the wrist and digits in adult mice (56). Recent work based on shared regulatory architecture has shown that ray-finned fish contain a homologous wave of late phase *Hox* expression, yet the role of late phase in pectoral fin development is unknown (32). To complement functional studies in mouse and provide a comparative means to assess

homology between fins and limbs, we performed knockout and lineage tracing experiments of late phase *Hox* expression in zebrafish. We find that zebrafish lacking *hoxd13* and *hoxa13a* have a severely reduced fin fold, and that the cells that receive late phase *hoxa* expression make up the skeleton of the dermal fin rays. These data suggest that the wrist and digits of tetrapods are developmentally homologous to the fin rays of fish fins, calling into question traditional morphological distinctions between fins and limbs. The evolution of the autopod may have involved a transition of distal cellular fates rather than the total loss of fin rays and acquisition of digits during the fin to limb transition.

5.2 Introduction

Tetrapod and fish appendages differ in a fundamental way--limbs contain three main segments (arm/stylopod, forearm/zeugopod, wrist and digit/autopod) that are composed entirely of endochondral bone, while fish fins contain a base of endochondral bone followed by long segments of dermal skeleton called fin rays (lepidotrichia). Traditionally, arguments concerning homology of distal elements of fins and limbs can be split into two main camps: those that see digits as an evolutionary novelty that arose in tetrapods (57, 58), and those that claim antecedents of digital structures in lobe-finned (sarcopterygian) fish (59, 60). Recent discoveries and reanalysis of the fossil record have supported the idea that lobe-finned fish (such as *Panderichthys* and *Tiktaalik*) have distal endochondral elements that are related to the autopod of tetrapods (7, 8). In contrast, comparisons between digits and the distal fin of extant ray-finned (actinopterygian) have remained enigmatic, as the fins of living fishes are composed mainly of dermal fin rays supported by a small base of endochondral bone. The apparent absence of an

autopod (based on morphology) in ray-finned fishes has led to a common classification of "fin" as "a vertebrate appendage lacking digits".

As morphological comparisons between the tetrapod limb and the fins of living fishes are difficult, developmental biologists have attempted to compare the molecular signatures of these structures to understand their origin. The lepidotrichia of fins are not preformed in cartilage, and thus have been classified as dermal bone and assumed to be neural crest in origin for decades (61, 62). A neural crest origin to fin rays suggests that the distal portion of fins represent a developmental module that is separate from that of the endochondral base, a hypothesis that fit well with the progressive loss of lepidotrichia seen in the fossil record (7). However, recent work based on lineage tracing experiments has shown that the lepidotrichia in caudal fins of zebrafish are not neural crest in origin but rather paraxial mesoderm, suggesting that fin rays are mesodermal in origin (63, 64).

Further insight into the origin of digits has come from detailed comparisons of *Hox* gene expression in fins and limbs. In the mouse forelimb, the autopod is built via late phase *Hoxd13* and *Hoxa13* expression (18), which is supported by two main lines of evidence: loss of these genes results in the complete deletion of the autopod (55), and *Hoxa13* positive cells in the developing autopod make up the skeletal cells of the wrist and digits of the adult mouse (56). Previous studies have identified the presence of homologous late phase *Hox* expression in fish, yet to date the function of late phase expression in fish is unknown (32). We reasoned that since late phase expression is homologous in fins and limbs, they are likely to build homologous structures. As late phase marks and builds the wrist and digits of tetrapod limbs, what is the function of late phase *Hox* in fish fins?

5.3 Results & discussion

We inactivated *hoxd13a* and *hoxa13a* from the zebrafish genome using the CRISPR/Cas9 system. Embryos homozygous null for both genes exhibited shorter and kinked posterior segments at 48 hours post fertilization (hpf), a defect consistent with the role of *Hox* genes in patterning the primary body axis. Surprisingly, these embryos contained pectoral fins with markedly smaller fin folds at 48 and 72 hpf, while the size of the endochondral disk remained normal (Figure 5.1).

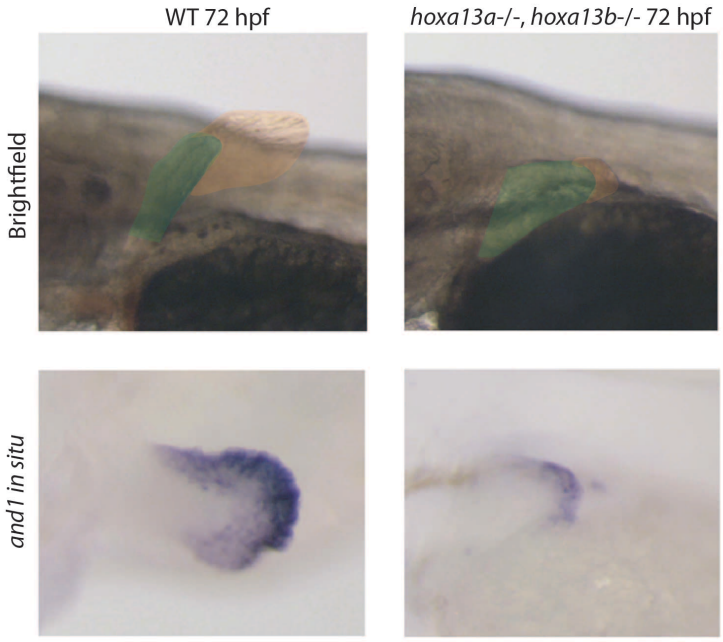


Figure 5.1. Knockout of all *hoxa13* genes in zebrafish results in a normal endochondral disc and severely reduced fin fold. (Top) *Hoxa13a* -/-:*hoxa13b* -/- zebrafish embryos have a small fin fold and WT endochondral disc at 72 hpf. (Bottom) *And1* in situs (which specifically mark the developing fin fold) are provided to clearly show the markedly different in size of the fin fold.

In order to visualize the extent of this loss, we performed *in situ* hybridization for the *and1* gene, which is necessary for lepidotrichia development and marks the developing fin fold. Staining for *and1* in 72 hpf was much weaker than in WT embryos, and was present only in the severely reduced fin fold (Figure 5.1). As deletion of *Hoxd13*/*Hoxa13* in mouse embryos results in the loss of the autopod, loss of the fin fold in *hoxd13a*/*hoxd13a* null zebrafish embryos suggests

homologous developmental mechanisms build the wrist/digits of tetrapods and the fin rays of fish.

In order to understand the fate of cells that experience both early and late phase *Hox* expression in fish fins, we performed lineage-tracing experiments in zebrafish. To trace early phase expression, we modified our previously reported transgensis vector to express CRE-recombinase driven by the zebrafish early phase enhancer CNS65 (32). This enhancer activates expression specifically in pectoral fins from 31 hpf to ~38 hpf, making it the ideal regulatory element for determining the fate of early phase *hoxd* cells. We used a 224 bp "core" of conserved sequence between fish and tetrapod genomes that was cloned in triplicate in order to drive CRE as robustly as possible, creating a final vector consisting of zebrafish CNS65x3 driving CRE (hereafter referred to as *Dr*-CNS65x3-CRE). Founder F1 fish expressing this construct were identified and crossed to *ubi*:*Switch* fish, a previously established lineage-tracing zebrafish line that permanently labels cells that express CRE with mCherry (65). At 6 days post fertilization (dpf), mesodermal cells that received early phase *hoxd* expression at 38 hpf make up the entire endochondral disk of the pectoral fin (Figure 5.2B). Surprisingly, we also found mCherry positive cells in the fin fold at 6 dpf, with filamentous protrusions facing distally and suggesting that cells are crawling from a proximal to distal direction in the developing fin (Figure 5.2B). The presence of early phase *Hox* positive cells in the developing fin fold verifies previous reports that the fold, and eventual dermal fin rays, are not neural crest in origin. These data suggest that the entire pectoral fins of zebrafish are built with a common pool of mesodermal cells from the lateral plate, and that the distal portion of these cells receive later signals to become dermal as opposed to endochondral bone. Thus, fish pectoral fins are composed of a single developmental domain that is mesodermal in origin, calling into question

the notion that the distal portion of fins (the rays) are a distinct module that aided the loss of fin rays over time during the fin to limb transition.

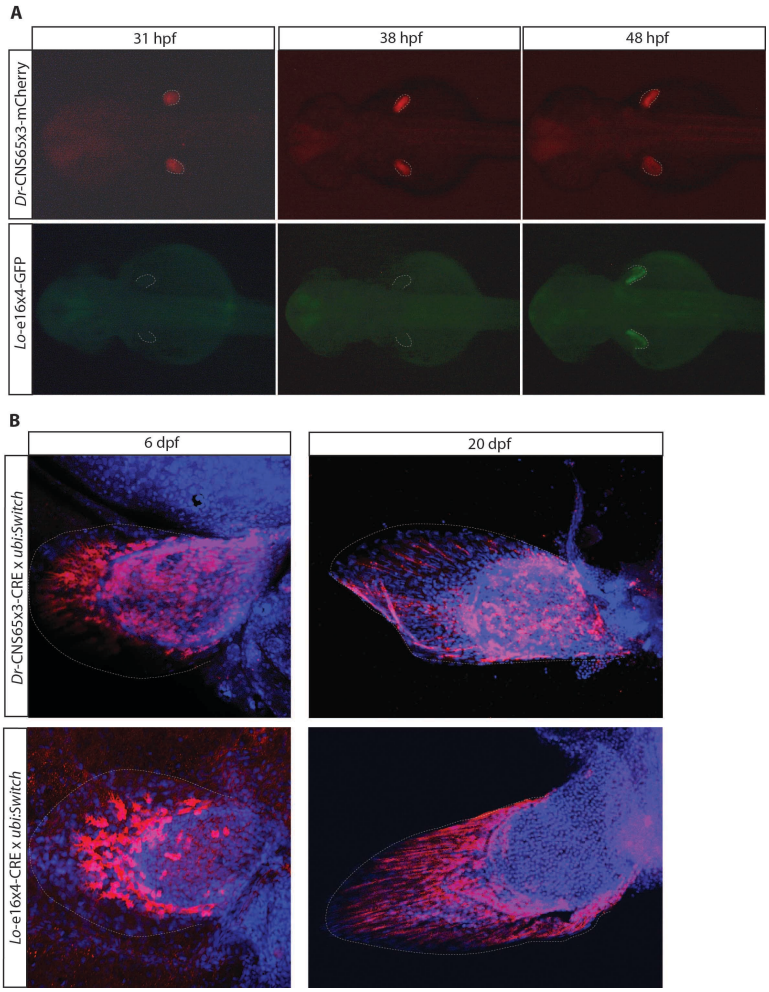


Figure 5.2. **Early phase *Hox* cells build the endochondral skeleton, while late phase *Hox* cells construct the skeleton of the dermal fin rays**. A) Expression dynamics of *Hox* enhancer constructs used for lineage tracing experiments. (Top) *Dr*-CNS65x3 drives expression (here mCherry) beginning at ~31 hpf and continues to be active in the endochondral disc until ~48 hpf, marking early phase expressing cells (Figure S5.1). (Bottom) In contrast, *Lo*-e16x4 drives activity (here eGFP) in a distal strip of cells in the endochondral disc beginning at ~48 hpf, and ending at ~72 hpf with only slight activity in the proximal fin fold, marking late phase expressing cells. B) Results of lineage tracing experiments. (Top) At 6 dpf, cells that have received early phase signals make up the future endochdondral disc, with a portion of the cells migrating distally out into the fin fold. At 20 dpf, the majority of early phase cells remain in the disc, while those that migrated to the fold have formed long lepidotrichia-like shapes. (Bottom) At 6 dpf, cells that received late phase *Hox* in the endochondral disc occupy a small portion of the distal disc, while the majority are actively migrating into the fin fold. At 20 dpf, nearly all late phase *Hox* positive cells occupy the fin fold as elongated lepidothrichia-like cells.

