Spring 2019

Molecular Homology & the Ancient Genetic Toolkit: How Evolutionary Development Could Shape Your Next Doctor's Appointment

Elizabeth G. Plender

MOLECULAR HOMOLOGY AND THE ANCIENT GENETIC TOOLKIT: HOW EVOLUTIONARY DEVELOPMENT COULD SHAPE YOUR NEXT DOCTOR'S APPOINTMENT

A thesis submitted to

Regis College

The Honors Program

In partial fulfillment of the requirements

For Graduation with Honors

By

Elizabeth G. Plender

May 2019

Thesis written by

Elizabeth Plender

Approved by

Thesis Advisor

Thesis Reader

Accepted by

Director, University Honors Program

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

Firstly, I would like to thank my thesis advisor, Dr. Max Boeck, and my reader, Dr. Bethany Lucas. Dr. Boeck took me in as a research orphan, pushed me to practice patience, and molded me into a confident scientist. Dr. Lucas fostered my obsession with developmental biology and showed me that nerdiness is *always* in style. You both have been infallible in your encouragement, and for that I am so grateful.

Next, I would like to thank Dr. Saulius Sumanas, Dr. Sanjeev Metikala, and my summer lab mates at Cincinnati Children's Hospital. I had more than my fair share of fun setting up mating tanks in the fish room and ditching work to go day-drinking.

Thank you to the Regis College Honors Program directors, Drs. Howe and Kleier, who have been part of my undergraduate journey since the first day that I stepped on campus. I'll never forget "the brick," *A Christmas Story*, our AJCU adventures, and the love for learning that both of you have instilled in me. How awesome it is to be part of a program where everyone is smart, but we're not a**holes about it!

Another thank you goes out to my science barbies, Josie Gruber and Angela Vu. You two have been the most tenacious, exuberant, and compassionate friends.

I would be remiss if I didn't thank my parents, the people whose DNA I share and whose love I cherish. Those hours of walking Crown Hill and studying for genetics, biochemistry, and developmental biology are sure to win me the Nobel Prize one day. You both have taught me how to love myself, a task which is much too complicated in this world.

Finally, I would like to thank Regis University, the place where I rediscovered how to have conversations with God. This community has made me a better person, and for that I am humbled. Here's to many memories of remarkable friendships, countless aha moments, and four years of tremendous happiness.

Chapter 1

INTRODUCTION

There's a rainbow shining in a bead of spittle
Falling diamonds in rattling rain
Light flexed on moving muscle
I stand here dazzled with my heart in flames
At this world of wonders

"World of Wonders"
--Bruce Cockburn

Our world is full of wondrous things. From the shimmer of a child's laughter to the intricate painting of a ladybug's shoulders, our home is colored with an infinite, incandescent kind of splendor. Thousands of years of poetry, artistry, and science have taught us that the natural world is nothing short of magnanimous. What better gift could we have been given than that of God's exquisite creation?

I've never been the kind of person who finds much spiritual solace in the church. I can't deny that gothic-style cathedrals with expansive acoustics and well-ordered pews do not astound me; it's the people, rather, that make me hesitant to find sanctuary in these places. History tells us that organized religion can be a fickle thing, an institution

wrought with human folly. Where do the rest of us fit in, the folks who struggle desperately to hear God over the mindless din?

When I first came to Regis, I had very little faith. I had spent my high school years fighting tooth and nail to belong, and, consequently, had little time to spend thinking about spirituality. I grew up among religious hypocrites, people who depicted God as a malicious, judgmental, and all-around unlikeable guy. While I found some comfort in my music, a passion that distracted me from the harsher parts of high school, I still had no clue how to relate to a God so distant. I was spiritually stuck—where was this God that I had heard about, a God who was supposedly a trusted advisor, confidant, and friend?

My first year at Regis was spent almost entirely ignoring God. I had been given the incredible gift of a scholarship in biology, a kind of "golden handcuffs" that pushed me in the direction of the natural sciences. I had always been good at science; so, naturally, I figured I should become a doctor. I kept on telling myself that *this* was what I wanted, that this would make God happy with me. Like many other honors students, I loaded my schedule with mindless activities and fruitless distractions. It was much too difficult to face the reality of my confusion.

I will never forget the day that I found my vocation. At the time, I had been taking genetics and had (quite literally) fallen in love with it. Years of biology lectures had finally culminated in some truly astounding "aha" moments—I felt as if my soul had finally been nourished. One bleak February afternoon, I was working in the biology prep room and had been tasked with the infamous job of cleaning up fly larvae for the labs. As

I meticulously washed each vial with a small scrubber and felt the hot water run over my hands, I hummed a tune that I had been learning in my voice lesson. I looked outside the prep room window to see snow falling peacefully on campus. In that moment, I felt a sort of simple, passionate peace while watching the water and soap circle the drain. I was finally allowing myself to feel the avalanche of emotion that I had bottled up for years. God was right there, nudging me to consider a future dedicated to the relentless pursuit of His creation, a future full of fly larvae, failed experiments, and ultimate wonder. In that moment, I felt a clarity of thought and emotion that demolished all of my old dreams; I knew I *needed* to become a scientist.

Much like listening to a good piece of music or seeing the face of a loved one, I felt myself opening up to new experiences over the next few years. As I devoured knowledge of the perfect structure of the helix, the ecosystem of a living cell, and the breathtaking homology of the animal kingdom, I cultivated a more beautiful image of God. I have come to realize one most precious thing—that the God of my childhood does not exist. He who revealed Himself to me that day in the prep room is the same God who made the cosmos, a universe both beautifully simple and vastly complex; namely, a God with one hell of a personality. The laboratory has become my church, the benchtop my pew, and the microscope a sacrament of my devotion to the divine. Forget the judgmental God that years of bible school shoved down my throat—give me nature's God, a God that made people like me who crave to know the infinite.

I found the inspiration to write about homology from the most unexciting of sources—my junior seminar on biology research literature. Tasked with the opportunity to write about any current topic in evolution, I found myself fascinated with the molecular similarities between members of the animal kingdom. I remember devouring Sean Carroll's publications on deep homology. How could anything be more fascinating than the concept of an ancient genetic toolkit? I couldn't believe that I had never heard of Urbilateria, the *Hox* clock*,* or convergent evolution. The possibility of studying life from both the molecular *and* organismic perspectives quickly drew me into an obsession even more intense than that of genetics—developmental biology.

In my need to know more, I spent the entirety of my winter break that year applying to almost every developmental laboratory across the country that offered a summer undergraduate internship position. The first invitation that I received ended up being the perfect fit, and I was all set to work at Cincinnati Children's Hospital in a zebrafish lab studying vasculogenesis. Armed with the knowledge that I had learned from my developmental biology lecture, I hopped onto a plane and embarked on my first solo adventure in the Midwest. At first glance, my project seemed to have nothing to do with evolution; by the end of the summer, however, I had realized that the proteins I was working with were in fact homologs in numerous other model organism systems. It all seemed to fall into place—I would be able to write about my fantastic summer research, all within the scope of my beloved evo-devo.

Contrary to what the seniors told me would happen my freshman year of honors, I have thoroughly enjoyed writing this thesis. Both intricately academic and profoundly personal, this last year of writing has helped me to develop a confidence in my chosen vocation that I couldn't have thought possible. I walked into this journey with a love for homology and now have a deeper understanding of its role in God's artistry. Many of the questions that I posed at the beginning of this process remain unanswered; it is through these uncertainties, however, that I have had the most fun.

Chapter 2

WHAT IS HOMOLOGY?

2.1: Introduction

We are in the midst of a genomic revolution. Beginning with the completion of the human genome project at the turn of the century, research across the life sciences has quickly become reliant on computational methodologies and bulk data analysis. This phenomenon of dependence has been seen across a wide array of subfields, with genomics informing our understanding of both conservation biology and biomedical inquiry. Genomics is to our generation what nuclear physics was to our grandparents—flashy, alluring, and full of possibilities.

In the presence of such remarkable computational power, biomedical discoveries entrenched in evolutionary biology may seem impractical. Why study comparative biology when one can simply genotype a patient? What is the value of continuing to study our evolutionary history when so many diseased phenotypes can be explored through computational measures? While genomic technologies have certainly opened up a floodgate of research possibilities, we simply cannot forget the value of the model organism system, an age-old institution that has provided the very foundation of all that

we know in the life sciences. Where would we be today if we had not marveled at the simple homology found throughout the animal kingdom?

The history of the use of model organisms in the biological sciences dates back to the infancy of the scientific method. In his publication *Generation of Animals,* Greek philosopher Aristotle provided numerous written accounts of his use of model organisms to study the comparative nature of animal development (Malpas, 2012). Since then, rodents, flies, worms, fish, amphibians, primates, and other countless organisms have diversified our understanding of the history of life on Earth. Very few subfields have relied on the merits of using a model system as intensely as that of developmental biology; when, then, did all of this developmental homology (i.e. "sameness") across species begin?

2.2: Urbilateria and the Cambrian Explosion

Developmental research concerning the morphology of the animal kingdom considers the origin of bilateral species as a crucial turning point in evolutionary history. Bilateral organisms first appeared on Earth between 535 and 525 million years ago during a period known as the "Cambrian Explosion." At this period in the fossil record, species diversification was at a remarkable high (DeRobertis, 2008). While Precambrian records include numerous complex organisms, such as that of sponges, radial animals, and small bilaterian species, it was not until the Cambrian Explosion that the majority of our modern bilateral phyla (levels of organismic classification) became evident and valuable to our understanding of animal diversification (Kirschner & Gerhart, 2006).

By the end of the Cambrian era, all but a select few of the modern metazoan (multi-cellular) phyla had evolved, as well as each of the thirty major bilateral animal phyla. The species that evolved during this period exhibited patterns of increased behavioral complexity as compared to their recent ancestral counterparts, supporting the notion that the Cambrian explosion was a time of immense evolutionary innovation (Knoll & Carroll, 1999). How this period of evolutionary history led to such profound diversification remains an enigma; however, the prevailing hypothesis, called the "Snowball Earth Scenario," stipulates that several periods of glaciation between 750 and 550 million years ago led to the survival of only those species that could survive extreme bottlenecking (a type of genetic isolation that leads to a loss in diversity). Upon warming, a late Protozoic increase in atmospheric oxygen levels most likely removed the environmental barrier that previously had prevented the diversification of bilateral, "higher" organisms (DeRobertis, 2008; Knoll & Carroll, 1999).

Perhaps the most exciting evolutionary outcome of the Cambrian explosion was the emergence of Urbilateria, the last common bilaterian ancestor of all animal species on Earth. Paleontological records indicate that Urbilateria emerged around 560 million years ago (right before the onset of the Cambrian explosion) and subsequently diverged into protostomes and deuterostomes (Kirschner & Gerhart, 2006; DeRobertis, 2008). Protostome and deuterostome species differ at one of the earliest stages in development; while the oral end of a protostome develops from the first opening of the blastopore (opening in the central cavity of an embryo), this same cavity of the deuterostome develops from the second opening. Some example protostomes include arthropods,

nematodes, and annelids, while deuterostomes include the well-known chordates (which later gave rise to the vertebrates). These distinctions represent the first major bifurcation in the "Animalia" kingdom, a seemingly small yet vastly important step in the cascade of our evolutionary history (Figure 2.1).

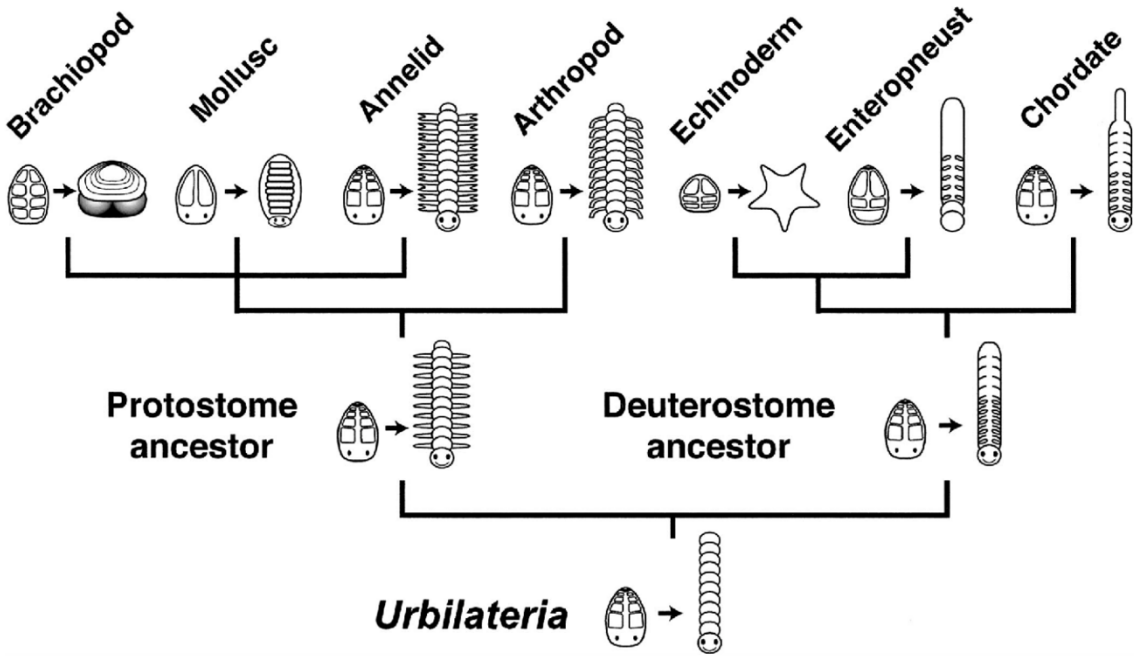

Figure 2.1. Branching of protostomes and deuterostomes from an Urbilaterian ancestor and subsequent lineages associated with these two groups. Humans (which are vertebrates) fall within the category of "Chordate" and are classified as Deuterostomes. Illustrations depict the emergence of diverse bauplan characteristics across the animal kingdom, as compared to the original morphology of Urbilateria. Adapted from Balavoine & Adoutte, 2003.