To determine the fate of late phase expressing cells, we employed the same strategy but using e16, a late phase *hoxa* enhancer from the spotted gar genome (32). We chose a *hoxa* enhancer because lineage-tracing data in mouse has shown that late phase *Hoxa13* cells in the limb make up the osteoblasts of the wrist and digits in the adult mouse, making it a *bona fide* marker of the autopod (56). The e16 enhancer from gar marks the distal portion of the fin of transgenic zebrafish at 48 hpf, and drives expression throughout the autopod in transgenic mice in a pattern that mimics endogenous murine *Hoxa13* expression (32). We cloned the conserved ~480 bp sequence of gar e16 in quadruplicate, again to ensure CRE was driven at sufficiently high levels (*Lo*-e16x4-CRE). When crossed to *ubi*:*Switch*, at 6 dpf we detected the majority of mCherry positive cells at the distal edge of the endochondral disk and in the developing fin fold (Figure 5.2B). As the enhancer marks a large portion of the distal endochondral disk at 48 hpf, yet we find only a small portion of lineage-traced cells in the disk at 6 dpf, late phase *hox*-positive cells are likely migrating from the endochondral portion of the fin into the fin fold. This endochondral-to-dermal domain cell movement is supported by fillipodia in mCherry positive cells projecting in the direction of the distal edge of the fin (Figure 5.2B). At 20 days post fertilization (dpf), the entire fin fold is composed of cylindrical mCherry positive cells that are likely actively forming lepidotrichia (Figure 5.2B). A portion of the distal endochondral disk is also marked by mCherry positive cells, suggesting that late phase *hoxa* cells also make up the future distal radials. As cells that receive late phase *HoxA* expression in mouse make up the wrist and digits, and in zebrafish compose the distal endochondral disk and the entirety of the fin fold, these anatomical structures are composed of a homologous cell population (Figure 5.3).

The data presented here represent the first functional tests of late phase *Hox* expression in fish fins, and have implications for understanding homology between skeletal structures between

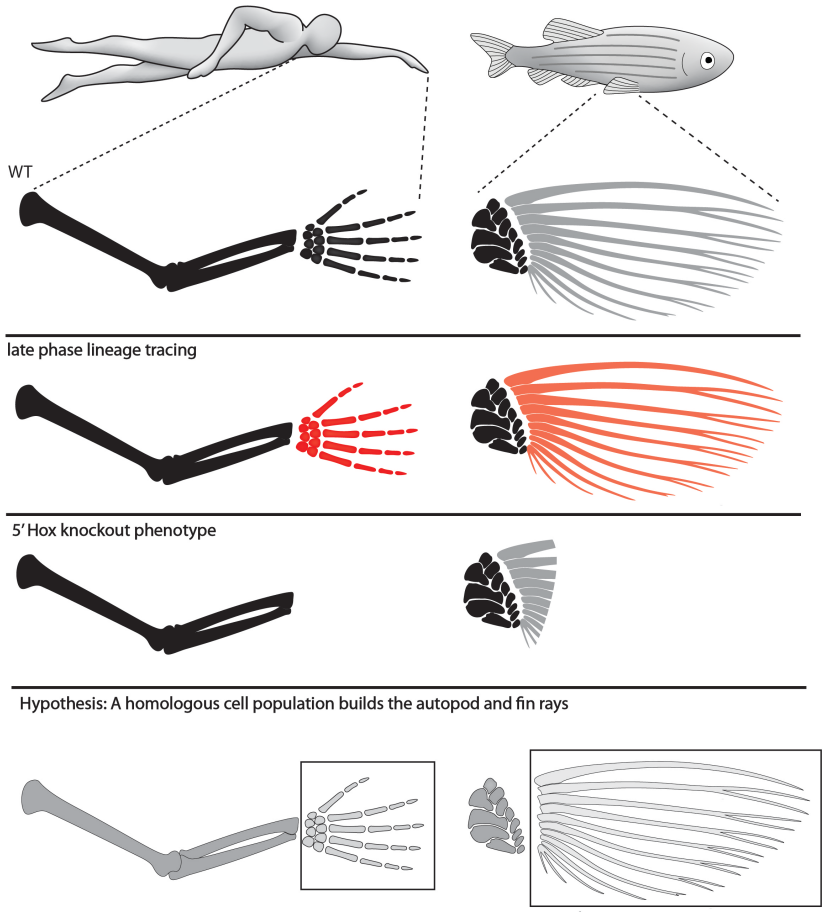


Figure 5.3. Summary and hypothesis: the digits of tetrapods and the fin rays of fish are built by homologous developmental mechanisms. In tetrapods, *Hoxa13* labels the wrist and digits, while late phase *hox* expression in fish labels fin rays. Knockouts of both *Hox13* genes in mouse results in the loss of the autopod, while in fish, *Hoxa13* knockouts result in the loss of the fin fold. Taken together, we suggest that a homologous wave of late phase Hox expression builds the autopod of tetrapods and the fin rays of fish fins.

fish and tetrapods and the fin to limb transition in general. Our findings suggest that both portions of fish fins (endochondral disk and dermal fin rays) are built using the same substrate of mesenchyme from the lateral plate mesoderm.

Thus, pectoral fins are not composed of two separate developmental modules that are free to evolve independently, e.g. a loss of distal domain combined with an expansion of

endochondral bones that has been suggested by the fossil record. Instead, our data better support a scenario where distal cells were progressively converted from a dermal fate to an endochondral one, as opposed to a simple loss of the fin fold cells over time. Finally our data show that the developmental mechanisms that build the wrist and digits of tetrapods are homologous to those that build the distal endochondral disk and fin rays of pectoral fins (Figure 5.3). These data are in contrast with previous studies that have suggested that fin rays did not play a major role in the evolution of digits, and instead point to a complex transition of the entire distal portion of fish fins in the evolution of the autopod.

5.4 Material & Methods

CRISPR-Cas9 design and synthesis. Two mutations were simultaneously introduced into the first exon of each *Hox* gene by CRISPR/Cas9 system as previously described in *Xenopus tropicalis* (66). Briefly, two gRNAs that match the sequence of exon one of each *Hox* gene were designed by *ZiFiT (http://zifit.partners.org/ZiFiT/)*. To synthesize gRNAs, forward and reverse oligonucleotides that are unique for individual target sequences were synthesized by Integrated DNA Technologies, Inc. (IDT). Each oligonucleotide sequence can be found in Table S5.1. Subsequently, each forward and reverse oligonucleotide were hybridized, and double stranded products were individually amplified by PCR with primers that include a T7 RNA promoter sequence, followed by purification by NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel). Each gRNA was synthesized from the purified PCR products by *in vitro* transcription with the MEGA script T7 Transcription kit (Ambion). Cas9 mRNA was synthesized by mMESSAGE mMACHINE SP6 Transcription Kit according to the manufacturers instructions (Ambion).

In situ hybridization. *In situ* hybridization for the CRE and *and1* genes were performed according to standard protocols (54) after fixation in 4% paraformaldehyde overnight at 4^{o} C. Primers can be found in Table S5.1.

Lineage tracing vector construction. In order to create a destination vector for lineage tracing, we first designed a random sequence of 298 bps that contained a SmaI sites to be used in downstream cloning. This sequence was ordered as a gBlocks fragment (IDT) and ligated into the pCR8/GW/TOPO TA cloning vector (Invitrogen). We then performed a Gateway LR reaction according to the manufacturers specifications between this entry vector and pXIG-cFos-GFP, which abolished an NcoI site present in the gateway cassette and introduced an SmaI site. We then removed the GFP gene with NcoI and BglII of the destination vector and ligated in CRE with (primers in Table S5.1), using the "*pCR8GW-Cre*-pA-*FRT-kan-FRT*" (kind gift of Maximiliano L. Suster, Sars International Center for Marine Molecular Biology, University of Bergen, Bergen, Norway) as a template for CRE PCR and Platinum Taq DNA polymerase High Fidelity (Invitrogen). In order to add a late phase enhancer to this vector, we first ordered four identical oligos (IDT gBlocks) of the core e16 sequence from gar, each flanked by different restriction sites. Each oligo was then ligated into pCR8/GW/TOPO, and sequentially cloned via restriction sites into a single pCR8/GW/TOPO vector. This entry vector was used a template to PCR the final *Lo*-e16x4 sequence and ligate into the CRE destination vector using XhoI and SmaI, creating *Lo*-e16x4-CRE. The early phase enhancer *Dr*-CNS65x3 was cloned into the destination vector using the same strategy. Final vectors were confirmed by sequencing. A full list of sequences and primers used can be found in Table S5.1.

Lineage tracing: All zebrafish work was performed according to standard protocols approved by The University of Chicago (ACUP #72074). *AB zebrafish embryos were collected from natural spawning and injected according to the Tol2 system as described previously (22, 32). Transposase RNA was synthesized from the pCS2-zT2TP vector using the mMessage mMachine SP6 kit (Ambion) (26). All injected embryos were raised to sexual maturity according to standard protocols (17, 26). Adult F0 fish were outcrossed to WT *AB, and the total F1 clutch was lysed and DNA isolated at 24 hpf for genotyping (see Table S5.1 for primers) to germline transmission of CRE plasmids in the F0 founders. Multiple founders were identified and tested for the strongest and most consistent expression via antibody staining and *in situ* hybridization (see below). One founder fish was identified as best, and all subsequent experiments were performed using these individual fish. Transgenic embryos from founders with fluorescent constructs were visualized using a Leica M205FA microscope.

CRISPR-Cas9: Two gRNAs targeting exon 1 of each *Hox* gene were injected with Cas9 mRNA into zebrafish eggs at the one cell stage. We injected ~2 nL of the injection solution (5 ul solution containing 500 ng of each gRNA and 500ng Cas9 diluted in nuclease free water) into the single cell of the embryo. Injected embryos were raised to adulthood, and at 3 months were genotyped by extracting DNA from tail clips. Briefly, zebrafish were anesthetized by Tricane (0.004%) and tips of the tail fin (2-3mm square) were removed and placed in an eppendorf tube. The tissue was lysed in standard lysis buffer (10 mM Tris pH 8.2, 10 mM EDTA, 200 mM NaCl, 0.5% SDS, 200 ug/mL proteinase K) and DNA recovered by ethanol precipitation. Approximately 100-500 bp of exon 1 from each gene was amplified by PCR according to primers found in Table S5.1. To determine if mutations were present, PCR products were

subjected to T7 endonuclease assay as previously reported (67), primers used can be found in Table S5.1. After identification of mutant fish by T7 assay, detailed analysis of mutation patterns were performed by sequencing at the Genomics Core at the University of Chicago.

Lineage tracing crossing and detection. Founder *Lo*-e16x4-CRE and *Dr*-CNS65x4-CRE fish were crossed to the *ubi*:*Switch* line (kind gift of Leonard I. Zon, Stem Cell Program, Children's Hospital Boston, Boston, MA). Briefly, this line contains a construct where a constituately active promoter (*ubiquitin*) drives expression of a loxP flanked GFP protein in all cells assayed of the fish. When CRE is introduced, the GFP gene is removed and the ubiquitin promoter is exposed to mCherry, thus permanently labeling the cell. We crossed our founder CRE fish to *ubi*:*Switch* and fixed progeny at different time points to track cell fate. In order to detect mCherry signal, embryos or adults were fixed overnight in 4% paraformaldehyde and subsequently processed for whole-mount antibody staining according to standard protocols (68) using the following antibodies and dilutions: 1^0 rabbit anti-mCherry/DsRed (Clontech #632496) at 1:250, 1^0 mouse anti-zns-5 (Zebrafish International Resource Center, USA) at 1:200, 2^0 goat anti-rabbit Alexa Fluor 546 (Invitrogen #A11071) at 1:400, 2^0 goat anti-mouse Alexa 647 (Invitrogen #A21235) at 1:400. Stained zebrafish were mounted under a glass slide and visualized using an LSM 710 confocal microscope.

5.5 Acknowledgements

We thank John Westlund for aid in figure construction. We thank Leonard I. Zon, Christian Mosimann, and Christian Lawrence for providing us with *ubi*:*Switch* fish, and Maximiliano L. Suster for the *pCR8GW-Cre*-pA-*FRT-kan-FRT* plasmid. We acknowledge Robert Ho and Steve

Briscoe for helpful discussions regarding lineage-tracing experiments. This study was supported by the Uehara Memorial Foundation Research Fellowship 2013, Japan Society for the Promotion of Science Postdoctoral Research Fellowship 2012-127, and Marine Biological Laboratory Research Award 2014 (to T.N.), National Institutes of Health Grant T32 HD055164 and National Science Foundation Doctoral Dissertation Improvement Grant 1311436 (to A.R.G.), and the Brinson Foundation and the University of Chicago Biological Sciences Division (N.H.S.).

CHAPTER VI:

CONCLUSIONS

6.1 Summary

In sum**,** this thesis provides data that show that fish and tetrapods share a homologous biphasic program of *Hox* gene expression to build proximal and distal elements of fins and limbs. The power of the work provided here is in its comparative regulatory approach: homology is not assumed simply by comparing expression patterns, but instead by comparing regulatory control of the loci of interest. By showing that both early and late phase *Hox* expression in fish fins are driven by the same regulatory systems as those in tetrapod limbs, we can conclude that these patterns of expression were present in their common ancestor and did not arrive through convergence.

By proving that these expression patterns are indeed homologous (based on shared regulatory architecture), these data suggest that the dermal fin rays of fish and the autopod of tetrapods are built by homologous developmental mechanisms. Through lineage tracing experiments, we show here that late phase *Hox* expression in fish is responsible for building the fin rays of zebrafish fins, where the orthologous expression pattern in tetrapods builds the wrist and digits. Thus, fin rays and digits are homologous, though the level at which this is true will be an interesting debate. I will discuss this issue, along with a number of additional conclusions that have arisen from this thesis in the following sections.

6.2 Expression patterns vs. regulatory mechanisms

As demonstrated in this thesis, comparing regulatory topology can elucidate the evolutionary history of gene expression patterns. However, it is important to ascertain whether

this strategy can be applied to other systems where organisms may be more distantly related or where comparisons of morphology are even more obscure. Not only can phylogenetic distance hinder comparisons of regulatory topology, but the inherit structure of chromatin and enhancer landscapes may bias interpretations as well.

A surprising discovery of the past 5-10 years has been that vertebrate distant regulatory elements do not necessarily "actively loop" to contact their target genes in a given cell type or tissue where the gene is active (69-72). Previous qualitative hypotheses of enhancer-promoter interaction supported a "facultative" landscape, where any given regulatory element likely remained separate and distant from the target gene when it is inactive (17). Upon binding of the appropriate TF inputs, an enhancer was thought to actively loop to contact the target gene promoter, an interaction that would become untethered in order to cease transcription. Instead, chromatin conformation experiments have shown that enhancer-promoter interactions are often "constitutive", where regulatory landscapes are locked in pre-formed loops hundreds of MB in size that are invariant across cell type and even species (73). Interestingly, these regulatory loops remain even when individual enhancers or critical binding sites are deleted (19, 74). There are numerous examples of both "facultative" (75, 76) and "constitutive" (18, 69, 77) regulatory landscapes from across the animal kingdom. How do these different modalities affect hypotheses of homology based on gene regulation?

A strictly facultative regulatory landscape is likely conducive to a comparative approach, as this type of topology is inherently less restricted across cell type and species. Thus, similar convergent expression patterns in two individuals are likely to utilize different enhancers in a facultative landscape. In contrast, constitutive TADs may require more nuanced analyses and a finer understanding of regulatory control to make meaningful evolutionary comparisons. For

instance, *HoxD* expression in mouse embryos provides an example of a constitutive landscape that could be problematic for comparison between taxa. *HoxD* genes are not expressed in the mouse brain, yet 4C-seq experiments on this tissue have revealed that a large number of the same enhancers are contacting *Hoxd* genes as in a tissue where they are active, e.g. limbs (9, 18). A similar conclusions was reached using the same methods in various *Drosophila* tissues, where genes appear to contact the same distant non-coding space regardless if the gene is active or not (69). This is problematic for a comparative approach, as chromatin contact experiments between species will likely yield the same profile may consistently bias the conclusion towards conservation rather than convergence. However, recent data has shown that while the overall landscapes may be conserved between species, subtle differences in sub-TADs may account for differences in expression (19, 70). It will be imperative for evolutionary studies to understand the dynamics of regulatory elements at the level of the enhancer (or preferably the nucleotide) when comparing distantly related species in order to avoid claims of conservation based on globally similar epigenetic profiles that are merely byproducts of TADs.

6.3 The utility of non-model organisms and the requirement of phylogenetic breadth

A common methodology of evo devo is to compare development in two taxa (i.e. zebrafish and mouse) in order to make an inference about their extinct past – if a feature is shared between the two, it was likely present in their common ancestor. Left unchecked, this mindset can lead to sweepingly broad statements concerning the evolution of morphology, e.g. the eyes or appendages of all protostomes and deuterostomes are homologous. Developmental biology is at an increased risk of insufficiently broad statements due to restrictions inherit to the

use of model organisms. For example, evo devo biologists must be constantly aware that mice do not represent all tetrapods, nor do zebrafish represent all fish.

This thesis represents an acute example of the importance of developing non-model systems and the use of as wide a phylogenetic breadth as possible in evolutionary biology. When comparing genomic sequence between mouse and teleosts, we were unable to identify early nor late phase enhancers by sequence conservation. Even when testing a variety of teleosts (zebrafish, pufferfish, medaka, stickleback), we failed to detect sequence orthology in every case. Only when introducing the spotted gar were we able uncover sequence conservation that remained hidden in teleost genomes. An important conceptual note here resides in the choice of the spotted gar as a model system. As all teleosts have undergone genome duplication, the additional sequence of any individual teleost would have been unlikely to unlock this cryptic orthology. The spotted gar is one of only a few living species of ray-finned fish fish (other examples include bowfin, polypterus, paddlefish) that are not teleosts and thus have an unduplicated genome (78). We hypothesize an unduplicated genome was crucial to our ability to detect *Hox* enhancer sequence orthology with mouse, and is a prime example of the utility of developing non-model organisms. Without the sequence of the spotted gar, we likely would have committed a type II error of a false negative by claiming that ray-finned fish do not contain orthologs of murine *Hox* enhancers.

A second example of the necessity of non-model organisms and expansive phylogenetic sampling lies in the use of the elephant shark genome in our search for ancient early phase *Hox* enhancers. We were unable to detect the second early phase *HoxD* enhancer, CNS39, even when including the sequence of the gar. In this case, we were able to detect the regulatory element by sequence conservation only when we included the genome of the elephant shark. Given the

previous success using gar as a stepping-stone to uncover hidden orthology in other genomes, it was tempting to declare that CNS39 did not exist in fish genomes and was thus a tetrapod synapomorphy. Only by considering a wider taxonomic breadth were we able to uncover CNS39, and thus leading us to a more precise understanding of the state of the early phase regulatory landscape in gnathostomes. The utility of both the gar and elephant shark genomes in this thesis are a testament to those considering sequencing future non-model systems, in that evo devo biologists (interested in the evolution of form) should sequence systems that are likely to be as phylogenetically informative as possible.

6.4 Developmental systems drift and inter-species transgenesis

The proper activation of an enhancer relies on the interaction between *cis* (the DNA binding site), and *trans* (the TFs that bind it). Decades of study in the evolution of gene regulation have shown that both the sequence of the enhancer and the properties of the TFs evolve, though recent studies show greater instances of changes in *cis* (34, 79). The standard and most widely used assay to determine the function of an enhancer is transgenesis, where the regulatory element of interest is placed upstream of minimal promoter and a reporter gene and injected to make a transgenic host organism. Interpreting transgenics in a native organism is more straightforward (i.e. an element from the mouse genome tested in a transgenic mouse), as the *cis* element is tested in its endogenous *trans* environment. However, evolutionary developmental biologists, who are by definition interested in a comparative approach, are faced with a more challenging interpretation of inter-species transgenesis. Biologists are limited to model systems that are amenable to the assay of interest, and reliable transgenesis is a technique

that often takes years to perfect in emerging systems. The result is a nearly limitless amount of donor sequences with only a few host species with which to test them. Thus, it is often impossible to test regulatory elements in their natural host, and the necessary evil of inter-species transgenesis must be relied upon. When testing a regulatory element in any organism other than the original host, careful consideration must be taken in what can be interpreted from the results.

Transgenic enhancer experiments involve two variables: donor (the *cis* DNA element), and host (the organism that will receive the *cis* element). As mentioned previously, when donor DNA and the host organism match (i.e. mouse in mouse), reproducible patterns of activity can be attributed solely to the regulatory element tested. When the donor and host species do not match (a "swap"), interpretations must take into account potential evolution in *cis*, *trans,* or both (34). The number and type of conclusions that can be made from a "swap" experiment have been reviewed in detail elsewhere (34). The most relevant type of swap for this discussion is one where the activity of the donor *cis* element in its native *trans* environment is either unknown or impossible to ascertain (often the case for non-model systems). Without knowledge of the enhancer capability in its native *trans* environment, no negative conclusion can be made concerning the activity of that element in another host organism. Foreign CREs may fail to drive expression in a transgenic host of a different species for a number of reasons that are irrelevant to the evolution of phenotype. In such "non-controlled" swap experiments, evolution in *trans* cannot be ruled out, and thus causality or direction of evolution cannot be inferred (34). The ability of underlying genetic systems of homologous characters to change in function is called "Developmental Systems Drift" (DSD), and is a phenomenon involving *cis* and *trans* coevolution that always must be considered when interpreting inter-species transgenic experiments (34, 80-82).

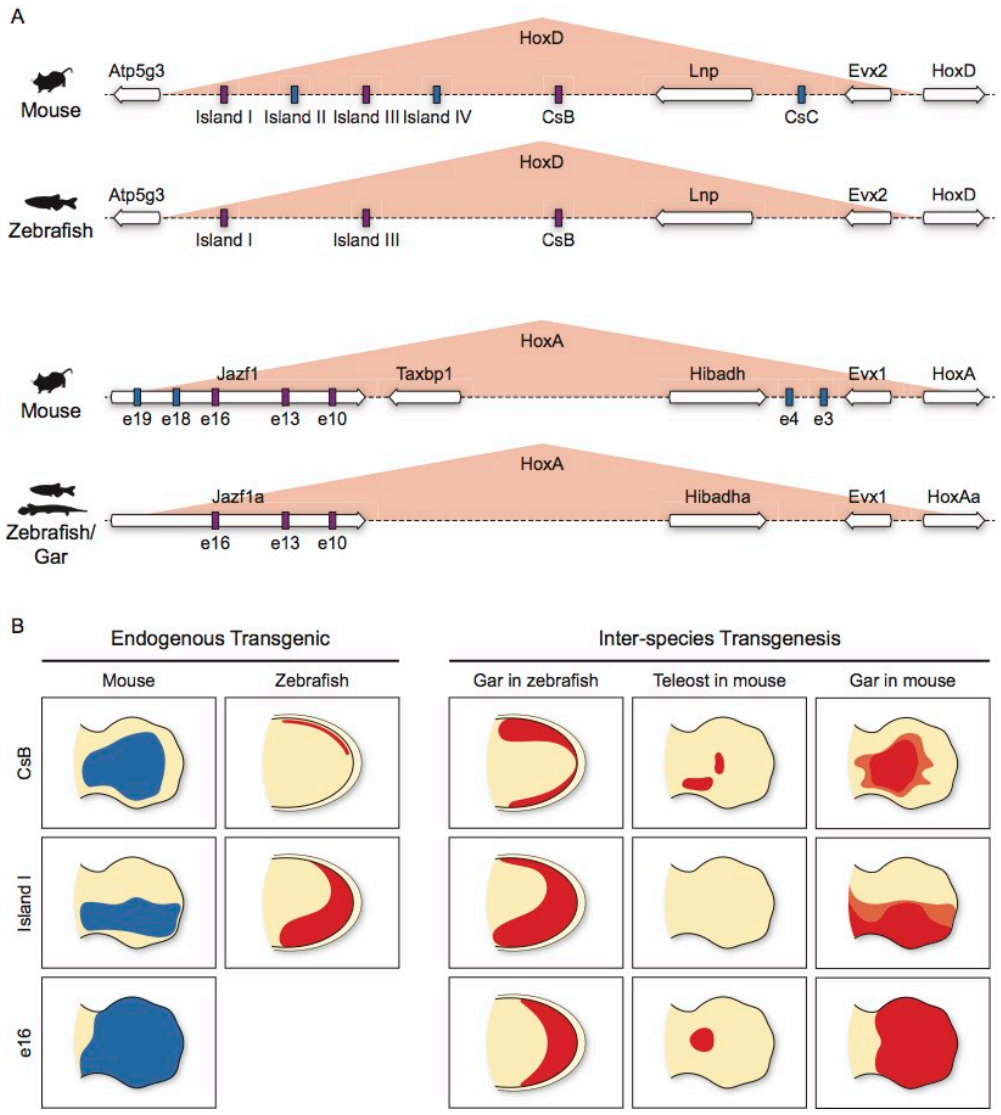


Figure 6.1. The evolutionary history of *Hox* gene regulation in distal appendages. A) Digit expression of *Hoxd13* and *Hoxa13* in mouse is driven by a series of enhancers that lie specifically on one side of the *Hox* cluter, spread over a distance of ~500 kb. These regulatory elements fall with topological associated domains (shown as red triangles) that define the regulatory landscape of *Hox* expression. Studies in zebrafish have shown that the same topological domains are present in fish, along with orthologs of specific mouse digit enhancers. Potential mouse-specific enhancers are shown in blue, those that are shared with fish are colored purple. For the zebrafish/gar *HoxA* cluster, the topological domain was found in zebrafish, while the orthologs of the enhancers represented are from the gar genome. B) Summary of transgenic experiments with *Hox* digit enhancers. *Cis*-elements from mouse are denoted in blue, those from fish in red.