Urbilateria has been characterized as a small, worm-like organism. Measuring at roughly 1 millimeter in length, it likely featured a through-gut, anterior head, three organizational germ layers, and multiple primitive sensory organs. It was also much more complex than its radial ancestor and probably housed a primitive nervous system (Kirschner & Gerhart, 2006). Inferences about the anatomy and physiology of Urbilateria

have been made through comparison of the characteristics present in the various modern animal phyla—if a conserved gene is found in both deuterostomes and protostomes, it stands to reason that the sequence was also present in the genome of Urbilateria. Recently, efforts have been made to computationally reconstruct the Urbilaterian genome through the tracing of gene duplications, deletions, and mutations across the larger animal phylogenetic tree. It has become evident that Urbilateria's genome likely featured the majority of developmental networks that distinguish the animal kingdom today (DeRobertis, 2008).

Thanks to comparative genomics, we now know that protostomes and deuterostomes share a remarkable number of developmental pathways that trace back to Urbilateria and the Cambrian explosion. Some of these pathways include the canonical *hox* genes, a plethora of growth factors, Wnt, the Hedgehog proteins, Notch ligands, and various proteins implicated in epidermal-growth factor signaling. These observations have supported the hypothesis that Urbilateria was the first organism to contain a complete and "modern" animal genome. While we are certainly limited in our reliance on extant (currently-living) animal genomes to draw inferences about the first bilateral genome, evolutionary biologists have nevertheless come to realize that, from a developmental perspective, members of the animal kingdom are highly similar at the molecular level (Knoll & Carroll, 1999).

2.3: Darwin, Natural Selection, & the Basics of Evolutionary Theory

In *The Descent of Man,* Charles Darwin claimed, "man still bears in his bodily frame the indelible stamp of his lowly origin" (Shubin, Tabin, & Carroll, 2009). While little was known about the molecular basis of heredity at the time of Darwin's famous expedition, he was remarkably prescient in his assertion that all members of the animal kingdom must share the fundamental basics of an evolutionary past. When asked why the Proterozoic-Cambrian boundary in the paleontological record reflects the sudden emergence of immeasurably diverse taxa, Darwin, like countless others, was stumped; why has there been so little change in bauplans (body plan morphologies) at the phylum and super-phylum levels since the Cambrian explosion, while the opposite has been observed within distinct classes (Knoll & Carroll, 1999; Davidson & Erwin, 2006)?

Before diving head-first into the robust "deep end" of modern evolutionary theory, it is important to consider the role of Darwinian evolution as foundational to all aspects of the field. Darwin's theory of natural selection stipulates that environmental constraints act upon individuals in a population such that only those that are most well-adapted to their environment will effectively reproduce and pass on their genetic material. Perhaps the most important element of Darwin's theory is its realization over evolutionary time – as environmental constraints change and become more selective over generations, the most "favorable" of phenotypes will accumulate within the gene pool and lead to the potential for long-term speciation (Stern & Orgogozo, 2009). At the molecular level, it would stand to reason that mutations with pleiotropic effects (genes linked to multiple phenotypic outcomes) would be less likely to support long-term

adaptivity because of their higher "sensitivity" to dramatic alterations; however, critics of the Darwinian model of evolution cite that selection need not be the only method in which evolutionary novelty may arise. The neutral theory of molecular evolution stipulates that most changes within and between species at the molecular level are in fact supported by the genetic drift (random changes in genetic variation in a population) of neutral, non-selective alleles (Ohta, 1992). Regardless of these differences, the modern perspective on evolution states that biased subsets of random mutations most often produce compounded phenotypes specific to various branches along the phylogenetic tree of the animal kingdom (Stern & Orgogozo, 2009).

The adaptive evolution of diverse species to aquatic, terrestrial, and aerial environments lead to the diversification of body plans and the pressing need for adaptability within the animal kingdom following the emergence of Urbilateria. Historically, evolutionary biologists and paleontologists focused on the homology of limb development and the diversification of bauplans within Cambrian records because these comparisons could be made with the naked eye. While discovery of the anatomical similarities of limb development has certainly shed light on the general patterns of homology, it wasn't until the advent of modern molecular technologies that biologists could truly witness the homology of development. Within the context of bauplan innovation since the Cambrian explosion, our best clues seem to lie in the developmental networks that have formed us all (Shubin & Tabin, 1997).

2.4: The History of Biological Homology

Since its conception, the evolutionary basis of homology has generated a remarkable amount of debate within and among communities of biologists. As of 1994, there were nineteen distinct definitions for the phenomenon, each specific to a subdiscipline of biology (Hall, 2003). The concept of homology was originally made prominent in 1843 by Richard Owen, who defined it as "the same organ in different animals under every variety of form and function" (Sansom & Brandon, 2007). As the number of studies concerning homologous morphology has grown, so has our collective understanding of the relationship between evolutionary time and the emergence of new body plans in species on Earth today.

Historically, biologists based their concept of homology on phenotypic, anatomical data directed at the organismic level (Brigandt, 2003; Wagner, 2015). The vast majority of these discoveries were grounded in synapomorphies, which are characteristics shared exclusively by species that have a common, direct ancestor. Biologists who tackled questions concerning homologous features were lumped under the subfield category "phylogenetic systematics" (Wake, 2003). Fervent supporters of the theory of natural selection, titled "Darwinian biologists," claimed that two characters could only be homologous if the origin of said characters could be explained by descent from a common ancestor (Sansom & Brandon, 2007; Amundson, 2005). This concept contrasts with that of "analogues," which are functionally homologous characters that do not share origins in a common ancestor (Hall, 2003). Thomas Huxley, a famous English biologist, was one of the first scientists to note that many homologous adult structures are

more obviously similar when considered during the stages of development (Gilbert, Opitz, & Raff, 1996). Richard Owen complicated this observation when he made the assertion that descent from a common ancestor is not sufficient to invoke the formation of homologous characters; rather, he proposed that two characters can only be homologous if they share an underlying mechanism of development (Amundson, 2005). These observations and those of developmental biologist Ernst Haeckel led to the creation of the famous theory of recapitulation, which states that the successive stages of development mirror the stages of evolution in that species' remote ancestors (Richardson & Keuck, 2002). While we now understand the inherent flaws in Haeckel's theory, little could be done in these times to test whether the phrase "ontogeny recapitulates phylogeny" was an entirely accurate assessment.

At the time of its conception, the identification of homologous characters among species was considered to be a useful form of "phenotype organization." We now know that studying homology is worth so much more than this—it is a key pattern in which evolution gives rise to new species (Wagner, 2015; Wake, 2003). This transition was largely influenced by discoveries in the field of molecular genetics. Thomas Hunt Morgan, a famous American geneticist and the man responsible for elucidating the role of chromosomes in genomic organization, was a proponent of the idea that genetics, rather than paleontology alone, is key to understanding the nuances of evolution. He claimed that the origins and development of higher taxa could be explained through many of the pattern that govern population genetics (Gilbert, Opitz, & Raff, 1996).

While different subfields of biology have distinct criteria in which to determine whether two characters are homologous, a common definition considers it as an "investigative kind concept"—it links objects of the same kind, even if the underlying mechanisms that govern said similarities remain unknown (Brigandt, 2003). Studies that concern evolutionary homology fall within the categories of genetic homology, developmental/ontogenetic homology, and organismic homology, as well as often take a taxic or transformational approach to study design (Hall, 2003; Wake, 2003). The taxic approach utilizes character and morphological data to make phylogenetic hypotheses, while the transformational approach tracks changes in homologous structures over evolutionary time and postulates their potential causes (Wake, 2003). Three general criteria are often considered in the discernment of homologous characters—these include whether they arose in similar positions within the organisms themselves (morphology), whether they have identical (or at least similar) functions, and/or whether they have similar intermediate and transitory forms during the life cycle (Rutishauser & Moline, 2005).

While much of the original criteria of evolutionary homology remain the same, the expansion and diversification of subfields in the biological sciences has led to the creation of numerous nuanced definitions. Juggling the differences in homology between subfields like genetics, anatomy/physiology, and phylogenetics presents a significant challenge for cross-disciplinary studies; however, an even more pressing problem is found in the need for selecting an appropriate hierarchical level at which to make such analyses (Wake, 2003). This brings up a most difficult and frustrating facet of

homology—the simple fact that it is entirely reliant on perspective. An anatomist's claim that two morphological body parts are canonically homologous may, at the molecular level, turn out to be entirely false; conversely, a lack of obvious homology at the organismic level may in fact hide the presence of a developmental network that is remarkably homologous (Hall, 2003). The ambiguous nature of homology, while frustrating and exhausting for biologists, remains a beautiful and humbling example of how very little we truly know about the patterns of evolution.

2.5: Subfield Definitions of Homology

As with other evolutionary principles, it comes as no surprise that defining homologous patterns between species is dependent on the level of hierarchy utilized by a given subfield and study design. Much of our current understanding of canonical, well-known homologous features is inherently linked to the flexibility in which we may define homology itself (Rutishauser & Moline, 2005). While there is no "correct" answer when it comes to defining homology, it is intriguing to note that the different subfields often define homology within the context of their own field's bias. A developmental biologist will consider homology in a different light than an evolutionary ecologist simply because the study of life spans numerous levels of organization.

Biological homology, the broadest of subfield definitions, seeks to understand the essence and cause of preserving a specific character across time, with the intention of discovering why homology is generally favored (Wake, 2003). This definition can be subdivided into that of the evolutionary and phylogenetic systematics specifications;

while evolutionary biologists consider two elements homologous if they are connected by a transformation series of *intermediate* homologues, phylogenetic systematic biologists consider homology as the principle defining characteristic of taxons (another term for phylogeny-based organism classifications). In comparative anatomy and physiology, characters are deemed homologous only if their relative position in a bauplan share similar structural, connective, and/or topological/histological similarities (regardless of a potential phylogenetic linkage). When considering the divergent definitions of homology in these subfields, it becomes apparent that the term is often defined by the general agenda of a group's study methodologies. Evolutionary biologists are primarily concerned with patterns of spatiotemporal speciation, phylogeneticists aim to discover the connections between current (rather than extant) species, and anatomists are concerned with patterns of body architecture (Brigandt, 2003).

Perhaps unsurprisingly, the task of defining homology at the molecular level appears to be just as difficult as that of the organismic level. Molecular biologists traditionally focus on the homology between molecules across taxa with the intention of comparing genetic sequence similarities and protein architecture. This perspective does not imply an evolutionary connection; rather, examples of molecular homology often consider the origin of entire molecular pathways as key units of evolutionary progress (Brigandt, 2003). A common criticism of the molecular approach to defining homology cites that some homologous molecules arise from independently-evolved sister taxa (termed "lineage splitting). Furthermore, paralogy (the process of gene duplication) can give rise to a "pseudo-homologous" network of proteins that in actuality are just

replicates of the same protein template. These counter arguments have helped formulate a more common definition of molecular/developmental homology, which focuses on the genetic and/or developmental basis of shared characters among species. Instead of simply focusing on gene families across an organism's entire life cycle, the zootype model stipulates that all species share a master genetic ground plan (dating, of course, to Urbilateria) that provides the basis in which homologues may form and evolve, irrespective of their mechanism of origin (Wake, 2003).

Regardless of these differences among subfields, there are five general levels of homology that have been consistently observed in paleontological and molecular records of speciation. The first, called historical homology, groups like characters based on evidence of their known common origin and ancestry. A canonical example of such a character is that of the notochord, a transient, embryonic midline structure that assists in ventral fate specification and vertebral development within all chordates (Stemple, 2005). These kinds of characters, which often have roots in earlier phylogenetic branches, tend to be more "dramatic" examples of the anatomical similarities in animal development. Certain subfields (namely that of evolutionary biology) consider this to be the only "true" level of homology; however, this rather limited view of homology fails to consider that all species are inherently linked by a common ancestor, even if this organism existed hundreds of millions of years ago.

To explain these instances of morphological sameness among related taxa that do not share an ancestor with the same trait (and thus do not fit into the historical category of homology), biologists have defined what is known as underlying/latent homology. This

concept, also termed “parallelism,” is often cited as the “reawakening” of homologous developmental mechanisms that simply were silenced in the common ancestor. A fascinating example of this phenomenon is found in the parallel evolution of anaerobic fermentation in various species of yeast; the Crabtree effect, which considers why yeast favor anaerobic fermentation in the presence of abundant oxygen, has been found to have evolved independently in two common yeast lineages through the loss of a genomic regulatory element implicated in respiration (Rozpędowska et al., 2011).

The third (and newest) level of homology, called molecular homology, is exclusively concerned with the comparison of protein sequences and architectures among species (Rutishauser & Moline, 2005). Recently, molecular biologists have realized that sequence homology is not sufficient for uncovering all homologous networks; rather, similar sequences can in fact elicit different phenotypes in species that are phylogenetically disparate (Sansom & Brandon, 2007). The task of parsing molecular homologues and relating them to their anatomical outcomes is undeniably arduous.