When we consider the *Hox* enhancer swaps presented in this thesis, the manner in which inter-species transgenic experiments are interpreted becomes crucial for making statements about

the evolutionary history of the limb (Figure 6.1). When assayed in transgenic mice, *hox* enhancers show a variety of different expression patterns depending on the phylogenetic position of the fish (32, 44) (Figure 6.1). While these regulatory elements have differing activities in the context of the developing mouse, all enhancers tested from fish are active in the pectoral fin of transgenic zebrafish (32, 43). Thus, phylogenetic distance between the donor and host organism is a factor that cannot be ignored, as only when *hox* enhancers from some fish are placed in the context of a different *trans* environment (mouse) does the function drift.

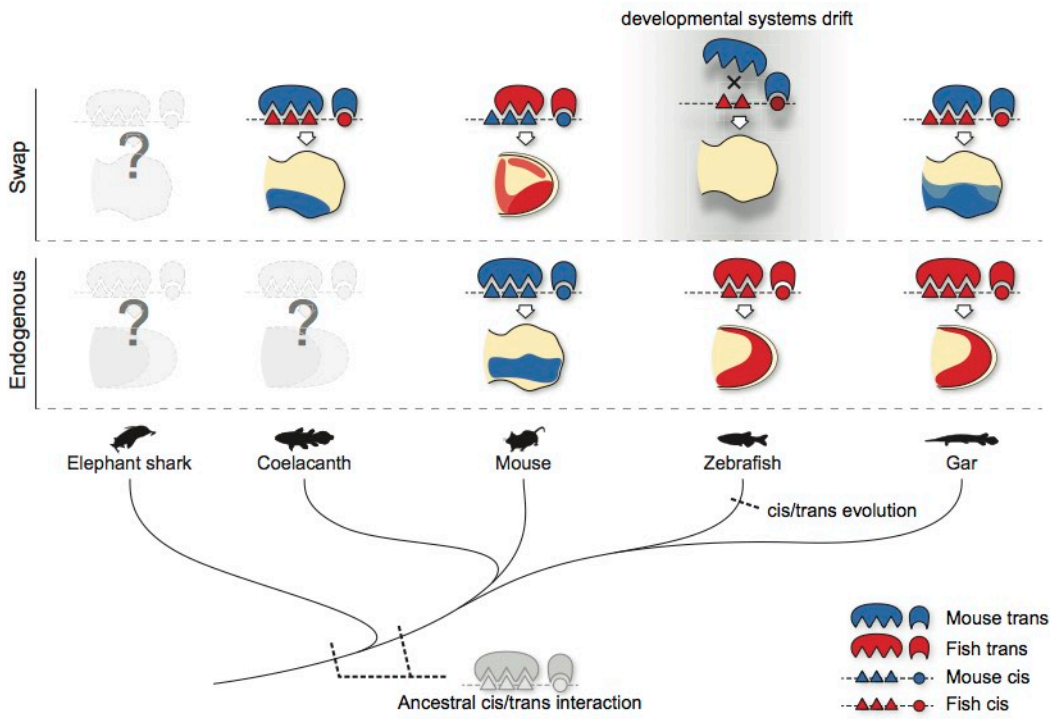


Figure 6.2. Interpreting inter-species transgenics in a phylogenetic context. A potential evolutionary scenario for the *cis*/*trans* interaction of the Island I *Hox* enhancer based on reciprocal transgenics. Here, zebrafish likely underwent evolution in both *cis* and *trans* (developmental systems drift), making the host *trans* environment of mouse incapable of interpreting the *cis* element of zebrafish.

The inability of some fish enhancers to drive expression in the autopods of transgenic mice is not due an intrinsic lack of function in these CREs, but developmental systems drift between donor and host species (Figure 6.2). Fish and tetrapods contain orthologous enhancers that may

underlie homologous structures, yet the *trans* system of the host cannot always properly decode donor *cis* in a "swap" due to local evolution in each species (Figure 6.2).

6.5 Homology and the fin to limb Transition

Chapters 2-4 of this thesis are concerned with providing regulatory support for homology between early and late phase *Hox* expression patterns in fish and tetrapods. These data comprise the necessary underpinnings for chapter 5, where we investigate the actual function of these homologous expression domains during development. We find that late phase *Hox* gene expression is responsible for building the dermal fin rays (and likely distal radials) of fish fins, while the orthologous gene activity in mouse builds the wrist and digits. With these data in mind, what exactly is homologous between fish fins and tetrapod limbs, and what do these interpretations mean for the fin to limb transition?

As mentioned in the introductory chapter, it is important to specify "levels" when considering homology (4, 5). On a coarse level based on these data, one might be tempted to designate the fin rays of ray-finned fishes as homologous to the tetrapod autopod. However, there are fundamental differences in the development and morphology of these structures that do not support the idea that they are homologous. The most obvious difference is that the fin rays of fish are composed of dermal bone, while tetrapod limbs are composed entirely of endochondral bone (21). While the data presented here show that both types of bone in paired appendages are derived from the lateral plate mesoderm, the two cell populations encounter different genetic programs in their development to instruct them to become either endochondral or dermal. A second counterpart to homology of these elements lies in the fossil record, which

shows a general reduction of the distal skeleton (7, 21). For instance, there are no fossils that show a transitional form where the dermal fin rays are fused/expanded to resemble digits. Instead, the fin rays appear to be gradually reduced until they are completely lost in the tetrapod lineage. Thus, there appears to be a dichotomy between data from development (homology) and morphology/fossils (non-homology) when considering the fin rays of actinopterygians. How can these be reconciled?

We propose that that wrist/digits of tetrapods and the fin rays of fish are built by homologous populations of cells during development. Thus, fin rays and digits are not homologous as structures, by are instead both built using a homologous group of mesenchymal cells from the lateral plate mesoderm that receive different downstream cues to become either endochondral or dermal bone. This interpretation has important implications for the fin to limb transition. Our data contrasts with the generally accepted paradigm for the origin of digits where fins rays are progressively lost and distal radials are either modified or digits are progressively added *de novo*. Instead, our data support a previously suggested model where the "loss" of fin rays creates a set of "spare" cells that provide fodder for the elaboration of distal endoskeleton (12). The data presented in this thesis (particularly Chapter 5) support the hypothesis by Ahn and Ho that fin rays and digits are comprised of the same population of mesodermal cells (12). We extend this hypothesis by showing here that these skeletogenic cells are comprised of cells that received late phase *HoxD* and *HoxA* expression in the endochondral disc and migrate into the fin fold to eventually build the rays.

In light of generally conserved developmental mechanisms, it has been suggested that changes in *Hox* gene expression likely did not play a causative role in the fin to limb transition (12, 32). The data in this thesis generally support this conclusion by providing regulatory

evidence for deeply conserved patterns of expression (e.g. there was likely no significant gain of an entire expression domain to produce digits, *contra* (83, 84)). However, we suggest that there is evidence for a gain of function scenario where late phase *Hox* expression was modified or "boosted" to play role in the acquisition/elaboration of digits. Zebrafish fins that have been exposed to an artificial overexpression of *Hoxd13a* show larger pectoral endochondral discs and reduced fin folds, showing that augmented late phase *Hox* expression can at least in principle produce more robust bone (36). We suggest a scenario where an increased amount of late phase *Hox* expression, potentially by the addition of enhancers in both the *HoxD* and *HoxA* clusters, caused an expansion of the endochondral program into the dermal domain of fish fins. In this model, a gradual "growth" of the endochondral disc was fed by cells of the fin fold, causing the reduction of fin rays found in the fossil record (*sensu* (12)). It will be interesting (in this model) to determine if the expansion of the disc was accompanied by an elaboration of the distal radials, or if "new" endochondral bone is produced, in order to determine if digits are indeed "novel" structures. Future studies that examine the precise contribution of late phase *Hox* expression to radials should aid to elucidate this question.

6.6 Future Directions: a phylogenetic perspective on functional genomics

This thesis can be classified as a study in Evolutionary Developmental Functional Genomics. This study uses regulatory topology as groundwork for understanding the evolutionary history of *Hox* gene expression and the implications for the fin to limb transition. This differs from more traditional evo devo studies that compare patterns of expression in

disparate embryos. If expression patterns are insufficient or problematic for inferring evolutionary history (15) (traditional evo devo), what is the future of evo devo functional geno?

The development of high-throughput, whole-genome technologies that require a relatively small (or even single) input of cells has a potentially transformative impact on evolutionary studies (85-90). Developmental biologists interested in the evolution of phenotype have the potential to complement the work done in model species, such as chicks and mice, with work on an ever-increasing diversity of species that was previously impossible due to technical limitations of bulk tissue needed. Armed with these tools, virtually any species has the potential to become a "model" of sorts. By being able to select taxa with relevant phenotypes, phylogenetic histories, and genomic structures, biologists can bring increasing granularity to comparisons, thereby allowing ever more refined tests of hypotheses of the genetic basis of morphological transformations. Finally, a greater understanding of regulatory landscapes across taxa will identify meaningful targets for direct genomic manipulation in order to test hypotheses of evolutionary scenarios (91). Functional genomics, performed in the context of as wide a phylogenetic breadth as possible, will provide valuable insight for the future of evo devo

REFERENCES

1. Carroll SB (2008) Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* 134(1):25-36.
2. Panganiban G*, et al.* (1997) The origin and evolution of animal appendages. *Proc Natl Acad Sci U S A* 94(10):5162-5166.
3. Rincon-Limas DE*, et al.* (1999) Conservation of the expression and function of apterous orthologs in Drosophila and mammals. *Proc Natl Acad Sci U S A* 96(5):2165-2170.
4. Shubin N, Tabin C, & Carroll S (2009) Deep homology and the origins of evolutionary novelty. *Nature* 457(7231):818-823.
5. Wray GA & Abouheif E (1998) When is homology not homology? *Curr Opin Genet Dev* 8(6):675-680.
6. Wagner GnP (2014) *Homology, genes, and evolutionary innovation* (Princeton University Press, Princeton ; Oxford) pp xiii, 478 pages.
7. Shubin NH, Daeschler EB, & Jenkins FA, Jr. (2006) The pectoral fin of Tiktaalik roseae and the origin of the tetrapod limb. *Nature* 440(7085):764-771.
8. Boisvert CA, Mark-Kurik E, & Ahlberg PE (2008) The pectoral fin of Panderichthys and the origin of digits. *Nature* 456(7222):636-638.
9. Andrey G*, et al.* (2013) A switch between topological domains underlies HoxD genes collinearity in mouse limbs. *Science* 340(6137):1234167.
10. Zakany J & Duboule D (2007) The role of Hox genes during vertebrate limb development. *Curr Opin Genet Dev* 17(4):359-366.
11. Nelson CE*, et al.* (1996) Analysis of Hox gene expression in the chick limb bud. *Development* 122(5):1449-1466.
12. Ahn D & Ho RK (2008) Tri-phasic expression of posterior Hox genes during development of pectoral fins in zebrafish: implications for the evolution of vertebrate paired appendages. *Dev Biol* 322(1):220-233.
13. Davis MC, Dahn RD, & Shubin NH (2007) An autopodial-like pattern of Hox expression in the fins of a basal actinopterygian fish. *Nature* 447(7143):473-476.
14. Freitas R, Zhang G, & Cohn MJ (2007) Biphasic Hoxd gene expression in shark paired fins reveals an ancient origin of the distal limb domain. *PLoS One* 2(8):e754.
15. Woltering JM & Duboule D (2010) The Origin of Digits: Expression Patterns versus Regulatory Mechanisms. *Dev Cell* 18(4):526-532.
16. Wray GA (2007) The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet* 8(3):206-216.
17. Wray GA*, et al.* (2003) The evolution of transcriptional regulation in eukaryotes. *Mol Biol Evol* 20(9):1377-1419.
18. Montavon T*, et al.* (2011) A regulatory archipelago controls Hox genes transcription in digits. *Cell* 147(5):1132-1145.
19. Berlivet S*, et al.* (2013) Clustering of tissue-specific sub-TADs accompanies the regulation of HoxA genes in developing limbs. *Plos Genet* 9(12):e1004018.
20. Zakany J & Duboule D (2007) The role of Hox genes during vertebrate limb development. *Curr Opin Genet Dev* 17(4):359-366.