A level of homology that continues to be hotly debated among biologists is that of serial homology. This form of homology considers the origin and diversification of iterated parts in a single organism over time. An example of this form of homology is found in the two sets of limb appendages that are common to many animal species. Also called “iterative homology” or “homonomy,” this topic is sometimes excluded from discussions of homology because it compares the whole body and its repeating parts, as opposed to two organisms or species separately. Without consideration of this form of homology, however, developmental biologists would be unable to understand how

similar molecular bauplans give rise to repeated elements among species. A crucial example of this is found in the consideration of vertebrate somites, which are blocks of mesoderm that develop in tandem alongside the neural tube and give rise to the basic elements of the vertebral bauplan (such as the dermis, skeletal muscle/tissue, and the vertebrae) (DeRuiter, 2010).

From a developmental perspective, serial homology is absolutely essential to "evo-devo"—since a whole organism is built upon its parts, the development of these parts must be related to the restrictive governing of the whole. In fact, some of the earliest discoveries in developmental homology were linked to instances of serial homology in model organism systems, such as that of the *hox* genes and body axis morphology among bilateral species. These transcription factors are conserved throughout the animal phylogenetic tree and are responsible for specifying the anterior-posterior body axis (Shubin, Tabin, & Carroll, 2009).

The *hox* genes, a canonical example of the functional consequences of molecular homology, are also an example of the fifth and final category of homology—that of deep homology. Deep homology compares divergent taxa that have conserved molecular and developmental mechanisms yet do not have homologous corresponding phenotypes. This is where Urbilateria and the evo-devo toolkit become relevant; comparisons of species at opposite ends of the phylogenetic tree, such as arthropods versus vertebrates, are nearly impossible unless one considers that Urbilateria's genetic regulatory apparatus has left an important handprint on each of our genomes. Perhaps the most fundamental question of evolutionary development thus asks—do morphological homologies have the same

underlying molecular genetic machinery, given one approaches the problem from the right level of perspective (Rustishauser & Moline, 2005; Shubin, Tabin, & Carroll, 2009)?

While each of the subfields certainly provide valuable insights into the nuances of homologous features, it is the perspective afforded by developmental biologists that most thoroughly integrates the organismic *and* molecular influences on evolutionary homology. Phenotype and morphology alone are not enough to reveal the mechanisms behind character conservation; rather, comparison of the developmental networks employed by vastly divergent species hints at the presence of an ancient regulatory circuit that, given the right perspective, makes everything homologous (Shubin, Tabin, & Carroll, 2009). Instead of producing developmental networks *de novo* at each junction in the phylogenetic tree, evolution seems to have optimized those networks that simply "work." Yet again, it appears as if homology is entirely dependent on perspective.

2.6: Homology vs. Homoplasy—Is There Actually a Difference?

Another popular and controversial evolutionary term that tends to complicate the definition of homology across subfields is that of "homoplasy." Put broadly, homoplastic characters are morphological and/or molecular phenotypes shared by two species that are absent in their most recent common ancestor. Traditional evolutionary biologists prefer to classify homoplasy as "false homology" because, from a phylogenetic perspective, the derived similarity is not the result of obvious and immediate common ancestry (Wake, 2003). A canonical example of homoplasy in the animal kingdom compares the

appearance of wings in bird and bat species; considering that the most recent common ancestor of these species did not have wing-like appendages, this indicates that the homology witnessed between these species must have been attained via different phylogenetic routes (Shubin, Tabin, & Carroll, 2009).

The evolution of novel appendages, such as that of bird and bat wings, can be categorized into three general evolutionary pathways—that of divergent evolution, parallel evolution, and convergent evolution. Divergent evolution occurs when different traits can be traced back to the same source, such as the tracing of human arms and bat wings to a common mammalian ancestor (Shubin, Tabin, & Carroll, 2009). Parallel evolution occurs when the same trait evolves independently in different species and can be traced back to a distinct ancestral trait; a great example of this is found in the comparison of similar adaptive traits in marsupials and placental mammals, even though the evolutionary environments of these species were vastly different (Hall, 2003). Finally, as the direct antithesis to divergent evolution, convergent evolution occurs when similar traits evolve and yet cannot be traced back to a common ancestor. The most dramatic example of convergence implicates, once again, the evolution of wings in the animal kingdom; even though birds, bats, and insects diverged a remarkably long time ago, it appears as if wings were advantageous enough that evolution decided to favor their selection more than once (Shubin, Tabin, & Carroll, 2009).

There are also documented instances of traits having evolved via convergent evolution in species that share a not-so-distant ancestor. In Old World monkeys, trichromatic color vision is dependent on the presence and expression of multiple opsin

genes, whereas most New World monkeys have only one opsin gene. Due to these differences, new world males are dichromatic and only females with a different copy of the opsin gene on the X chromosome are trichromatic. Interestingly, howler monkeys (new world) support trichromatic vision in both sexes, indicating that gene duplication in this species evolved in the same manner as that of the Old World monkeys. Unlike their New World counterparts, trichromatic vision in howler monkeys appears to have evolved via convergent evolution with that of Old World primates, such as macaques, baboons, and mandrills (Jacobs, Neitz, Deegan, & Neitz, 1996).

If certain morphological characteristics witnessed across the phylogenetic tree tend to reappear across evolutionary time, even without a common ancestor that had the trait in question, then where did the trait originally come from? Do convergent characters arrive *de novo* with each branching of the tree, or is there something deeply homologous, especially at the molecular level, that allows such traits to "appear" in later, further adapted progeny species? Should we look far back enough, it may be that parallel and convergent characters arise from the scaffolding of deeply homologous sequences that date back to the original animal genome. When considered from the perspective of homology as opposed to homoplasy, characters such as these may be examples of "latent homology" in which previously silenced (but ever-present) developmental programs become active once more (Hall 2003).

What does homoplastic evolution look like from a molecular and genetic perspective, as compared to that of the more canonical, anatomical examples? Molecular parallelism tends to occur via the accumulation of small changes in homologous genes

over long stretches of time, such as through single amino acid mutations (Stern & Orgogozo, 2009). When considered in conjunction with the phenotypic outcomes of such evolutionary events, parallelism implicates Urbilateria's deeply homologous developmental network as a potential evolutionary constraint (DeRobertis, 2008). It is possible that the regulatory mechanism "template" that originated in Urbilateria's genome is what allows parallel characteristics to emerge at different evolutionary times, even if the mechanism remains "dormant" in one species for a longer period than another (Shubin, Tabin & Carroll, 2009). This logic holds true when considering the molecular origins of convergent characteristics as well—we expect convergent characteristics to have different underlying development mechanisms *unless* we can determine that the origin of said "mystery" characteristic was actually contained in the ancestral genome long before it was outwardly expressed (Hall, 2003).

A notable example of the relationship between the evolution of similar phenotypes in dramatically different species and their corresponding molecular morphologies considers the evolution of echolocation in bats and whale species. Echolocation, which is the ability to locate a target object through the use of reflective sound waves, is made possible by microscopic cilia (hair-like structures) found in the inner ear. Cilia are known to predate all mammals, and yet a certain "version" of the prestin protein, which is critical for sensitive hearing, is unique to the bat and whale lineages. When compared to other organisms that produce prestin protein and do not have the ability to echolocate, the protein sequences of bat and whale species share a unique structure simply because they evolved the same protein function (Venema, 2013).

Examples such as this in the animal kingdom open up the possibility that, given the right genetic predisposition *and* the right environmental pressures, speciation at the molecular level may be entirely dependent on the broader principles of homology.

2.7: Descent with Modification—a Molecular Consideration of Evolution

While it is true that many of our claims about the homologous pathways contained in Urbilateria's genome involve conjecture, discoveries in comparative genomics have indicated that, at the molecular level, homologous proteins provide a framework with which we may sequentially reconstruct the "original" animal genome. It is shocking that we share so much of our genomes with that of other species; for example, humans and fruit flies share sixty percent of their coding genes, even though their anatomy and physiology would lend one to believe that such is not a possibility! It is foolish to think that evolution carefully hand-crafted each and every animal genome at all of the phylogenetic junctions; instead, deeply homologous networks and their resulting morphologies have supported independent innovation across evolutionary time. While millions of years of evolution cannot be broken down to a single biological pattern, it would seem as if the perpetual persistence of deeply homologous networks has attested to their utility and robustness in supporting large-scale speciation (Hall, 2012).

Should evolution thus be characterized as the collaboration of deliberate, careful patterns of craftsmanship, or instead as a "hands-off, trial-and-error" set of environmental innovations? The latter model of evolution, called "Descent with Modification" by Charles Darwin, stipulates that speciation is the result of subtle modifications within the

confines of environmental and selective pressures (Shubin, Tabin, & Carroll, 2009; Stern & Orgonzo, 2009). This perception of evolution is often contrasted with an image of selection as the result of "bursts" of entirely *de novo* innovation. At the molecular level, it is probable that exposure to similar environmental cues has allowed already diverged species to become "activated" in a specific regulatory network across evolutionary time, even when these regulatory networks have lain phenotypically dormant. From this framework, phenomena like homoplastic characters become nothing more than nuanced examples of biological homology (Hall, 2003). Taking such a broad perspective allows one to see that quiescent molecular phenotypes have the potential to give rise to homologous organismic phenotypes, given that the right environmental pressures are exploited. This concept also aligns with the theory of punctuated equilibrium, which considers periods of stasis along the phylogenetic tree to be of evolutionary relevance. Speciation may have occurred through sudden "bursts" at specific spots in the fossil record simply because the right environmental pressures at that geological time activated favorable, latent pathways (Gould & Eldridge, 1993).

While our understanding of evolutionary homology is certainly limited because multicellular organisms have been forced to retain the same basic "building blocks" in order to successfully evolve, this does not detract from the notion that homologous characters have been favored because of their adaptability and favorability across time. Each time a new phenotype arises in the timeline of evolution, it is considered an evolutionary novelty; however, it is not until that novelty transitions into homology that it can be truly considered a favorable trait (Wagner, 2015).

2.8: The Clues Found in Developmental Homology

Along with ascribing to a more robust definition of homology, the field of evolutionary development is unique in that it focuses on those molecular pathways that are exceptionally common among members of the animal kingdom. Comparative studies of animal anatomy during development indicate that phenotypic divergence is more or less linearly correlated with time—at the beginning of development, embryos from different species are almost entirely indistinguishable. While phenotypic and molecular divergence among animal species becomes markedly more pronounced after development, the "reawakening" of developmental pathways later in life (whether through normal or abhorrent mechanisms) provokes evolutionary biologists to consider how developmentally homologous networks can impact adult homeostasis. Evo-devo uniquely explains how structures can emerge in ontogeny (development), why they appear and function as they are, as well as why and how they have been conserved and transformed through phylogeny by integrating molecular and anatomical data (Brigandt, 2003; Hall, 2012). In the quest to distinguish real versus superficial homology among comparative species, evo-devo takes a unique approach by formulating a more concrete picture of homology; instead of speculating about the origin of homologous characters, evolutionary developmental biologists have the opportunity to inspect all levels of animal organization such that the underlying mechanisms of our "sameness" become readily apparent (Wagner, 2015).

Before considering the molecular mechanisms in which protein evolution has shaped the homology of animal development, it is important to ask how evo-devo

grapples with the problem of integrating molecular and organismic data on homologous features. There is no denying that the molecular and anatomical definitions for homology are often difficult to translate, leading to subfield-specific definitions that exclude broader criteria; however, one could also argue that it is the non-biased exploration of how molecular phenotypes directly create anatomical phenotypes during development that sets evo-devo's definition of homology apart. Evolutionary development is also uniquely able to assess the role of intra-organismic homologies (such as that of serial homology) within the broader scope of evolutionary development. Through the integration of molecular and anatomical data on homologous characters, evolutionary developmental biologists are able to uniquely construct wholistic depictions of the plausible constraints that led to the diversification of Urbilateria's genome, a topic which will be explored in the next chapter (Brigandt, 2003; Sansom & Brandon, 2007).

2.9: Chapter Two References

Amundson, R. (2005). *The changing role of the embryo in evolutionary thought: roots of evo-devo*. Cambridge University Press.

Balavoine, G., & Adoutte, A. (2003). The segmented Urbilateria: a testable scenario. *Integrative and Comparative Biology*, *43*(1), 137-147.

Brigandt, I. (2003). Homology in comparative, molecular, and evolutionary developmental biology: the radiation of a concept. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, *299*(1), 9-17.

Davidson, E. H., & Erwin, D. H. (2006). Gene regulatory networks and the evolution of animal body plans. *Science*, *311*(5762), 796-800.

De Robertis, E. M. (2008). Evo-devo: variations on ancestral themes. *Cell*, *132*(2), 185-195.

DeRuiter, C. (2010). Somites: Formation and Role in Developing the Body Plan. *The Embryo Project Encyclopedia.* Retrieved from https://embryo.asu.edu/pages/somites-formation-and-role-developing-body-plan.

Gehring, W. J. (1998). *Master control genes in development and evolution: The homeobox story*. Yale University Press.

Gilbert, S. F., Opitz, J. M., & Raff, R. A. (1996). Resynthesizing evolutionary and developmental biology. *Developmental biology*, *173*(2), 357-372.