21. Schneider I & Shubin NH (2013) The origin of the tetrapod limb: from expeditions to enhancers. *Trends Genet* 29(7):419-426.
22. Fisher S*, et al.* (2006) Evaluating the biological relevance of putative enhancers using Tol2 transposon-mediated transgenesis in zebrafish. *Nat Protoc* 1(3):1297-1305.
23. Braasch I*, et al.* (2014) A new model army: Emerging fish models to study the genomics of vertebrate Evo-Devo. *Journal of experimental zoology. Part B, Molecular and developmental evolution*.
24. Amores A*, et al.* (1998) Zebrafish hox clusters and vertebrate genome evolution. *Science* 282(5394):1711-1714.
25. Amores A, Catchen J, Ferrara A, Fontenot Q, & Postlethwait JH (2011) Genome evolution and meiotic maps by massively parallel DNA sequencing: spotted gar, an outgroup for the teleost genome duplication. *Genetics* 188(4):799-808.
26. Taylor JS, Braasch I, Frickey T, Meyer A, & Van de Peer Y (2003) Genome duplication, a trait shared by 22000 species of ray-finned fish. *Genome Res* 13(3):382-390.
27. Kimmel CB, Ballard WW, Kimmel SR, Ullmann B, & Schilling TF (1995) Stages of embryonic development of the zebrafish. *Dev Dyn* 203(3):253-310.
28. Suster ML, Abe G, Schouw A, & Kawakami K (2011) Transposon-mediated BAC transgenesis in zebrafish. *Nat Protoc* 6(12):1998-2021.
29. Braasch I*, et al.* (2016) The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. *Nat Genet*.
30. Venkatesh B*, et al.* (2014) Elephant shark genome provides unique insights into gnathostome evolution. *Nature* 505(7482):174-179.
31. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, & Greenleaf WJ (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* 10(12):1213-1218.
32. Gehrke AR*, et al.* (2015) Deep conservation of wrist and digit enhancers in fish. *Proc Natl Acad Sci U S A* 112(3):803-808.
33. Didier DA, Leclair EE, & Vanbuskirk DR (1998) Embryonic Staging and External Features of Development of the Chimaeroid Fish, Callorhinchus milii (Holocephali, Callorhinchidae). *J Morphol* 236:25-47.
34. Gordon KL & Ruvinsky I (2012) Tempo and mode in evolution of transcriptional regulation. *Plos Genet* 8(1):e1002432.
35. Gehrke AR & Shubin NH (2016) Cis-regulatory programs in the development and evolution of vertebrate paired appendages. *Seminars in cell & developmental biology*.
36. Freitas R, Gomez-Marin C, Wilson JM, Casares F, & Gomez-Skarmeta JL (2012) Hoxd13 contribution to the evolution of vertebrate appendages. *Dev Cell* 23(6):1219-1229.
37. Zhang Y*, et al.* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome biology* 9(9):R137.
38. Dekker J, Rippe K, Dekker M, & Kleckner N (2002) Capturing chromosome conformation. *Science* 295(5558):1306-1311.
39. Hagege H*, et al.* (2007) Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat Protoc* 2(7):1722-1733.
40. Noordermeer D*, et al.* (2011) The dynamic architecture of Hox gene clusters. *Science* 334(6053):222-225.

41. Splinter E, de Wit E, van de Werken HJ, Klous P, & de Laat W (2012) Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation. *Methods* 58(3):221-230.
42. Montavon T & Duboule D (2013) Chromatin organization and global regulation of Hox gene clusters. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 368(1620):20120367.
43. Schneider I*, et al.* (2011) Appendage expression driven by the Hoxd Global Control Region is an ancient gnathostome feature. *Proc Natl Acad Sci U S A* 108(31):12782-12786.
44. Woltering JM, Noordermeer D, Leleu M, & Duboule D (2014) Conservation and divergence of regulatory strategies at hox Loci and the origin of tetrapod digits. *PLoS Biol* 12(1):e1001773.
45. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, & Greenleaf WJ (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* 10(12):1213-+.
46. Lara-Astiaso D*, et al.* (2014) Immunogenetics. Chromatin state dynamics during blood formation. *Science* 345(6199):943-949.
47. Chiang C*, et al.* (2001) Manifestation of the limb prepattern: limb development in the absence of sonic hedgehog function. *Dev Biol* 236(2):421-435.
48. Litingtung Y, Dahn RD, Li Y, Fallon JF, & Chiang C (2002) Shh and Gli3 are dispensable for limb skeleton formation but regulate digit number and identity. *Nature* 418(6901):979-983.
49. Gonzalez F, Duboule D, & Spitz F (2007) Transgenic analysis of Hoxd gene regulation during digit development. *Dev Biol* 306(2):847-859.
50. Force A*, et al.* (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151(4):1531-1545.
51. Bessa J*, et al.* (2009) Zebrafish enhancer detection (ZED) vector: a new tool to facilitate transgenesis and the functional analysis of cis-regulatory regions in zebrafish. *Dev Dyn* 238(9):2409-2417.
52. Kawakami K, Shima A, & Kawakami N (2000) Identification of a functional transposase of the Tol2 element, an Ac-like element from the Japanese medaka fish, and its transposition in the zebrafish germ lineage. *Proc Natl Acad Sci U S A* 97(21):11403-11408.
53. Suster ML, Abe G, Schouw A, & Kawakami K (2011) Transposon-mediated BAC transgenesis in zebrafish. *Nat Protoc* 6(12):1998-2021.
54. Thisse C, Thisse B, & Postlethwait JH (1995) Expression of snail2, a second member of the zebrafish snail family, in cephalic mesendoderm and presumptive neural crest of wild-type and spadetail mutant embryos. *Dev Biol* 172(1):86-99.
55. Fromental-Ramain C*, et al.* (1996) Hoxa-13 and Hoxd-13 play a crucial role in the patterning of the limb autopod. *Development* 122(10):2997-3011.
56. Scotti M, Kherdjemil Y, Roux M, & Kmita M (2015) A Hoxa13:Cre mouse strain for conditional gene manipulation in developing limb, hindgut, and urogenital system. *Genesis* 53(6):366-376.
57. Holmgren N (1933) On the origin of the tetrapod limb. *Acta Zoologica* 14(2-3):185–295.

58. Holmgren N (1939) Contribution on the question of the origin of the tetrapod limb. *Acta Zoologica* 20(1):89–124.
59. Watson DMS (1913) On the primitive tetrapod limb. *Anat. Anzeiger* 44:24–27
60. Gregory WKR, H.C. (1941) Studies on the origin and early evolution of paired fins and limbs. *Ann. N.Y. Acad. Sci.* 42:273-360.
61. Smith M, Hickman, A., Amanze, D., Lumsden, A., and Thorogood, P. (1994) Trunk neural crest origin of caudal fin mesenchyme in the zebrafish Brachydanio rerio. *Proc. R. Soc. Lond. B Biol. Sci.* 256:137-145.
62. Smith MM & Hall BK (1990) Development and evolutionary origins of vertebrate skeletogenic and odontogenic tissues. *Biol Rev Camb Philos Soc* 65(3):277-373.
63. Lee RT, Thiery JP, & Carney TJ (2013) Dermal fin rays and scales derive from mesoderm, not neural crest. *Curr Biol* 23(9):R336-337.
64. Lee RT, Knapik EW, Thiery JP, & Carney TJ (2013) An exclusively mesodermal origin of fin mesenchyme demonstrates that zebrafish trunk neural crest does not generate ectomesenchyme. *Development* 140(14):2923-2932.
65. Mosimann C*, et al.* (2011) Ubiquitous transgene expression and Cre-based recombination driven by the ubiquitin promoter in zebrafish. *Development* 138(1):169-177.
66. Nakayama T*, et al.* (2013) Simple and efficient CRISPR/Cas9-mediated targeted mutagenesis in Xenopus tropicalis. *Genesis* 51(12):835-843.
67. Cho SW, Kim S, Kim JM, & Kim JS (2013) Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nature biotechnology* 31(3):230-232.
68. Asharani PV*, et al.* (2012) Attenuated BMP1 function compromises osteogenesis, leading to bone fragility in humans and zebrafish. *American journal of human genetics* 90(4):661-674.
69. Ghavi-Helm Y*, et al.* (2014) Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* 512(7512):96-100.
70. Lonfat N, Montavon T, Darbellay F, Gitto S, & Duboule D (2014) Convergent evolution of complex regulatory landscapes and pleiotropy at Hox loci. *Science* 346(6212):1004-1006.
71. Pope BD*, et al.* (2014) Topologically associating domains are stable units of replication-timing regulation. *Nature* 515(7527):402-405.
72. Remeseiro S, Hornblad A, & Spitz F (2015) Gene regulation during development in the light of topologically associating domains. *Wiley interdisciplinary reviews. Developmental biology*.
73. de Laat W & Duboule D (2013) Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* 502(7472):499-506.
74. Amano T*, et al.* (2009) Chromosomal dynamics at the Shh locus: limb bud-specific differential regulation of competence and active transcription. *Dev Cell* 16(1):47-57.
75. van de Werken HJ*, et al.* (2012) Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nat Methods* 9(10):969-972.
76. Vernimmen D, De Gobbi M, Sloane-Stanley JA, Wood WG, & Higgs DR (2007) Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. *Embo J* 26(8):2041-2051.
77. Dixon JR*, et al.* (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485(7398):376-380.

78. Braasch I*, et al.* (2015) A new model army: Emerging fish models to study the genomics of vertebrate Evo-Devo. *Journal of experimental zoology. Part B, Molecular and developmental evolution* 324(4):316-341.
79. Wittkopp PJ, Haerum BK, & Clark AG (2004) Evolutionary changes in cis and trans gene regulation. *Nature* 430(6995):85-88.
80. True JR & Haag ES (2001) Developmental system drift and flexibility in evolutionary trajectories. *Evol Dev* 3(2):109-119.
81. Verster AJ, Ramani AK, McKay SJ, & Fraser AG (2014) Comparative RNAi screens in C. elegans and C. briggsae reveal the impact of developmental system drift on gene function. *Plos Genet* 10(2):e1004077.
82. Wang X & Sommer RJ (2011) Antagonism of LIN-17/Frizzled and LIN-18/Ryk in nematode vulva induction reveals evolutionary alterations in core developmental pathways. *PLoS Biol* 9(7):e1001110.
83. Sordino P, van der Hoeven F, & Duboule D (1995) Hox gene expression in teleost fins and the origin of vertebrate digits. *Nature* 375(6533):678-681.
84. 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.
85. Schmidl C, Rendeiro AF, Sheffield NC, & Bock C (2015) ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. *Nat Methods*.
86. Cao Z, Chen C, He B, Tan K, & Lu C (2015) A microfluidic device for epigenomic profiling using 100 cells. *Nat Methods*.
87. Buenrostro JD*, et al.* (2015) Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* 523(7561):486-490.
88. Cusanovich DA*, et al.* (2015) Epigenetics. Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* 348(6237):910-914.
89. Brind'Amour J*, et al.* (2015) An ultra-low-input native ChIP-seq protocol for genome-wide profiling of rare cell populations. *Nat Commun* 6:6033.
90. Jin W*, et al.* (2015) Genome-wide detection of DNase I hypersensitive sites in single cells and FFPE tissue samples. *Nature*.
91. Pieretti J*, et al.* (2015) Organogenesis in deep time: A problem in genomics, development, and paleontology. *Proc Natl Acad Sci U S A* 112(16):4871-4876.