Gould, S., & Eldredge, N. (1993). Punctuated equilibrium comes of age. *Nature*, *366*(6452), 223.

Hall, B. K. (2003). Descent with modification: the unity underlying homology and homoplasy as seen through an analysis of development and evolution. *Biological Reviews*, *78*(3), 409-433.

Hall, B. K. (2012). Parallelism, deep homology, and evo-devo. *Evolution & development*, *14*(1), 29-33.

Jacobs, G. H., Neitz, M., Deegan, J. F., & Neitz, J. (1996). Trichromatic colour vision in New World monkeys. *Nature*, *382*(6587), 156.

Kirschner, M., & Gerhart, J. C. (2006). *The plausibility of life: Resolving Darwin's dilemma*. Yale University Press.

Knoll, A. H., & Carroll, S. B. (1999). Early animal evolution: emerging views from comparative biology and geology. *science*, *284*(5423), 2129-2137.

Malpas, J. (2012). Developmental Biology. *The Stanford Encyclopedia of Philosophy*. Retrieved from https://plato.stanford.edu/entries/biology-developmental/#HisCon

Ohta, T. (1992). The nearly neutral theory of molecular evolution. *Annual review of ecology and systematics*, *23*(1), 263-286.

Richardons, M.K., & Keuck, G. (2002). Haeckel's ABC of evolution and development. *Biological Reviews, 77*(4), 495-528.

Rozpędowska, E., Hellborg, L., Ishchuk, O. P., Orhan, F., Galafassi, S., Merico, A., ... & Piškur, J. (2011). Parallel evolution of the make–accumulate–consume strategy in Saccharomyces and Dekkera yeasts. *Nature communications*, *2*, 302.

Rutishauser, R., & Moline, P. (2005). Evo-devo and the search for homology ("sameness") in biological systems. *Theory in Biosciences*, *124*(2), 213-241.

Sansom, R., & Brandon, R. N. (Eds.). (2007). *Integrating evolution and development: From theory to practice*. MIT Press.

Shubin, N., & Tabin, C. (1997). Fossils, genes and the evolution of animal limbs. (Cover story). Nature, 388(6643), 639

Shubin, N., Tabin, C., & Carroll, S. (2009). Deep homology and the origins of evolutionary novelty. *Nature*, *457*(7231), 818.

Stemple, Derek L. (2005). Structure and function of the notochord: an essential organ for chordate development. *Development,* 132, 2503-2512.

Stern, D. L., & Orgogozo, V. (2009). Is genetic evolution predictable? *Science*, *323*(5915), 746-751.

Venema, D. (2013). Evolution Basics: Convergent Evolution and Deep Homology. *Biologos.* Retrieved from https://biologos.org/blogs/dennis-venema-letters-to-the-duchess/evolution-basics-convergent-evolution-and-deep-homology.

Wagner G.P. (2015) Reinventing the Organism: Evolvability and Homology in Post-Dahlem Evolutionary Biology. In: Love A. (eds) Conceptual Change in Biology. Boston Studies in the Philosophy and History of Science, vol 307. Springer, Dordrecht

Wake, D. B. (1991). Homoplasy: the result of natural selection, or evidence of design limitations?. *The American Naturalist*, *138*(3), 543-567.

Wake, D. B., & WAKE, D. B. (2003). Homology and homoplasy. In B. K. Hall, & W. M. Olson (Eds.), *Keywords and concepts in evolutionary developmental biology*. Cambridge, MA: Harvard University Press. Retrieved from http://dml.regis.edu/login?url=https://search.credoreference.com/content/entry/hupedb/homology_and_homoplasy/0?institutionId=2022

Chapter 3

THE HOMOLOGOUS PATTERNS OF DEVELOPMENT

3.1: Introduction

Developmental biologists are uniquely primed for studies concerning homology—in the quest to understand how form meets function, embryologists must be intimately familiar with the mechanisms in which molecular processes translate into organismic phenotypes. We have already seen how different perspectives of biological inquiry may give rise to varied definitions of homologous characters; unlike other fields of the life sciences, however, evo-devo biologists possess the most impartial tools for understanding the many levels of animal homology within the context of evolutionary novelty. With a concept so open to interpretation, it would seem as if studying homology at any level of animal organization (let alone across all of them) would be a tremendous feat—how, then, do evo-devo biologists do it?

3.2: The Time-Honored Exploitation of Genetic Material

In order to understand the vantage point from which evo-devo biologists compare the spatiotemporal patterns that govern animal development, it is important to understand the general facets of embryogenesis that are shared by all species (Hoekstra & Coyne,

2007). There is no question that development on the whole is intricately complex at both the molecular and anatomical levels; however, as organ systems become established, the molecular mechanisms that govern animal embryogenesis tend to become more nuanced and species-specific. Regulatory processes at the onset of development are often comparatively less complex and have a more dramatic pleiotropic effect on phenotype because they establish the basis of bauplan architecture—thus, the beginning and middle stages of development are where we find the majority of homologous molecular networks (Davidson & Erwin, 2006). These patterns, coupled with the claim that changes in genomic regulation act as key drivers of evolution, potentially explain how homologous characters develop on a molecular scale (Hoesktra & Coyne, 2007).

It would stand to reason that a species with a larger genome should be more phenotypically complex than that of a small genome. While this pattern is certainly evident across the larger tree of life, there are numerous instances in which genome size does not positively correlate with phenotypic complexity (Levine & Tijan, 2003). Considering that the biochemical signaling cascades important to cell functionality tend to be rather large and complex, small changes in an organism's genome often lead to exponentially large phenotypic consequences (Hoekstra & Coyne, 2007). This relationship between an organism's genome and biochemistry further implicates the role of regulatory elements in evolutionary novelty. A telling example of how animal complexity has been achieved through changes in genomic regulation lies in a comparison of the nematode *Caenorhabditis elegans* and the common fruit fly (*Drosophila melanogaster)*; *C. elegans,* a phenotypically simple animal, has a genome

comprised of roughly 20,000 genes, while that of *D. melanogaster* (a more complex organism) features only 14,000 genes. Furthermore, the human genome contains approximately 20,000 genes, and yet there is little question as to whether or not we feature more dynamic tissue systems than that of a worm (Levine & Tijan, 2003)!

If increasing phenotypic complexity through speciation is not always positively correlated with the size of an organism's protein-coding genome, then how has evolution "exploited" such a small number of genes, many of which can be traced back to Urbilateria's genome? Previously, two mechanisms of diversification have been proposed by molecular biologists—that of alternative splicing and DNA rearrangement. In the process of creating mRNA, the molecule that directs protein synthesis in the cell, a single stretch of genomic material has the potential to code for numerous proteins via the mechanism of alternative splicing, thus allowing for a smaller genome to code for a larger proteome (Levine & Tijan, 2003). A rather dramatic example of this phenomena is found in the *D. melanogaster* genome, which features a gene called *dscam* that directs neuron pathfinding and codes for roughly 18,000 different proteins in a single stretch of base pairs (Kashyap & Tripathi, 2008). The other proposed mechanism, simply described as "DNA rearrangement," encompasses those mutational events in which various parts of the genome rearrange in sequential order and frequency such that spatiotemporal expression levels of various proteins becomes more complex over evolutionary time.

Increases in physiological and behavioral complexity along the phylogenetic tree have been largely accredited to evolutionary changes in the expression of genes, rather than changes in the number of protein-coding genes themselves. The mechanisms that

describe many of these changes are cis-regulatory sequences and the diversification of multi-protein transcription complexes, two processes that will be explored in depth later in this chapter (Levine & Tijan, 2003). When examining the evolutionary history of complex developmental traits, it is important to keep in mind that the appearance of favorable characters is largely determined by what is possible at the molecular level, what is probable within the genomic constraints situated in an animal's genome, and what is permissible via selective environmental pressures (Carroll, 2008). It is likely that the accumulation of slight genomic alterations over evolutionary time has given rise to the many distinct phenotypic traits associated with extant species on earth today (Stern & Orgogozo, 2009).

3.3: The Structure of the Protein-Coding Genome

Before diving into the molecular mechanisms that have given rise to diversification throughout the animal kingdom, it is important to understand how the genome is organized and regulated. Only two percent of the metazoan genome (multicellular, eukaryotic organisms) features protein-coding sequences; besides those conserved repeat sequences that have unknown functions, the rest of the genome most likely contains information for the many types of regulatory mechanisms characteristic of higher organisms. It is estimated that a *third* of the animal genome codes for cis-regulatory sequences that direct the binding of regulatory molecules implicated in spatiotemporal gene expression. These sequences are connected to the larger processes of chromosome condensation, pairing, segregation, and expression patterning. In higher

organisms, each protein-coding gene contains ten kilo-base pairs of DNA on average, with only a fraction of these nucleotides comprising the gene itself. *Ascidiacea,* a phylogenetic class that includes simple sea squirts, have an estimated ten to twenty-thousand tissue-specific regulatory sequences called enhancers, even though they are phenotypically simple organisms (Levine & Tijan, 2003).

Out of the two percent of the metazoan protein-coding genome, five to ten percent of coding capacity is dedicated to producing proteins that directly regulate transcription. These proteins include many classes of sequence-specific DNA binding proteins, such as transcription factors, large multi-protein RNA polymerase chain complexes, and countless chromatin remodeling complexes. Protein-coding genes are traditionally organized into four parts—first, a sequence called an enhancer flanks the front of the gene (a cis-regulatory element, may also be found after the gene body) and is responsible for binding transcription factors; next lies the promoter, a sequence which, upon upstream binding of transcription factors, supports the binding of RNA-polymerase and the activation of transcription; third lies the sequence that directly codes for the protein in question; and finally, there is a termination in the coding sequence and the potential for another cis-regulatory element called a silencer, which can bind to other regulatory molecules and conversely downregulate transcription of the gene in question. Another cis-regulatory element that adds a layer of complexity to genome expression is that of an insulator; while the relative occurrence of insulators in the human genome remains

unknown, insulators are known to inhibit the binding of enhancers and silencers through spatial constriction of the 3D architecture of DNA (Raab & Kamakaka, 2010).

Figure 3.1. Schematic of the general architecture of a protein-coding gene. Regulatory sequences upstream and downstream of the transcription start site (depicted in yellow) are locations of transcription factor binding, which leads to differential patterns of expression. The gene body, depicted in red, features the sequence of nucleotides responsible for the eventual production of the protein in question. Adapted from Shafee & Lowe, 2017.

Most animal genes have multiple enhancers that are located before or after protein coding sequences. Sometimes found within the introns (parts of the gene that are excised after transcription) of a gene itself, they help to direct spatiotemporal patterns of transcription without directly altering the mRNA's code. Normally, an enhancer is dedicated to the direct activation of a gene in a tissue-specific manner such that the expression of a protein may look dramatically different in one tissue system as compared to another. Enhancers often have multiple sites for transcription factor binding and sometimes couple these regulatory activities with the use of multiple promoter sequences (Buecker & Wysocka, 2012). In the early stages of development, embryos must transition from utilizing mom's "pre-loaded" transcription factors and signaling molecules to those that it assembles itself; compared to other protein-coding sequences, these genes

disproportionately rely on more than one promoter. The complexity of these regulatory networks is made even more pronounced when multi-subunit cofactor complexes are considered, as these conglomerates of proteins require the binding of multiple proteins to invoke silencing and/or transcriptional enhancement activity. Upon considering these factors, it appears as if phylogenetic complexity is directly related to transcriptional plasticity; while the yeast genome, the simplest of eukaryotic organisms, contains approximately three hundred sequences coding for transcription factors, the genomes of *C. elegans* and *D. melanogaster* contain over one thousand transcription factors (Levine & Tijan, 2003).

When comparing species-specific exomes, which include only protein-coding sequences and ignore cis-regulatory elements, it appears as if proteins have remained remarkably homologous over evolutionary time. This has allowed molecular and evolutionary biologists to track the evolution of protein "families" and their comparative phenotypes across species. While molecular function seems to be relatively consistent along the phylogenetic tree, phenotypic function can vary dramatically. The disruption of developmental genes often leads to the appearance of unique phenotypes that, without molecular comparisons, appear to be entirely divergent. It is possible that upon knockout of specific genes, phenotypic outcomes may vary simply because each species supports its own cell signaling networks with proteins that are implicated in unique biochemical "steps" of a network. A current dilemma in evo-devo lies in the need for laboratory methodologies that concretely recognize equivalent molecular phenotypes between different model organisms, even if these phenotypes may not share the same anatomical

outcomes. These phenotypes, deemed "phenologs," are often related by the orthology (spatiotemporal development) of associated genes in two different organisms, suggesting that evolutionarily conserved characters can sometimes convergently arise through different disruptions of the *same* set of conserved genes. These challenges recapitulate the need in evo-devo studies to reconcile examples of homology at both the molecular and organismic levels (McGary et al., 2010).

More recently, studies concerning comparative protein domains and sequence conservation have indicated that homology at the molecular level does not always translate to shared phenotypic characters (and vice versa). A protein domain is a conserved stretch of amino acids that is shared among proteins in similar families and classes. Canonical, well-known examples of protein domains that are shared among functionally-similar proteins include kinase domains and cytoplasmic domains, only two groups of over fifty-three million unique examples in the animal kingdom. Domain-specific studies are often used as tools in modern genomics and developmental biology because they discretely identify homologous functions between two proteins. It is not surprising that proteins with the same domain composition are likely to be ancestrally related; however, like almost every other example of molecular homology, there are instances in which some comparative proteins feature similar sequences with different domains (and vice versa). The study of protein evolution, just like the study of evo-devo, is *not* clear-cut, and even though it is hypothesized that close to all of the protein families contained in the animal genome today originated from a small number of domains, there

is still much work to do in uncovering just how similar our protein-coding genome is to that of Urbilateria's (Caetono-Anolles, 2011).