APPENDIX I

SUPPORTING INFORMATION FOR CHAPTER III

Cm-CNS39

5′- CATCACTGCTTGTTGTGTATCCTATAACCTTTGAACGTAATAATCCCTCCATAATAAATGGGCCATTGAAGAGGAACAATAAATCATGAACAAT
TTTTAAACATTTCTCCATTCTACACCTAGGTAATATTAGTAATAAATACTACTAATAATCCTTGCATTTGATATAGTGGCTTATCACATCTTCGAGACG
TCTCAAAGCACTTCACAGAATATACTGTATATTGATATAGCATATACATGATATATTGATATAAGATAAAGGATGGATGAAAATACTTCCATCTTTAGA
GGATATTGCACAATTACGATTTGAAGACATAAGGTCCAACACAACACAAAACAACACAATGCTGCAGTTCTGAAAACTCATTGACTAAAACTAAA
ATAAAGTCCAACACAACAGGACAGAATCAACTCTTTACCCAGTGTGCTTTGAACACATTTAAAACTTGGGTCCAAAGAATTTGGAGAACTATCAG
TGCAGGATGTTTTCCTATGTGATTTGTTGTGTTTTGCTTTGTCTGTGTTTGGTTTATTCTGACTTGATGCAGGACCTTTGGGTCTGTTGTCTAGGTTT
GTGGTACTCACTGAGTTATCTCTGAATGGGTGACAAGGAGACATCTGGCTGTCAGCTACAATTACATCCCCGCACATGAAATTGTTTATAAATACTC
GGGTTTCCCTTTGAACGGAGATCTGTGACAAATAGCAGTTTGGGGTCATTAATGTAAGGGGTTGAGTCAATTTAAGGATTAAACAGAATTAAAAA
GAGATTTATGAGAGGTGGTGTATTCCATTTTAACTGAGTAGGAATCTCTTTTCTAGCATTATCTCAAAGATGCTGTTTGCTTTTACGGCCTACAAAAC
CAGGCCTTGAACATAGCAAGGCTTGTATCTGAAATAGCTATTGAAACTGCAAAGAGGAGTTCAGAACTGTCTGAATGAAAGCAATACCTGAATTTC
CGCCAAGAAATTTCTGCTTCACATTTTAAATTTGTACTGCCACCAGAAAGACTTGGATAATAAATACAGAAATTCTTTTTATACAGACATCACTAGTC
CAGGTGCCAAAGCTGGCGGGGAAAATAGGAAACCAATGCAGCGGCTTATGACATCAATGCTTTCGGACAATTTCAAAGGACTGAATATTTGTATTT
GTGACACAAACTACAATAGTCCAAATTGACTGCTGTTTCATAAACAGAAACTCTGTGGTTTGTGCTGGACCATTGAGCTCCTGATGAGTTCAGCTCC
AAGTTAAAAGATTCCCACATCCCAGGCAAGTTTCAAACTGCCTGAGAATCAGCTTGTTCAGCACCAGAACGACATAAAGAGACCAACGCAAATTCT
AACCCATCTCGGAGACCACAGAACTGGCACAGCATGTGTTGGAATGTTCTTAACCTATTCCAGGGCTCAGTTCAGTGGTAGAGCTGTATTGAAAGCA
CTAGCGCAAGTAAATATTGTTTGAATCCATGGAATCTAAATTAAATCCTCGCTGTAAACAGGATGTGAAGAATACTCTGCTGAAATATTTGAGTTGTAA
GTTTGATGCCTGATGTTTACCAGTTCTCCTTCTGTTAGTGGAGTCTGTTATCAGCATCAGTATGAACCAAGCCTATGTAATGTAAGGCAAGAAAGTGAC
CCATCCCACTACTTAGGAATAACAAAGAAACAGATTCACACAGAGATAGACCAGTGGCCCATTTGGACTTACCCCTTCAACTTCCCAAAAGATTGCTT
TGACACATCTATCTCATCATGTAATAACTTGGAATCCTTTAGGCAGTCTCTCTTACCCTCAAAGGCTGCATGGAATTTATTTTTTCAAAAGCTTCCTTTTA
TTTTCCACCCTCATTTCTACCCAGAGAGCAACCACCGTATTGTTATCAGACCCTTCAGATGTTCCCTTGCTAATCAGATCCCTTTTCTTAGTTGAAACCA
GATCCACTATTGTATTTTTTCTGGTTAGTTCGCTCACATTGTTTCAGAAATGA-3′

Cm-CNS65

5′- ATACAGGCCAGTATATGTCTGTTATAGTTCATAACCACTTCTAAATCTGTCATCACTACATAAGCACTGTACTAAATCCCTTCCATCAACTTTAAACA
GGAAGAGAATGAAGGAATATATCAATTACTGACATGCTTATATTGGGTTCAATATTCAGCACTGCAATATACTGACAATTCAAAGCTACCTTGGTAGCTG
GCTATTATACAGTGCAGTTGTACAAGGGTTTCTGTAATTTGGAAGCCTAAATAATGACTTTAATGTTCTGTTCACAACAAAAATAACCCAGGCTCCATTC
GGTTTAAAATGAAACCTAGTAATCACAGTTACTAATCCCAAGTGACATTAATTATATATTTTTTTGAATTGGCGGGACCTTTTCTGTTCCTTGCTATGTCT
TGTGTTAAATTATTAGAAGCGATTGAACAACGCCTGCCCTTCATTCAGTGGCATTTTGCAAAAAACAAGCTGAACATTGAACATGTTAAACCGTGCTG
GTATCTGAACTAACCCTGTCTTTGGCAGCTCCCAACTCTCTCCTAATTTGTATCAACCCCCCACTAACAATTGGTTTCATTTTCAATTAAATTATCAAACC
TGATTTTATAGTTGTCCTATACTGCTATATAAATCCTCATTGTAACTTTATATGTGTTTTTGTTAAACCTGGAAAATATAATTTCTTTGGCGTGTTTGATGTTC
TTGAGAAGAGTTACAGCAGATATTTATTGGACATAAATTCAGAAATGTTCCACTTTATAATGACTTTCCAGTGCCAGAACTTTTCTATGTACAATTTGTAT
TGTGATGAAATGCACTTAAAATATTCAGAAACACATTGACCCAGCTGTAGAAGGCACTGAATGAGCTTGAAAGAATCCTGAGTAAGGACAGAACCACA
TAGCTGTAAGCATTCTGTATTGTCACAGTTAATGAGGTCCATGTGTGGTGCCGTGTGTCATAAATAGTCCGCTCCTGGCTCTCAGTACACTTCATCACCAC
TTGTTAAATTCAGCAGAGGAAATGCTGGGGGAAGTTTAAAAAAGCGGCAGAGCTGGTGGTAGTAGTTTGGCCCTGGGGTAGCCCAGTAGCACAAAG
TGGTAAAGCATTTGCCTTGTTATATTTCACAGTTGCAGAGGGGCTCATCTGTCGGATCAGAGTCCTGATCAGTTGAGTTTGGTCACTGACTCCTGGGAA
CACTAGTTGCAATTGTCATCTAGTACAGAAAGGCTGAAATCCCTCCTGTTGATGGTGGTGGTGTTGGGCGGTGGGGAAAAAATAAGGCGGGGTGTATG
AAACAGGCTGGTGCTCACAGAGCTAGGGTCAGGCATGACTGTGGAGGGATTTAAAAGGTGAGAATGAGGATTTTTGAACCCGATCCTCTGGCGGATG
GGGTACCAAATGGAGCCGAGCCAGAGTGGCAGTGATAGGCTCGTGGAGCCTGGTGCGGGAAAGGACATGGACTGCAAGTTCTGGACATTACGCAGC
TTGTGCAGGGAATAGCTGGG-3′

Lo-CNS39-F

5′- CCAAGACCCCTTTCATGTGT - 3′

Lo-CNS39-R

5′- TCAAACACAATGTTACATGCTCA -3′

Table S3.1. List of oligos used for Chapter III

APPENDIX II

SUPPORTING INFORMATION FOR CHAPTER IV

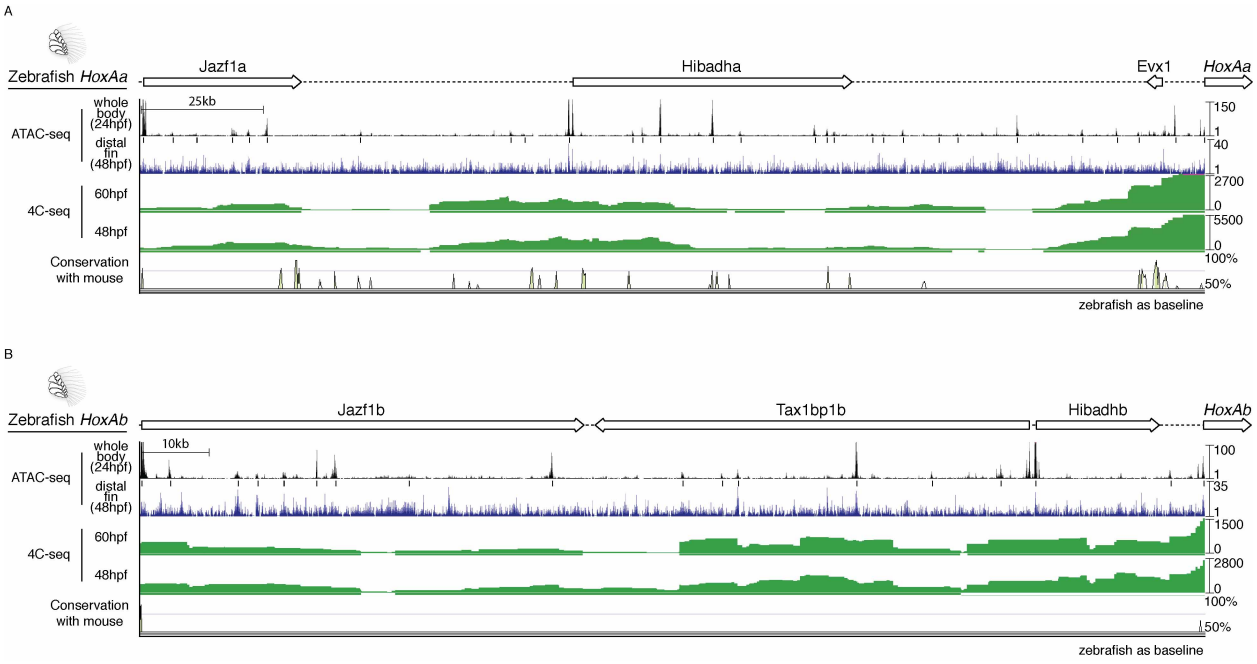


Figure S4.1. Epigenetic profiling and sequence conservation of the zebrafish *HoxAa* and *HoxAb* clusters. Schematic representations of the zebrafish two zebrafish *HoxA* clusters are shown at the top. ATAC-seq for both whole body and distal fin are shown, combined with 4C-seq data reveals areas of interaction with the *hoxa13a* and *hoxa13b* genes.