3.4: The Mutations that Altered Ubilateria's Genome

When it comes to understanding those parts of the animal genome that have been selectively altered over the course of evolutionary time, it is important to understand that conserved mutations are not always equally distributed. Even though each mutational event is itself random, segments of the genome that are repeatedly and preferentially mutated, termed "hotspot genes," disproportionately belong in the category of developmental regulatory proteins. Of this group of protein-coding segments, those that are deeply homologous (i.e. deep conservation across species) tend to be patterning genes that dictate development in more than one cell type. It is perhaps no surprise, then, that many of our canonical examples of deeply homologous sequences contain families of transcription factors that function within large protein networks in the earliest stages of animal development (Stern & Orgogozo, 2009).

If sections of an animal's exome are thus preferentially targeted for evolutionary innovation, are there certain types of mutational events that are also under positive selection? The five types of mutations that most likely gave rise to the alteration and diversification of Urbilateria's genome include that of structural mutations, gene expansions, gene duplications, gene deletions, and cis-regulatory mutations. The first of these includes all alterations of the genome in which the order of specific protein-coding segments is changed. While these mutations do not affect the protein's functionality per

se, they do alter the spatiotemporal expression of that protein in relation to its neighbors. This type of mutation becomes important when a protein is translocated and placed on a non-synonymous chromosome; in this novel location, the spatiotemporal transcription of the protein may become differentially upregulated or downregulated based on the expression patterns of its new neighbors (DeRobertis, 2008).

A similar phenotypic effect to structural mutations is found in gene duplications and expansions. In duplication events (and in its large-scale version, dubbed gene expansions), the emergence of multiple copies of a protein allows the subsequent mutation and innovation of one copy without loss of the original sequence (and thus its original functionality). Duplications and expansions are often favored in large families of transcription factors such that many developmental regulatory networks may contain some degree of functional overlap. In the sea urchin genome, for example, there are hundreds of duplicated protein-coding sequences for Toll-like receptors and leucine-rich proteins that helped support the diversification of autoimmunity across species. In most developmental cases, duplications and expansions have either allowed the tandem duplication of protein exons or have supported increased opportunity for alternative splicing (DeRobertis, 2008).

The direct antithesis to gene duplications and expansions is that of gene deletions. In regard to evolutionary innovation, gene deletions are often utilized as molecular adaptations to new ecological niches through immediate loss of protein function. This type of mutational event is comparatively dramatic—from a developmental perspective, these mutations can lead to the loss of whole organs, such as eye degradation and

subsequent development of albinism in cave-dwelling salamanders. A rather extreme and fascinating example of the loss of an entire developmental system through gene deletion is found in the disappearance of segmentation patterns in nematode and planarian flatworm species, even though their common ancestor was undoubtedly a segmented organism (DeRobertis, 2008).

The final type of mutational event that led to the diversification of Urbilateria's genome was that of cis-regulatory mutations. In the spirit of Neo-Darwinian natural selection, these mutations are likely to have had a dramatic impact on animal diversification and innovation across evolution because they can support the subtle compounding of small regulatory alterations over time. Unlike structural mutations, duplications, expansions, and deletions, cis-regulatory mutations do not require novel mutations in coding regions to be "tested out" for positive functionality. Due to their role in regulation, a single-point mutation in a cis regulatory region (CRE) has the capacity to single-handedly impact the spatiotemporal expression of unrelated proteins in numerous divergent regulatory networks. CREs provide the ultimate platform for evolutionary innovation—by mutating a single CRE, the genome is able to alter its ability to bind a target transcription factor, which has the potential to exponentially change the regulation of an entire cascade of proteins without directly affecting any of these proteins, including that of the original transcription factor (DeRobertis, 2008; Levine & Tijan, 2003).

3.5: The Lewis Model & Hox Genes—How Duplication Diversified the Metazoan Body Plans

Perhaps the most remarkable step in the evolution of species on earth was that of the emergence of multicellular organization. In order to step out of the limitations of unicellularity, molecular innovation needed to acquire the capacity to support cell-type divergence and coordination. Much of our understanding of the evolutionary development of metazoans depends on the study of sponges, the most ancient and primitive multicellular organisms still on earth today (Coutinho et al., 2003). For a body part to develop, the domain (location) of that part must be specified at the molecular level, followed by the morphological patterning that gives rise to such a phenotype. Comparative studies have indicated that the coordinated development of body plans among all animal bauplans dates back to regulatory mechanisms contained in Urbilateria's genome (Davidson & Erwin, 2006).

Perhaps the most canonical example of how gene duplications have driven the homologous patterns of animal development lies in the *Homeobox* genes, a group of proteins that direct body axis development (Akam, 1989). Studies concerning the *Homeobox* genes largely propelled the field of developmental biology into the spotlight of biomedical inquiry. *Hox* genes were first discovered as master control genes in *D. melanogaster* and were later identified in other vertebrate model systems (Carroll, 2008). As developmental directors of cell fate, proliferation, and diversification, *Hox* genes act as transcription factors that preferentially bind to target cis-regulatory elements and initiate developmental cascades related to body segmentation and limb development.

Even sponges, which retain a relatively amorphous bauplan, have genomes that contain the *Hox* genes *efh-1* and *prox2*, indicating that these sequences are deeply homologous (Coutinho et al., 2003; Akam, 1989).

If *Hox* genes were present in Urbilateria's genome, then through which mutational mechanisms did they help diversify the animal kingdom? A comparison of *Hox* copy number in the genomes of *D. melanogaster* (fruit fly)*, D. rerio* (zebrafish), and *M. musculus* (mouse) indicate that these homeotic genes were diversified through duplication and subsequent diversification events. Contrary to cis-regulatory mutations, *Hox* genes serve as an example of the Lewis Model, which contends that genetic innovation across evolutionary time has been principally achieved through the employment of duplication events (Gehring, 1998). Besides this increase in copy number, which has led to the emergence of different *Hox* frequencies across the phylogenetic tree, *Hox* transcription factors are also unique in their spatiotemporal expression during development. *Hox* genes in all model organism systems are expressed in a reverse collinear relationship called the "*Hox* clock"—genes that are expressed earlier in development and are at the 3' chromosomal position are related to anterior axis specification, while those that are expressed later in development and at 5' chromosomal positions direct posterior processes (Shubin, Tabin, & Carroll, 2009; Akram, 1989). This order of expression relates to the temporal and spatial coordination of stem cell differentiation in developing embryos, as the collinearity of the *Hox* genes allows *Wnt* signaling to induce axial progenitor cells to direct spatial coordinates (Deschamps & Duboule, 2017). By duplicating and then "tweaking" these transcription factors along the

phylogenetic tree, it seems as if evolution has allowed for the slow and steady diversification of body axis patterning within the confines of selective constraints.

Besides the duplication and diversification of individual *Hox* genes within species, comparison of the mouse and fruit fly genomes (representing that of invertebrates and vertebrates) indicate that vertebrates also experienced whole group duplications of these genes. The groups, called "*Hox* clusters," also seem to follow a linear pattern of regulation—entire clusters are often co-expressed in a 3'-5' fashion along their location on a given chromosome. The regulatory use of these clusters in tandem with the collinear pattern of individual gene expression provides a narrative for the developmental diversification of vertebrates, as this additional layer of regulatory complexity follows species along the phylogenetic tree. Regardless of regulatory differences, perhaps the most surprising characteristic of *Hox* genes is that they exhibit acute sequence similarities, regardless of species diversification. Later in development, comparative roles of the genes across species indicate that they support analogous (rather than homologous) phenotypes at the organismic level (Akam, 1989). It is absolutely remarkable that unique combinations of sequence conservation, individual gene duplications, *and* cluster duplications were able to give rise to such a robust family of deeply homologous transcription factors; *Hox* genes, like many other examples of evolutionary novelty in development, further implicate Darwin's notion that evolution functions via descent with modification.

3.6: The Homologous Regulatory Patterns of Evo-Devo

While the single-protein mutational perspective certainly sheds light on the patterns of selective pressure that have guided speciation at the molecular level, the study of evolutionary animal development would not be complete without a consideration of the selection that has dictated which cellular protein networks have stood the test of time. Homology is not solely observable from the microscopic view of singular genes and their coded proteins; rather, it appears that homologous networks are also often found in examples of the comparative cellular regulatory machinery that direct animal development. Pleiotropy, ancestral genetic complexity, functional equivalence, heterotropy, and cis-regulatory elements are the most common mechanisms in which homologous protein networks have driven animal diversification over time (Carroll, 2008).

One major challenge in uncovering the evolutionary origins of important developmental protein networks has been in answering the question of how two identical proteins are capable of giving rise to different phenotypic outcomes in comparative species. Homologous examples of pleiotropy, a term to describe instances in which a single locus affects multiple traits, are unique in evo-devo simply because the vast majority of developmental proteins participate in numerous independent pathways. The use of pleiotropic loci in development logically coincides with that of the "descent with modification" theory of evolutionary novelty—instead of repeatedly supporting a *de novo* creation of proteins to accommodate selective pressures, evolution has favored the integration of master regulatory proteins into new networks over time (Carroll, 2008). An

example of a protein that is known to give rise to vastly divergent phenotypes at the intra-organismic level is that of Sonic hedgehog—this protein is known to play a role in the networks that govern neurulation, digit polarity and number, floor plate development, and even feather bud formation (McMahon, Ingham, & Tabin, 2003). Proteins such as this are often deemed "mosaic" because they play integral roles outside of the confines of a single cell type or germ layer and function as conserved toolkit proteins. While individual proteins in the same developmental networks across species may not share direct sequence conservation, it appears as if proteins like Hedgehog act as a conserved example of the homology contained in master regulators (Carroll, 2008).

The second pattern observed among homologous developmental protein networks considers the notion of ancestral gene complexity. In this category falls the *Hox* genes; like other large developmental gene families, the *Hox* proteins are remarkably similar among vertebrates and invertebrates. The vast majority of ancestral gene families are made up of transcription factors, which further indicates that speciation at the molecular level has been shaped by small changes in the animal regulatory genome (rather than through quick and large genome changes). When considering toolkit proteins and their ancient origins, it is important to note that many of these proteins, if replaced by an orthologous protein from a vastly divergent species, are capable of "taking over" and directing normal development. Murine *Pax6,* a transcription factor that directs eye development, is able to direct normal ommatidium formation in both its host species and in *Drosophila,* indicating that the protein has not been dramatically altered since chordate divergence. Examples of functional equivalence such as this indicate that there must have

been considerable evolutionary constraints enacted upon these specific proteins over time and that much of the basic animal machinery is homologous across phenotypically diverse animal species (Carroll, 2008). It is possible that we are unable to elucidate the exact mechanisms that gave rise to these toolkit proteins simply because they predate Urbilateria itself.

Contrary to other large gene families, duplications of prominent toolkit proteins have been observed at a surprisingly low frequency on average. While it was previously thought that protein innovation is always preceded by gene duplication, studies of master regulators like the *Hox* family have found that no new *Hox* members have arisen in the mammalian lineage. Interestingly, the tetrapod lineage has actually lost some *Hox* members over time. A possible explanation as to why duplication is not a prerequisite for novelty within this subset of proteins implicates the dosage-sensitivity of most developmental processes (Carroll, 2008).

The final two patterns observed across homologous protein networks in development include that of heterotropy and cis-regulatory diversification. From a developmental context, heterotropy is a term to describe comparative spatial changes in development across evolutionarily related species. A conceptual example of this can be found in the different limb locations of mammals; these morphological changes are often not caused by different molecular players themselves, but rather are enacted by different mechanisms of spatiotemporal gene expression during key stages of development. This points to the role of cis-regulatory elements in initiating cascade-effects on protein networks, considering that cis-regulatory elements allow the accumulation of small

mutations over time without disrupting toolkit sensitivity. All toolkit transcription factors, with their plethora of molecular functions in a developing embryo, are inherently at the mercy of multiple modular cis-regulatory elements that have the potential to tweak their activity at distinct time points in development. Again, the notion of descent with modification comes into play—instead of invoking dramatic, protein-specific mutations, selection has the potential to both maintain chromatin stability *and* "tinker" with CREs (Carroll, 2008).

While these homologous regulatory patterns make conceptual sense when considered on their own, how can these collective patterns provide a holistic explanation of evolutionary development at the intermolecular level? In 2006, Davidson and Erwin constructed a machine learning algorithm to answer this exact question. They postulated that homologous master regulators that govern animal body plan development (such as the *Hox* family of transcription factors) function as "evolutionarily inflexible subcircuits," i.e. they perform essential upstream functions that are vital to viable embryonic development, while "certain small subcircuits" act as developmental signaling proteins (often not homologous across species) that are smaller "plug-ins" for networks. These smaller players, which they later titled "gene regulatory network (GRN) kernels," have been co-opted repeatedly by selection to take on diverse developmental purposes. When combined, their proposed "kernel architecture" indicates that evolutionary innovation through the diversification of animal development is governed by both the conserved master regulators at the top of the kernel and the lower plug-in proteins that are amenable to alteration over time (Davidson & Erwin, 2006).