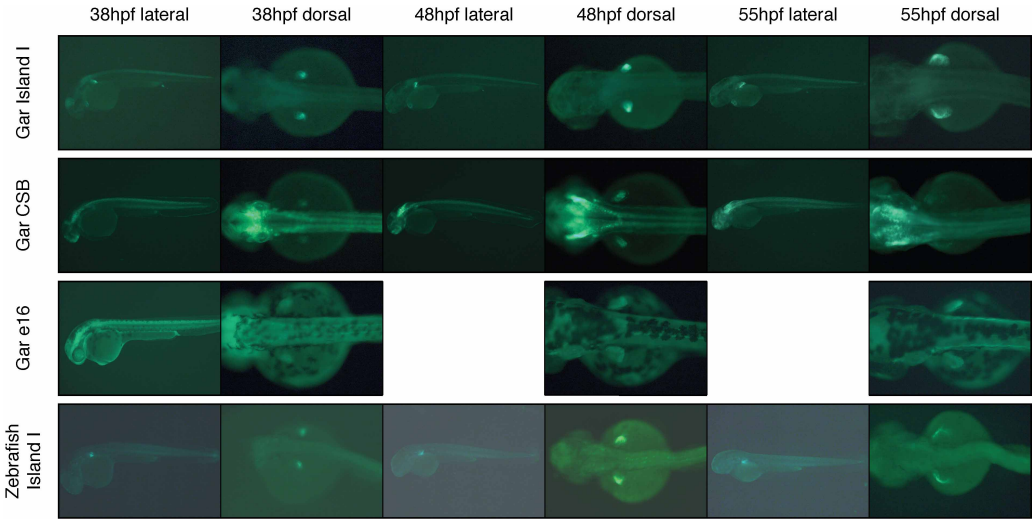


Figure S4.2. Whole body views of transgenic zebrafish. Lateral and dorsal views are shown for transgenic animals at 38, 48, and 55 hpf.

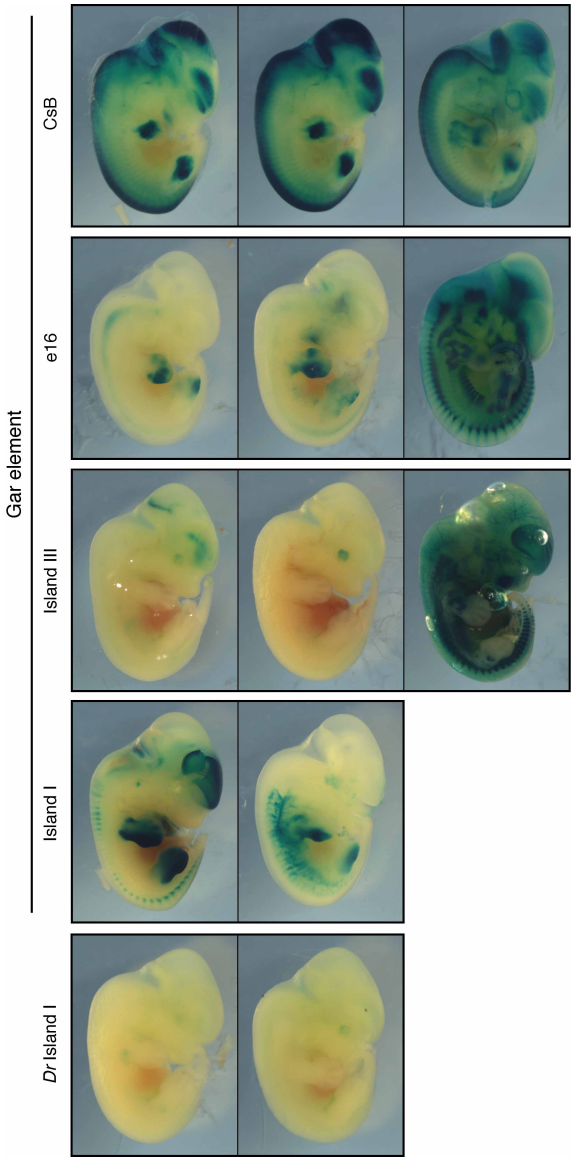


Figure S4.3. Summary of mouse injections for chapter IV. All injected mouse embryos that were positive for LacZ staining are provided at stage e12.5.

Mouse

Peaks	Overlap with ATAC-seq
H3K4Me1 (LICR ENCODE)	26210/52705 (49.7%)
H3K27ac (Cotney et al, 2013)	18082/25861 (69.9%)
Vista Enhancer Database (limb)	54/67 (80.6%)
Vista Enhancer Database (all)	138/250 (55.2%)

Zebrafish

Peaks	Overlap with ATAC-seq
H3K27ac (Bogdanovic et al, 2012)	20387/21839 (93.3%)

Table S4.1. Comparisons between mouse and zebrafish ATAC-seq and previously published reports.

Genomic region	# injected	stage screened	# fin positive	# fin negative	% w/fin signal	raise for line?
Lo CsB	172	48hpf	66	106	62.2	Yes
Lo Island I	152	48hpf	21	131	13.8	Yes
Lo e16	183	48hpf	26	157	14.2	Yes
Lo Island II	160	48hpf	4	156	2.5	No
Lo Island III	186	36, 48hpf	12	174	6.4	No
Lo Island IV.1	171	48hpf	6	165	3.4	No
Lo Island IV.2	168	48hpf	4	164	2.4	No
Dr Island I	136	48hpf	3	133	2.2	Yes

Table S4.2: Summary of zebrafish injections.

Name	Locus	Sequence
Zebrafish		
Dr_Island_I_F	Zebrafish Island I	AGCAACGCATGTCTTTCAACA
Dr_Island_I_R	Zebrafish Island I	ATAACGTTGTGTGCCTGCTG
Gar		
Lo_Island_I_F	Gar Island I	TGGCCTACAACACCAGTGAA
Lo_Island_I_R	Gar Island I	CAGATTTTGTGCGTTTCTCCT
Lo_CsB_F	Gar CsB	GGAGTCTCCCACAAGGTGAA
Lo_CsB_R	Gar CsB	CGAAGGCTCTGCACTACTCA
Lo_Island_II_F2	Genomic "area" of mouse Island II	GAGGTTGTGGGCTGTCCAAA
Lo_Island_II_R2	Genomic "area" of mouse Island II	CCACATTTGTGGAAATTCCTG
Lo_Island_IV_F2	Genomic "area" of mouse Island IV part 1	GCTTGAAAGCAACTGCATC
I4_Split_R	Genomic "area" of mouse Island IV part 1	GAGATGGCAACGCCTTATGT
I4_Split_F	Genomic "area" of mouse Island IV part 2	CCGTGTGATCCAAAGCAATA
Lo_Island_IV_R	Genomic "area" of mouse Island IV part 2	TTGGGCTGACCTGCTTTTAT
Lo_e16_R	HoxA enhancer e16	GTGATTTTCTCGGCATTTGG
Lo_e16_F	HoxA enhancer e16	CACCGACTTTGCTGTGTCAT
Gar Island III	Lo_Island_III_Long_F	GTGAGCCATGAGATGTACCG
Gar Island III	Lo_Island_III_Long_F	GTAAAACACTCCGGCACCTT
Mouse		
Tg1_F	Mouse Island I	TTCAGACTAGGCCCTCCAGA
Tg1_R	Mouse Island I	GACAGTGGGGACAACCCTAA
Tg2_F	Mouse Island II	GTGTGTGCGCTAGGGATTCT
Tg2_R	Mouse Island II	AGAGAGGGCTCGTCACTCAA
Tg4_F	Mouse Island IV	GAGTTAGCCACTCAGCCATGT
Tg4_R	Mouse Island IV	TGGTGTCTTTCCTCCATTCTG

Table S4.3. List of primers used for Chapter IV.

APPENDIX III

SUPPORTING INFORMATION FOR CHAPTER V

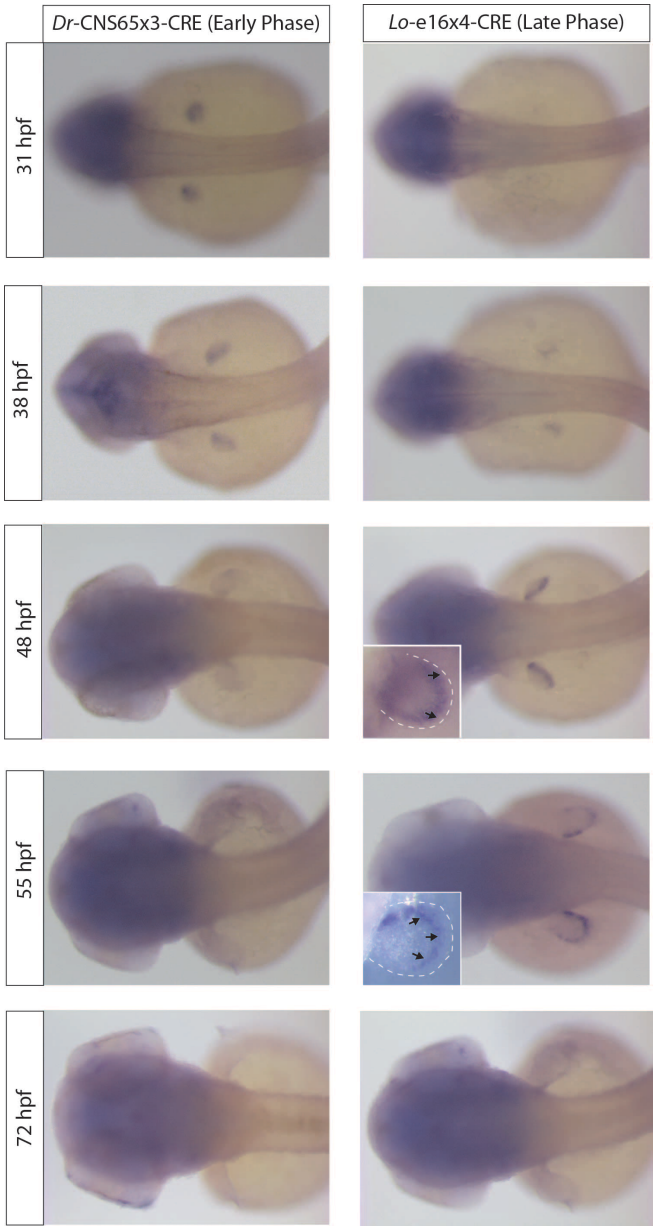


Figure S5.1. CRE expression dynamics in transgenic founder fish as detected by *in situ* hybridization. (Left) Fish transgenic for *Dr*-CNS65x3-CRE express the CRE transcript at ~31 and ~38 hpf throughout the whole fin, with transcription ceasing after ~38 hpf. (Right) *Lo*-e16x4-CRE transgenic fish express CRE robustly in the distal edge of the fin beginning at ~48 hpf, and ceasing after ~55 hpf. Inserts show close up view of the fins to show that at 48 hpf, active transcription takes place within the endochondral disc of the pectoral fin. At 55 hpf, CRE is weakly expressed in the proximal portion of the fin fold, likely as active transcription is in the process of ceasing. Fin outlines are marked by a white dotted line, and black arrowheads point to the distal edge of the endochondral disk.