According to their algorithm, there are distinct properties that govern whether or not a GRN kernel is evolutionarily feasible from the perspective of body plan innovation. First, their model stipulates that genes at the top of the protein cascades were almost always master transcription factors. Second, they determined that besides executing the spatial commands for the development of a given body part, the smaller kernels were dedicated to a specific temporal location during development and did not execute any other function. Finally, and perhaps most intriguingly, they found that the proteins within the kernels had to be continually transcribed for the kernel to continue functioning (a positive feedback loop). Interference with the first kernel proteins would inevitably result in the destruction of the kernel's function altogether. These combined observations led to the computational confirmation of a long-accepted concept in evo-devo—that protein networks are conserved across development to the point of being categorized as *deeply* homologous (Davidson & Erwin, 2006). Future investigations into the homologous patterns of evolutionary development through computational approaches may be key to understanding how homology has shaped animal novelty.

3.7: Are Cis-Regulatory Elements the Missing Link?

There remains much to be explored on how cis-regulatory elements (CREs) drive development and impact the preservation of homologous protein networks across evolutionary time. The prevailing hypothesis considers whether mutations in CREs can enact a genomic "snowball" effect—upon mutation, important transcription factors may differentially bind to target CRE elements and support varied patterns of gene expression

and protein translation. Interestingly, most novel CREs in extant animal genomes have been found to arise via transposition, a type of copy number variation in which a whole section of the genome is relocated to a new chromosomal location. These transposable elements are under uniform selective pressure, indicating that transposons may act as primitive CREs that change the spatiotemporal expression of a cascade of proteins. Loss-of-function mutations in CREs have also been observed to downregulate entire networks of development, further implicating the snowball effect of these regulatory elements. Perhaps the most exciting facet of CRE modification is that they have the potential to alter pathways in a variety of mechanistic directions, such that evolutionary novelty may arise through a plethora of epigenetic forms (Carroll, 2008).

With only two percent of the human genome coding for proteins, it comes as no surprise that much of cellular functionality depends on the spatiotemporal control of gene expression. Is it realistic to hypothesize that small, accumulated changes in cis regulatory elements could account for the magnitude of evolutionary innovation seen in fossil records since the emergence of Urbilateria? While there is little doubt that homologous CREs during development certainly impacted the long-term interplay between speciation and phylogenetic branching, it is not likely that they alone can account for the diversification of the animal kingdom. Instead, it is probable that structural *and* regulatory mutations have supported speciation since the Cambrian explosion. A major criticism of evolutionary development in recent years has cited that evo-devo research focuses too heavily on the homologous patterns of body plan morphology; instead, the missing link between structural and regulatory mutations may be found in the study of

homologous behaviors, metabolisms, and physiologies of comparative species. While the study of form versus function is certainly more feasible when consulting fossil records, the advent of modern molecular technologies has opened up the opportunity to consider just how much of the modern animal bauplan is dependent on the interplay of structure and regulation (Hoekstra & Coyne, 2007).

3.8: Transcription Factors & the Regulation Conundrum

As was validated in Davidson & Erwin's machine learning model on the nature of protein networks in driving homologous patterns of development, it is well-known that transcription factors play a substantial role in cell functionality and differentiation (Davidson & Erwin, 2006). Compared to other major groups of proteins, transcription factors tend to be conserved in structure and function because they are tasked with the job of maintaining protein scaffolding and timed expression. Because they interact with CREs and other regulatory enzymes, transcription factors hint at a potential mechanism in which small mutational events bifurcate and multiply differential gene expression across evolutionary time. It is foolish to assume that changes in epigenetic regulation can account for all of the important novelty in the animal phylogeny; however, through the "radical co-option of function" afforded by transcription factors and their corresponding CREs, the animal genome may well have accumulated selectively favorable mutations over time (Hoekstra & Coyne, 2007). When considered at the molecular level and beyond, these sorts of mutations give meaning to the comparative types of homology discussed in chapter two—if homology did not exist at the most basic molecular level,

biologists would most certainly struggle in finding examples today of organismic homology that inspired previous generations to coin the term.

3.9: Chapter Three References

Akam, M. (1989). Hox and HOM: homologous gene clusters in insects and vertebrates. *Cell*, *57*(3), 347-349.

Buecker, C., & Wysocka, J. (2012). Enhancers as information integration hubs in development: lessons from genomics. *Trends in Genetics*, *28*(6), 276-284.

Caetano-Anollés, G. (2011). *Evolutionary genomics and systems biology*. John Wiley & Sons.

Carroll, S. B. (2008). Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell, 134(1), 25-36.

Coutinho, C. C., Fonseca, R. N., Mansure, J. J. C., & Borojevic, R. (2003). Early steps in the evolution of multicellularity: deep structural and functional homologies among homeobox genes in sponges and higher metazoans. *Mechanisms of Development*, *120*(4), 429-440.

Davidson, E. H., & Erwin, D. H. (2006). Gene regulatory networks and the evolution of animal body plans. *Science*, *311*(5762), 796-800.

Deschamps, J., & Duboule, D. (2017). Embryonic timing, axial stem cells, chromatin dynamics, and the Hox clock. *Genes & development, 31*(14), 1406-1416.

De Robertis, E. M. (2008). Evo-devo: variations on ancestral themes. *Cell*, *132*(2), 185-195.

Gehring, W. J. (1998). *Master control genes in development and evolution: The homeobox story*. Yale University Press.

Hoekstra, H. E., & Coyne, J. A. (2007). The locus of evolution: evo devo and the genetics of adaptation. *Evolution: International Journal of Organic Evolution*, *61*(5), 995-1016.

Kashyap, L. & Tripathi, P. (2008). Alternative splicing: how one gene can make many proteins. *Bioscience,* 4(1), n.p.

Levine, M., & Tjian, R. (2003). Transcription regulation and animal diversity. *Nature*, *424*(6945), 147.

McGary, K. L., Park, T. J., Woods, J. O., Cha, H. J., Wallingford, J. B., & Marcotte, E. M. (2010). Systematic discovery of nonobvious human disease models through orthologous phenotypes. Proceedings of the National Academy of Sciences, 107(14), 6544-6549.

McMahon, A.P., Ingham, P.W., and Tabin, C.J. (2003). Developmental roles and clinical significance of hedgehog signaling. Curr. Top. Dev. Biol. 53, 1–114.

Raab, J. R., & Kamakaka, R. T. (2010). Insulators and promoters: closer than we think. *Nature Reviews Genetics*, *11*(6), 439.

Shafee, T., & Lowe, R. (2017). Eukaryotic and prokaryotic gene structure. *WikiJournal of Medicine*, *4*(1), 2.

Shubin, N., Tabin, C., & Carroll, S. (2009). Deep homology and the origins of evolutionary novelty. *Nature*, *457*(7231), 818.

Stern, D. L., & Orgogozo, V. (2009). Is genetic evolution predictable? *Science*, *323*(5915), 746-751.

Chapter 4

HOMOLOGY IN ACTION – A PAC CASE STUDY

4.1: The Homology of the ETS Family

The *Hox* genes and *Pax6* function as canonical, well-known examples of deeply homologous proteins in development—as drivers of body plan morphology, they have functioned as excellent tools in bridging the gap between older anatomical studies in fossil records and that of newer molecular investigations. But what about less well-known examples of molecular homology during the later stages of animal development, specifically those in which families of transcription factors drive tissue-specific development?

Unlike other canonical examples of deeply homologous sequences, the ETS family of transcription factors has not been as thoroughly explored from an evolutionary perspective. In terms of embryogenesis, members of the ETS family function as master transcription factors in the development of the animal circulatory system across species. Most research to date on these proteins depicts the biochemical parameters that govern them as a distinct group. Originally named after the E26 transformation-specific family in the avian erythroblastosis virus, the ETS family of transcription factors features 11 distinct subfamilies that have been conserved across a remarkably wide array of species.

The subfamilies are primarily differentiated based on their subdomain compositions and the position of the DNA-binding domain within their primary sequences (Sato, 2001). In accordance with other homologous transcription factors, ETS homologs have been identified in mice, fish, worms, flies, and humans, with most of these proteins acting as transcriptional activators and/or repressors. At the subcellular level, ETS variants have been implicated in cell migration, differentiation, and proliferation during development and adulthood (Craig & Sumanas, 2016). Interestingly, the ETS family was originally used as a model for understanding general transcriptional control and has helped elucidate the epigenetic mechanisms that determine tissue differentiation in model organism systems (Graves & Peterson, 1998). At least thirteen of these transcription factors have been identified as molecular players in hematopoietic (blood cell) and endothelial (blood vessels) stem cell differentiation (Craig & Sumanas, 2016).

ETS transcription factors (TFs) can be easily classified through careful observation of their similar architectures and sequence homology. Each TF features a conserved "ETS" domain made up of 85 amino acids which has a low-complexity binding target of the GCA sequence motif; this relatively small and simple motif highlights the potential functional plasticity of this family such that ETS proteins may have a wide array of targets (Craig & Sumanas, 2016). In Ets1, the binding domain is near the C-terminus, while in Elf1 it resides in the middle of the sequence and at the N-terminus in Elk and Sap1. Rather surprisingly, this family of transcription factors lacks the canonical helix-turn-helix motif for substrate binding that is found in most other TF families. ETS family members can also bind to a wide array of multi-subunit complexes,

suggesting that there are likely other regulatory mechanisms that these proteins may participate in. The different ETS subfamilies likely arose via multiple gene duplication events, as Fli/Ets1 and Erg/Ets2 are closely linked on separate chromosomes. The presence of very similar homologs in *Drosophila melanogaster* (fruit flies, invertebrates) and humans (vertebrates) suggests that these gene duplications were likely ancient events (Wasylyk, Hahn, & Giovane, 1994). There is a possibility that there is also some functional overlap between ETS members within individual organisms, and studies generating knockouts of these genes must often consider the multiple homologs present in each model organism genome (Craig & Sumanas, 2016).

While the ETS family of transcription factors may not be an obvious choice for the study of evolutionary developmental homology, it is important to consider examples of molecular homology outside of the traditional focus of body plan morphology. As key regulators of vascular development, ETS proteins serve as a unique example of how master regulatory molecules specific to the development of individual organ systems may further complicate our understanding of the general facets of animal homology. In the summer of 2018, I was fortunate to have worked in Dr. Saulius' Sumanas' laboratory on a project that focused on the ETS family of proteins in zebrafish vascular development. Model organism systems such as that of *Danio rerio* provide an excellent opportunity for biological researchers to consider how homology has impacted our own species' evolution and health.

Identification of a Novel Vascular Endothelial Progenitor Cell Population in *Danio rerio*

4.2: Project Introduction

Vertebrate vascular development functions as a remarkable example of the diverse molecular mechanisms that support embryogenesis. As one of the first organ systems to fully develop, the vascular system is tasked with directing gas exchange and nutrient acquisition throughout an organism's lifetime. Many studies concerning these processes have used *Danio rerio* as a model species; while blood circulation commences at roughly 24hpf (hours post- fertilization), zebrafish embryos do not rely on circulation until much later in development. This morphological characteristic gives researchers the opportunity to observe phenotypic mutants that would otherwise perish prematurely (Ellertsdottir et al., 2010).

Vascular development in vertebrates is temporally governed by two broad mechanisms—that of vasculogenesis and angiogenesis. Vasculogenesis, the first step in vessel formation, is the *de novo* assembly of primary blood vessels from angioblasts, a progenitor population of cells derived from the posterior lateral plate mesoderm (Sumanas & Lin, 2005; Ellertsdottir et al., 2010). In response to Sonic Hedgehog (SHH) and VEGF signaling, angioblasts migrate bilaterally to the embryonic midline and coalesce, forming the dorsal aorta (DA) and posterior cardinal vein (PCV) (Casie Chetty et al., 2018; Ellertdottir et al., 2010). These cellular movements occur in two waves, with the first giving rise to the dorsal aorta and the second to the cardinal vein (Figure 4.1).

Specification of arterial versus venous fates is accomplished before vasculogenesis is complete (Kohli et al., 2013).

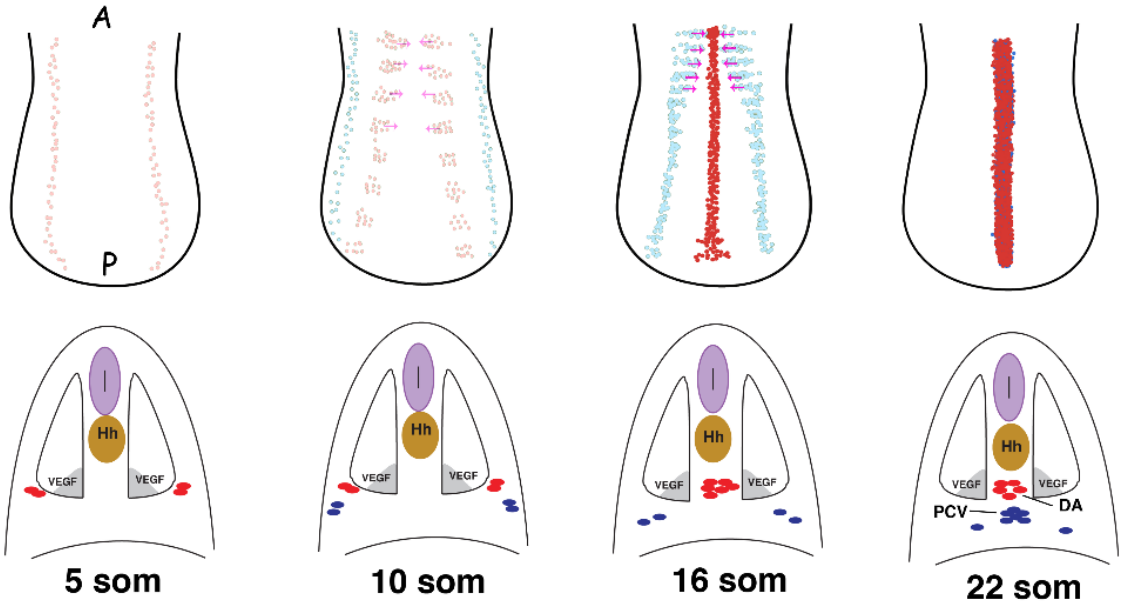

Figure 4.1. Angioblast migration during vasculogenesis. (Top) The primordial formation of the dorsal aorta occurs before that of the posterior cardinal vein. Cells in red depict the first wave of progenitor cells migrating and coalescing to form the dorsal aorta, while blue cells depict the second wave that give rise to the posterior cardinal vein. (Bottom) SHH and VEGF gradients instruct arterial and venous angioblasts to migrate bilaterally. The top, purple structure represents the notochord, the middle represents the expression gradient of SHH, and the sides represent the expression gradients for VEGF. Red cells depict DA progenitor populations, while blue cells depict PCV progenitors. Adapted from Casie Chetty et al., 2018.

At roughly 22hpf, angiogenesis begins to take over as the primary mechanism of blood vessel formation in the zebrafish trunk. Unlike vasculogenesis, angiogenesis is the elaboration of the vasculature from the scaffolding of pre-existing vessels. Angiogenesis begins when aortic sprouts migrate dorsally and form the aortic intersegmental vessels (ISVs). In keeping with the two-wave model of angioblast movement, cells from the

posterior cardinal vein sprout later than those from the dorsal aorta (Figure 4.2). Interestingly, sprouts from the PCV are not restricted to a segmental fate; instead, they may contribute to the lymphatic vasculature or migrate ventrally to become the subintestinal vasculature (Casie Chetty et al., 2018). It is currently thought that after the primary vessels have formed in the trunk, all new vessels thereafter are formed exclusively via angiogenesis; it is not clear, however, if new vascular progenitors are capable of incorporating into the existing vasculature at this time in development.

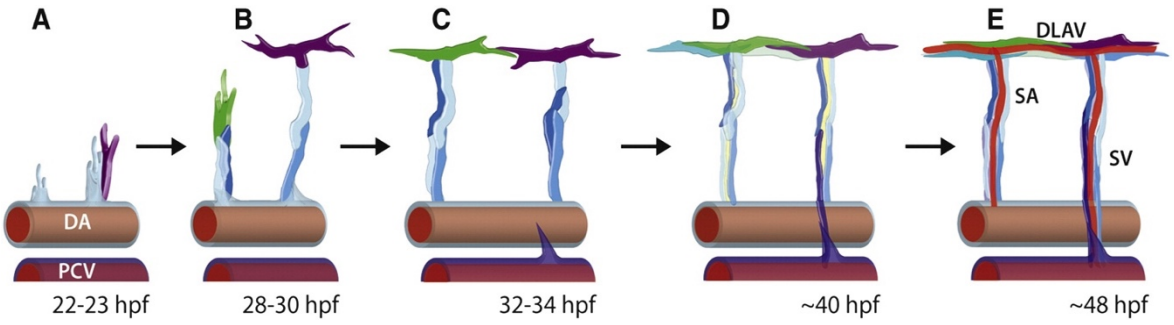

Figure 4.2. Timeline of angiogenic sprouting from the Dorsal Aorta (DA) and Posterior Cardinal Vein (PCV). After the establishment of the DA and PCV following vasculogenesis, sprouts first migrate out of the DA at 22-23hpf, followed by those from the PCV at about 32-34hpf. The Dorsal Longitudinal Anastomotic Vessel and associated segmental vessels/veins are established in the trunk vasculature by around 2 days post fertilization. SA: Segmental Artery; SV: Segmental Vein; DLAV: Dorsal Longitudinal Anastomotic Vessel. Adapted from Ellertsdottir et al., 2010.

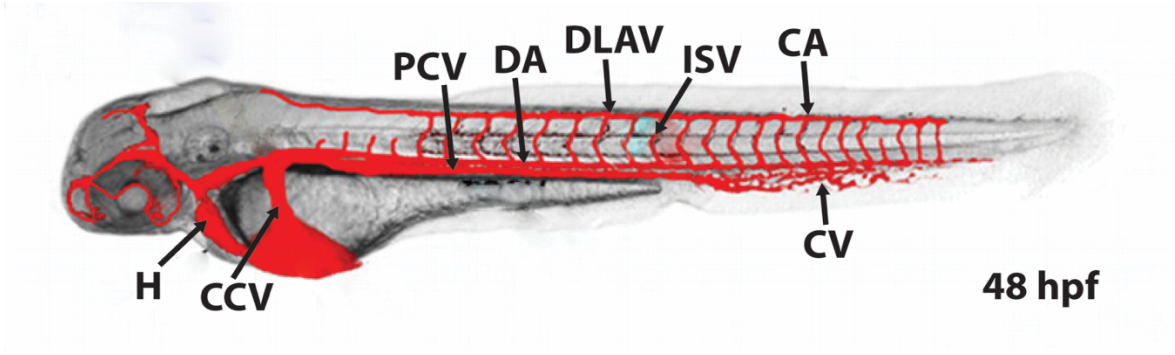

Figure 4.3. The major vessels of the zebrafish trunk at 48hpf. H: Heart; CCV: Common Cardinal Veins; DA: Dorsal Aorta; PCV: Posterior Cardinal Vein; SV: Segmental Vein; SA: Segmental Artery; DLAV: Dorsal Longitudinal Anastomotic Vessel; ISV: Intersegmental Vessels; CA: Caudal Aorta; CV: Caudal Vein. Adapted from Bolcome et al., 2008.

At the molecular level, several important TFs have been implicated in the regulation of vasculogenesis and angiogenesis. One such regulatory molecule is *etv2*; this member of the ETS family of transcription factors has been shown to direct the early differentiation of vascular endothelial cells in the zebrafish trunk beginning at the 10-somite stage (Sumanas & Lin, 2005). *etv2* expression in the vasculature generally decreases at 24hpf; however, our laboratory has demonstrated that a population of progenitor cells exhibit high expression of this TF between 24 and 48hpf. Our laboratory named these cells "Pronephros-associated cells" (PACs) due to their proximity to the pronephros (primitive kidney) (Figure 4.4).

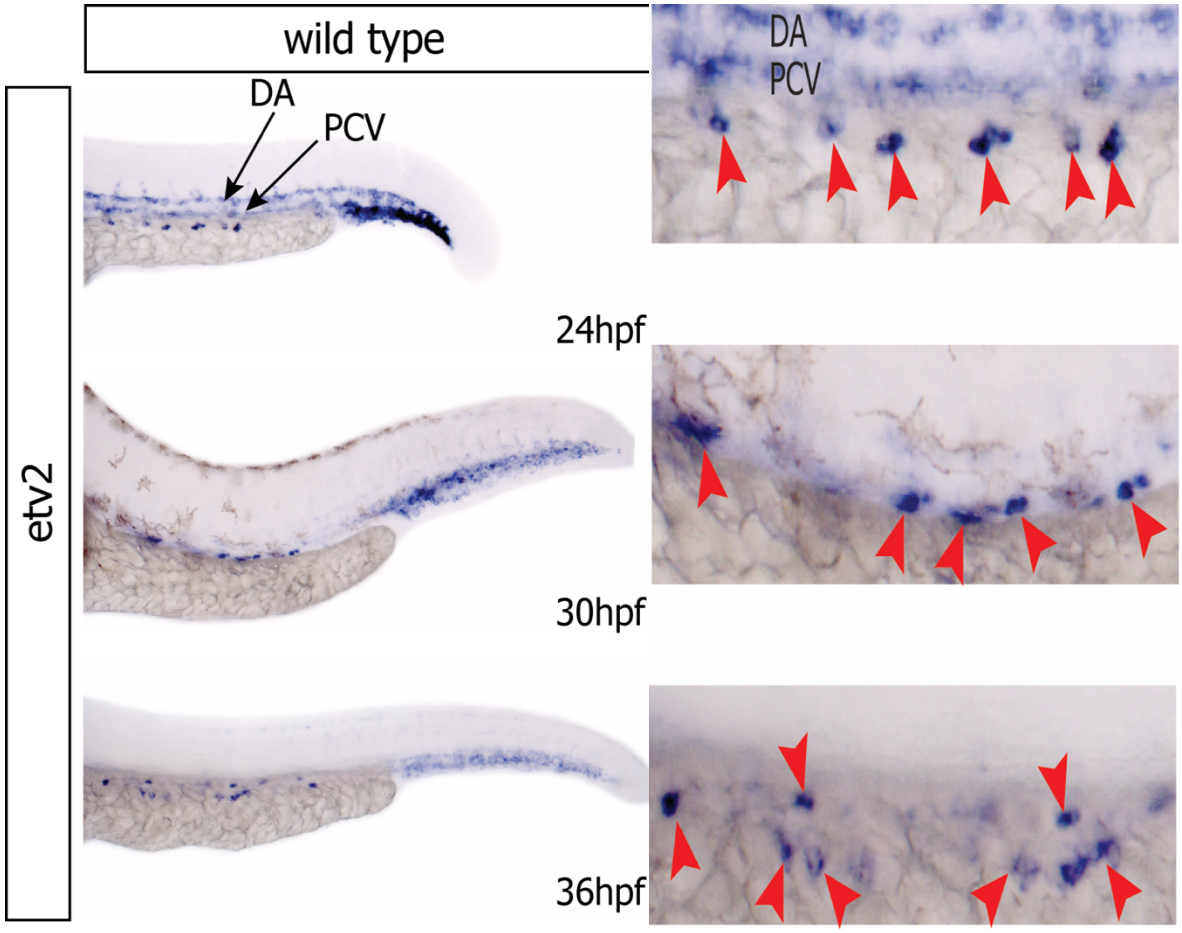

Figure 4.4. PACs expression of *etv2* in the zebrafish trunk. Whole-mount *in situ* hybridization imaging of wild type expression at 24, 30, and 36hpf. Red arrows indicate locations of PACs in the trunk at corresponding time points. DA: Dorsal Aorta; PCV: Posterior Cardinal Vein. Images courtesy of Sanjeev Metikala, Sumanas Lab, Cincinnati Children's Hospital Department of Developmental Biology.

PACs are most numerous in the embryonic trunk between 32hpf and 38hpf (Figure 4.5). Upon migration into the midline, time-lapse imaging indicates that PACs appear in the yolk extension and subsequently integrate into the PCV. Remarkably, this integration mechanism leaves blood circulation uninterrupted (Figure 4.6). Besides *etv2,* PACs also express other canonical vascular endothelial progenitor markers, such as *lmo2* (Lim

domain only 2) and *scl* (T-cell acute lymphocytic leukemia 1) (Figure 4.6). While useful in elucidating their role as a progenitor population, these markers are not PAC-specific, as *etv2* is a general vascular regulator, *scl* is a known hematopoietic marker, and *lmo2* is expressed in macrophages. We currently do not have a marker that is exclusive to PACs; however, upon analysis of bulk RNA-seq analysis and subsequent observation of expression in the yolk extension at 30hpf, *jam2b* (a junctional adhesion molecule) has been implicated as a potential regulatory molecule for PACs (Figure 4.7).

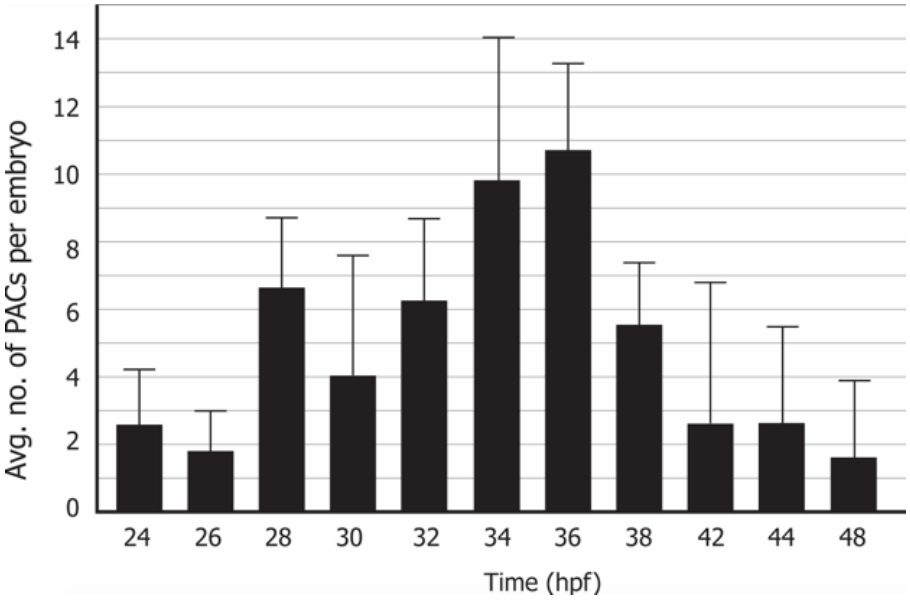

Figure 4.5. Histogram of average number of PACs per embryo at various stages of zebrafish embryonic development. Error bars represent +/- 1 standard deviation. Data courtesy of Andrew Koenig and Sanjeev Metikala, Sumanas Lab, Cincinnati Children's Hospital Department of Developmental Biology.