CRISPR gRNA oligos

zebraHoxa13a_gRNA1_F
5′- AATTAATACGACTCACTATAGGGCAATCACAACCAGTGGAGTTTTAGAGCTAGAAATAGC -3′

zebraHoxa13a_gRNA2_F
5′- AATTAATACGACTCACTATAGGCAGTAAAGACTCATGTCGGTTTTAGAGCTAGAAATAGC -3′

zebraHoxa13b_gRNA1_F
5′- AATTAATACGACTCACTATAGGATGATATGAGCAAAAACAGTTTTAGAGCTAGAAATAGC -3′

zebraHoxa13b_gRNA2_F
5′- AATTAATACGACTCACTATAGGACACTTCTGTTTCTGGAGGTTTTAGAGCTAGAAATAGC -3′

zebraHoxd13a_gRNA1_F
5′- AATTAATACGACTCACTATAGGCTCTGGCTCCTTCACGTTGTTTTAGAGCTAGAAATAGC -3′

zebraHoxd13a_gRNA2_F
5′- AATTAATACGACTCACTATAGGCGAACTCTTTAAGCCAGCGTTTTAGAGCTAGAAATAGC -3′

Zebrafish_gRNA_R
5′- AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCT
AGCTCTAAAAC -3′

T7 assay primers

zebraHoxa13a_Cont_F
5′- CTGCAGCGGGTGATTCTG -3′
zebraHoxa13a_Cont_R
5′- CTCCTTTACCCGTCGGTTTT -3′
PCR product: 810 bp

zebraHoxa13b_Cont_F
5′- GAAGCTTATCACTAGAATCTTTACAGC -3′
zebraHoxa13b_Cont_R
5′- TTTTTCTCAGGGCCTAAAGGT -3′
PCR product:1089 bp

zebrafishHoxd13a_Cont_F
5′- AGCTGCCCAATCACATGC -3′
ZebrafishHoxd13a_Cont_R
5′- CGATTATAAATTCAGTTGCTCTTTAG -3′
PCR product: 823 bp

Danio_and1_F
5′-ACCTGCTCCTGCTCCAGTTA -3′

Danio_and1_R
5′- CACATCCTCTTGAGGGGAAA -3′

Table S5.1. List of Oligos used for Chapter V

Lineage tracing oligos:

CRE_PCR_F_NcoI
5′- CGCCCTTCCATGGATGGCCAATTTACTGACCGTAC -3′

CRE_PCR_R_BglII
5′- GTTCTTCTGAAGATCTCTCTGGGGTTCGGGGCTGCAGG -3′

CRE_Genotype_F
5′- CGTACTGACGGTGGGAGAAT -3′

CRE_Genotype_R
5′- ACCAGGCCAGGTATCTCTGA -3′

CRE_Probe_F
5′- ATGGCCAATTTACTGACCGTAC -3′

CRE_Probe_R
5′- CTAATCGCCATCTTCCAGCAGGCG -3′

Random_Oligo_SmaI
5′- CTGCTCTGGTCAGCCTCTAATGGCTCGTTAGATAGTCTAGCCGCTGGTAATCACTCGATGACCTCGGCT
CCCCATTGGTGCTACGGCGATTCTTGGAGAGCCAGCTGCGATCGCTAATGTGAGGACAGTGTAATATTAG
CAAGCGATAAGTCCCCAACTGGTTGTGGCCTTTTGAAAAGTGAACTTCATAACATATGCTGTCTCACGCAC
ATGGATGGTTTGGACAAATTTGATTCAAGTCTGATCAACCTTCACTGCTCTAGAATCAAAAGCAGTGATCTC
CCGGGTGCGAAATAAA -3′

Lo-e16_Oligo_1_BamHI_SmaI
5′- CCCCCAAAAAATGACAAAACTCTTGGAATTTATTACGGCTTTGGCAATAGAGACCGCTTTTTGGGTGG
CTCAGTAAAAGGTTTGATGTTCACGTATCGCCTTTTAAATGCATTCATTCCTCTTTCATATGTGTGCAACTGT
TTAGATACATCATAAAAATGTCACCATTGAGGTTCCCCATTAGGCATCTACCCGTTCTCCTCCAGGCCATGGA
GATAAATTTGGACCAGGTGATCCCCTCCTAGAAGAGCCCTTGATGTCTTCTGGTAATGAGTTGAAAGCGGAA
GCTGTCAGCCTTCAGCAGGCATGAAGATGCAATTAGAGCTGCGTTCAAAGTGCCCAGGCAGTCTCATAAGG
AGCACTAGCCTTGGTGTAAGCTGCTTATTCACAGATCAGTTATGTAAGGGTACAGCAAAAAGGCAAGACACT
CGATTTTTGAATGACACAGCAAAGTCGGTGCGGATCCCGAGTTTGCCCGGGTAGCCC -3′

Lo-e16_Oligo_2_BamHI_SalI_SmaI
5′- CCCCCAAAAAATGACAAAAGGATCCGAATTTATTACGGCTTTGGCAATAGAGACCGCTTTTTGGGTGG
CTCAGTAAAAGGTTTGATGTTCACGTATCGCCTTTTAAATGCATTCATTCCTCTTTCATATGTGTGCAACTGT
TTAGATACATCATAAAAATGTCACCATTGAGGTTCCCCATTAGGCATCTACCCGTTCTCCTCCAGGCCATGGA
GATAAATTTGGACCAGGTGATCCCCTCCTAGAAGAGCCCTTGATGTCTTCTGGTAATGAGTTGAAAGCGGAA
GCTGTCAGCCTTCAGCAGGCATGAAGATGCAATTAGAGCTGCGTTCAAAGTGCCCAGGCAGTCTCATAAGG
AGCACTAGCCTTGGTGTAAGCTGCTTATTCACAGATCAGTTATGTAAGGGTACAGCAAAAAGGCAAGACACT
CGATTTTTGAATGACACAGCAAAGTCGTCGACTTCTCCGAGCCCGGGAAACTAGCCC -3′

Table S5.1 continued

Lo-e16_Oligo_3_SalI_BglII_SmaI
5′- CCCCCAAAAAATGACGTCGACCTTGGAATTTATTACGGCTTTGGCAATAGAGACCGCTTTTTGGGTGG
CTCAGTAAAAGGTTTGATGTTCACGTATCGCCTTTTAAATGCATTCATTCCTCTTTCATATGTGTGCAACTGT
TTAGATACATCATAAAAATGTCACCATTGAGGTTCCCCATTAGGCATCTACCCGTTCTCCTCCAGGCCATGGA
GATAAATTTGGACCAGGTGATCCCCTCCTAGAAGAGCCCTTGATGTCTTCTGGTAATGAGTTGAAAGCGGAA
GCTGTCAGCCTTCAGCAGGCATGAAGATGCAATTAGAGCTGCGTTCAAAGTGCCCAGGCAGTCTCATAAGG
AGCACTAGCCTTGGTGTAAGCTGCTTATTCACAGATCAGTTATGTAAGGGTACAGCAAAAAGGCAAGACACT
CGATTTTTGAATGACACAGCAAAGTCGAGATCTTCTCCGAGTCCCGGGAACTAGCCC -3′

Lo-e16_Oligo_4_BglII_SmaI
5′- CCCCCAAAAAATGAGATCTCTCTTGGAATTTATTACGGCTTTGGCAATAGAGACCGCTTTTTGGGTGGC
TCAGTAAAAGGTTTGATGTTCACGTATCGCCTTTTAAATGCATTCATTCCTCTTTCATATGTGTGCAACTGTTT
AGATACATCATAAAAATGTCACCATTGAGGTTCCCCATTAGGCATCTACCCGTTCTCCTCCAGGCCATGGAGA
TAAATTTGGACCAGGTGATCCCCTCCTAGAAGAGCCCTTGATGTCTTCTGGTAATGAGTTGAAAGCGGAAGC
TGTCAGCCTTCAGCAGGCATGAAGATGCAATTAGAGCTGCGTTCAAAGTGCCCAGGCAGTCTCATAAGGAG
CACTAGCCTTGGTGTAAGCTGCTTATTCACAGATCAGTTATGTAAGGGTACAGCAAAAAGGCAAGACACTCG
ATTTTTGAATGACACAGCAAAGTCCCCGGGTTCTCCGAGAAACTAGCCC -3′

Primers for final PCR to clone into destination vector:

e16x4_F_Xho1:
5′- CAGGCTCCCTCGAGCCCCCAAAAAATGACAAA -3′

e16x4_R_SmaI:
5′- CGAATTCGGTCCCGGGACTTTGCTG -3′

Dr-CNS65_Oligo_1_BamHI_SmaI
5′- GAGGTTCACCTTTAACCACAACACGTAACAAATCAGATCTCAGAAGACAAGCCGCTTCAGAAGTCGT
GCTCAGTGTTGCATTCAAGCGTGTGTGATTTTCCAGACTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGT
GTGTGTGTGTGTGTGTGCTCTCAGAGATCTTTCATTGGGGAATCTTTCCTGTGTGAGAGCTGCGGTCTCA
GCGGCTGATTTATGGCGCTCCGCAGCTATGCTCATGCTACGCTAACAATGCTCATTAAAAAGAGGATGTC
ATCACTCCGCGACACCGCAGGACTCGTATGTGTCACATGCATCCTCAATACAGCGAACCGCTGACCAATA
CCGTCCACAACATCCTGTAAATCTGTCATCGCCAGCATGGCCGCGGAAACACACACACACACACACACC
ATTAGAGTGCAGTAATAGAGGATCAGAGGTTAATGTGGAGCTGTTTGCTGGTGTTTAGTTTTGTATTAGA
GGATTTCACGTGCTTACAGCTATGTGTGTGTGTTTGAACAGTAAAGAAAGTATAAAAAGTAAAATATTATA
ATCTTAAGCCACTCGTAATCTTCAAAAAACACTAAAATGCAAGAATAACGGATCCCTTTCACACTAGAGC
CCGGGAAAGTGAGCGTT -3′

Dr-CNS65_Oligo_2_BamHI_SalI_SmaI
5′- GAGGTTCACCTTTAGGATCCACACGTAACAAATCAGATCTCAGAAGACAAGCCGCTTCAGAAGTCGTG
CTCAGTGTTGCATTCAAGCGTGTGTGATTTTCCAGACTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTG
TGTGTGTGTGTGTGCTCTCAGAGATCTTTCATTGGGGAATCTTTCCTGTGTGAGAGCTGCGGTCTCAGCGGC
TGATTTATGGCGCTCCGCAGCTATGCTCATGCTACGCTAACAATGCTCATTAAAAAGAGGATGTCATCACTCC
GCGACACCGCAGGACTCGTATGTGTCACATGCATCCTCAATACAGCGAACCGCTGACCAATACCGTCCACAA
CATCCTGTAAATCTGTCATCGCCAGCATGGCCGCGGAAACACACACACACACACACACCATTAGAGTGCAGT
AATAGAGGATCAGAGGTTAATGTGGAGCTGTTTGCTGGTGTTTAGTTTTGTATTAGAGGATTTCACGTGCTTA
CAGCTATGTGTGTGTGTTTGAACAGTAAAGAAAGTATAAAAAGTAAAATATTATAATCTTAAGCCACTCGTAAT
CTTCAAAAAACACTAAAATGCAAGAATAAGTCGACCCTTTCACACTAGAGCCCGGGAAAGTGAGCGTT -3′

Table S5.1 continued

Dr-CNS65_Oligo_3_SalI_SmaI
5′- GAGGTTCACCTTTAGTCGACACACGTAACAAATCAGATCTCAGAAGACAAGCCGCTTCAGAAGTCGTG
CTCAGTGTTGCATTCAAGCGTGTGTGATTTTCCAGACTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTG
TGTGTGTGTGTGTGCTCTCAGAGATCTTTCATTGGGGAATCTTTCCTGTGTGAGAGCTGCGGTCTCAGCGGC
TGATTTATGGCGCTCCGCAGCTATGCTCATGCTACGCTAACAATGCTCATTAAAAAGAGGATGTCATCACTCC
GCGACACCGCAGGACTCGTATGTGTCACATGCATCCTCAATACAGCGAACCGCTGACCAATACCGTCCACAA
CATCCTGTAAATCTGTCATCGCCAGCATGGCCGCGGAAACACACACACACACACACACCATTAGAGTGCAGT
AATAGAGGATCAGAGGTTAATGTGGAGCTGTTTGCTGGTGTTTAGTTTTGTATTAGAGGATTTCACGTGCTTA
CAGCTATGTGTGTGTGTTTGAACAGTAAAGAAAGTATAAAAAGTAAAATATTATAATCTTAAGCCACTCGTAAT
CTTCAAAAAACACTAAAATGCAAGAATAACCCTTTCACACTAGAGCCCGGGAAAGTGAGCGTT -3′

Primers for final PCR to clone into destination vector:

CNS65x3_F_XhoI:
5′- GCAGGCTCCTCGAGGAGGTTCACCTTTAACCA -3′

CNS54x3_R_SmaI:
5′- AACGCTCACTTTCCCGGGTCTAGTGT -3′

Table S5.1 continued