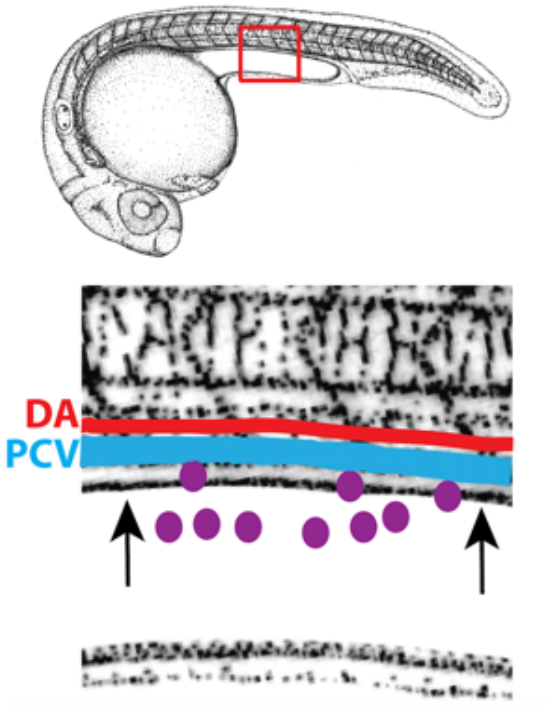

Figure 4.6. Illustration of PAC integration into the PCV. Figure courtesy of Sanjeev Metikala, Sumanas Lab, Cincinnati Children's Hospital Department of Developmental Biology.

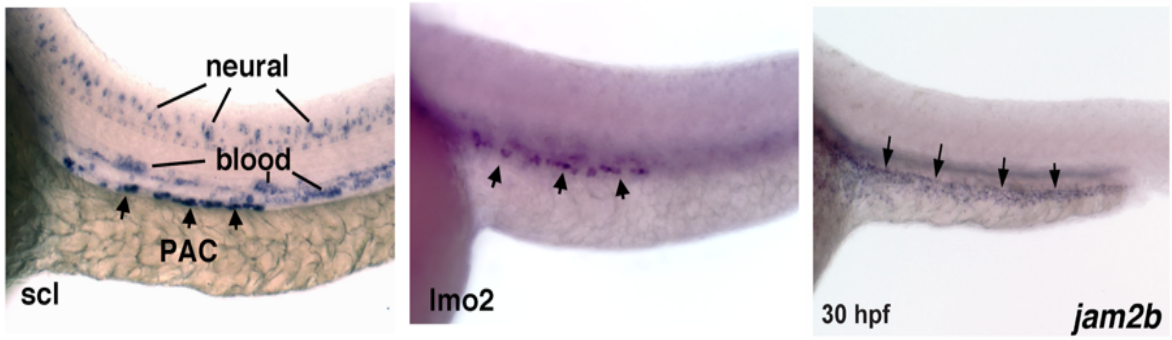

Figure 4.7. scl, lmo2, and *jam2b* expression in the yolk extension. *scl* and *lmo2* are both expressed in PACs at 30hpf. Expression of *jam2b* at 30hpf is exclusively localized to the yolk extension, the same location as PACs at this stage in zebrafish development. Images courtesy of Sanjeev Metikala, Sumanas Lab, Cincinnati Children's Hospital Department of Developmental Biology.

Due to their behavior as a progenitor population, our lab hypothesized that PACs are capable of giving rise to new endothelial cells in the zebrafish vasculature. In this project, we investigated if PACs are capable of repopulating the trunk vasculature, as well as whether integration is vessel-specific. Furthermore, we predicted that *jam2b* is a molecular regulator of PACs because of its spatiotemporal expression in the yolk extension at 30 hpf. We aimed to elucidate if *jam2b* loss-of-function affects PAC frequency in the trunk.

4.3: Conditional Cell Ablation- Methodology

In order to investigate whether PACs are capable of giving rise to new endothelial cells upon integration into the major trunk vessels, targeted cell ablation (death) was induced in vascular cells. Traditional methods of invoking cell-specific apoptosis include surgery, laser-mediated ablation, and the creation of transgenic lines with DTA (diptheratoxin A); however, many of these techniques are labor-intensive, time-consuming, and hard to reproduce in a large number of embryos (Curado et al., 2009).

A novel method for conditional ablation that has been optimized in *D. rerio* is the NTR-MTZ method. This protocol capitalizes on *E. coli* nitroreductase (NTR), an enzyme that catalyzes the conversion of the nontoxic prodrug metronidazole (MTZ) into a cytotoxic substrate. The NTR method uses engineered, transgenic lines to sequester the MTZ conversion effect to a target population of cells (Figure 4.8); for our experiment, we inserted the NTR gene under an etv2 promoter in an etv2:Gal4 line, giving rise to the *Tg(etv2:Gal4, UAS:gfp, UA:ntr-mCherry)* genotype (Curado et al., 2009). Upon MTZ

treatment, we thus expected to observe that all vascular cells had been ablated before "rescue," i.e. removal of treatment.

Transgenic organisms, such as those in our specific reporter line, are organisms that have experienced some form of genomic alteration. In a developmental laboratory that utilizes a live model organism system, it is imperative that specific cells (such as those of the zebrafish vasculature) can be visualized apart from those of other tissues in an actively growing organism. *gfp,* a gene derived from jellyfish that produces a fluorescent product when expressed in other organisms, has become an excellent tool for accomplishing such a task. When placed under the promoter of a regulatory transcript that is expressed in a unique cell type, it allows for the fluorescent detection of target cells in an otherwise incoherent biological landscape (Amsterdam, Lin, Moss, & Hopkins, 1996). In the case of our transcriptional reporter line, *gfp* was placed under the promoter of *etv2* because this ETS family member is exclusively expressed in vascular progenitor cells during specific developmental time points of interest. When visualized under a fluorescent compound microscope, cells expressing GFP will glow and highlight those cells that are specific to the vascular system (Figure 4.8).

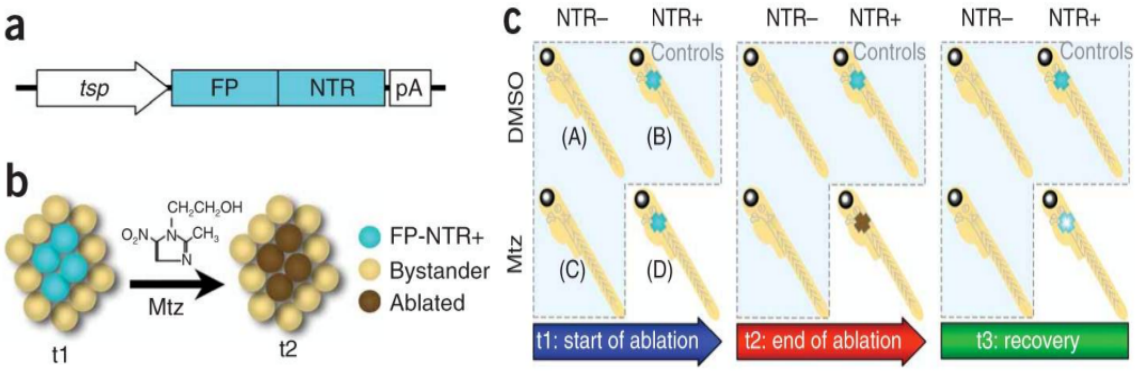

Figure 4.8. Experimental Design for the NTR-MTZ method of targeted cell ablation in *D. rerio.* (A) Schematic of template transgene for a cell-specific promoter (tsp: transcription start point; FP: fluorescent protein). (B-C) Using a cell-specific promoter allows for conditional ablation and leaves bystander populations untouched in transgenic embryos. Figure adapted from Curado et al., 2007.

In order to acquire the necessary quantities of embryos within the *Tg(etv2:Gal4, UAS:gfp, UA:ntr-mCherry)* for our experimental purposes, males and females of the same *etv2:gal4/UAS:ntr-mcherry* genotype were sorted for mating once a week. Eggs were collected from tanks and stored in fish water (60µg/ml salt water) at 28.5°C, which is the optimum temperature for embryo development. Once the embryos reached 50% epiboly (about 6 hpf), MTZ treatment commenced. Epiboly is one of the major types of cell movements that occur during gastrulation, an early phase in animal development in which a single-layered embryo becomes reorganized into multiple germ layers. At 50% epiboly, hemangioblasts begin to differentiate and prepare for the onset of *etv2* expression, a beginning marker of the onset of vasculogenesis (Casie Chetty et al., 2018).

Embryos were allowed to continue developing under the MTZ treatment conditions until developmental characteristics consistent with 24hpf were observed; at

this time, MTZ treatment solution was replaced with DMSO fish water (a control solution) and eggs were returned to the incubator. Fluorescence imaging of live embryos was conducted at 38hpf and 60hpf within the following days, followed by corresponding *in situ* hybridization fixation.

In situ hybridization (ISH) is a widely-used experimental protocol that allows for the targeted visualization of a specific transcript (mRNA) within a fixed organism/embryo. We utilized ISH to supplement our fluorescence imaging because it allows for the observance of a target transcript's expression at an arrested stage of development. For this histological protocol, we utilized an *etv2* probe and imaged the expression of this transcription factor in embryos fixed at 38hpf and 60hpf under a bright field compound microscope (procedure adapted from Thisse & Thisse, 2008).

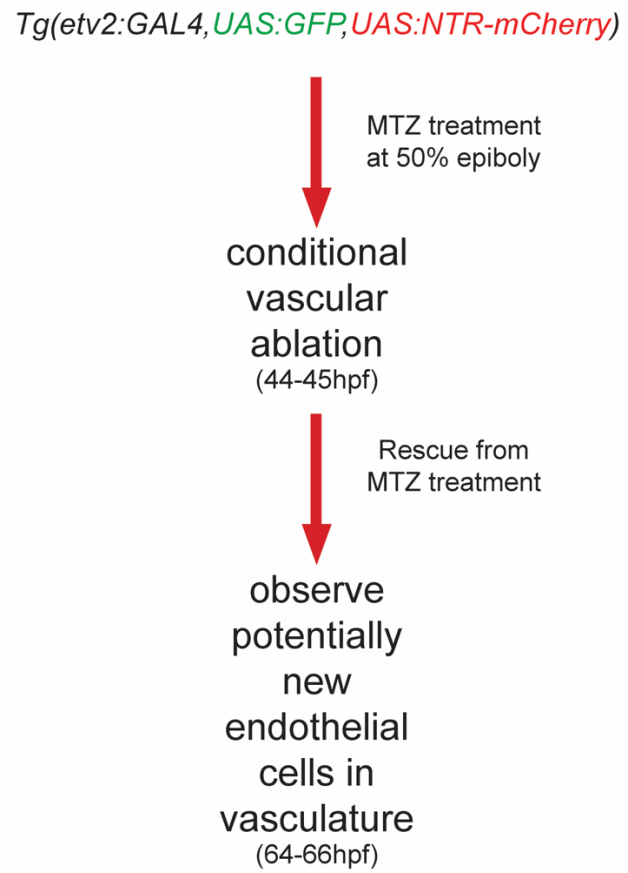

Figure 4.9. Workflow diagram of MTZ treatment outcomes in transgenic embryos versus controls in our experiment.

4.4: Conditional Cell Ablation- Results

Before investigating the rescued phenotype in the *Tg(etv2:Gal4, UAS:gfp, UA:ntr-mCherry)* line, we optimized the concentration of MTZ needed for complete vascular ablation. While MTZ is not known to induce dramatic side effects in treated zebrafish embryos, we did note that treatment incurred a developmental delay of 1-3 hours within the majority of our embryos, suggesting a potential for toxic side effects. An initial concentration of 2.5 mM MTZ led to reasonable vascular ablation; however,

embryos showed traces of vascular cells that had escaped the effect, suggesting that this concentration was too low for the 38-hour treatment cycle. We determined that 10 mM MTZ solution was sufficient to induce ablation without incurring toxic side effects to the embryos, as deemed by whether the embryos were able to reach wild-type developmental milestones within the expected margin of time (Figure 4.10).

Upon rescue from MTZ treatment at 44hpf, embryos were placed in 0.2% DMSO (a control solution without any treatment) until 60-64hpf. Fluorescence imaging indicated that new endothelial cells appeared in the primary trunk vessels after treatment had been removed. This was confirmed through *etv2*-mediated ISH, as PACs are the only cells remaining in the vasculature at 64hpf that express this vasculogenic TF. Furthermore, because PACs arrive in the zebrafish trunk later on in development and are a progenitor population, they are more likely to escape MTZ-mediated ablation and regenerate (Figure 4.11a-b). Time-lapse imaging is needed to confirm that these endothelial cells are entirely derived from PACs within this region of the vasculature.

After witnessing the migration of new endothelial cells into the ablated zebrafish vasculature post-rescue at 66hpf, we were curious as to whether these cells integrated into the posterior cardinal vein, dorsal aorta, and subintestinal vasculature non-randomly. These cells appeared to favor migration into the regions of the dorsal aorta and cardinal vein over that of the subintestinal vasculature, although replication of these results through time-lapse imaging would be needed to confirm this observation (Figure 4.11c).

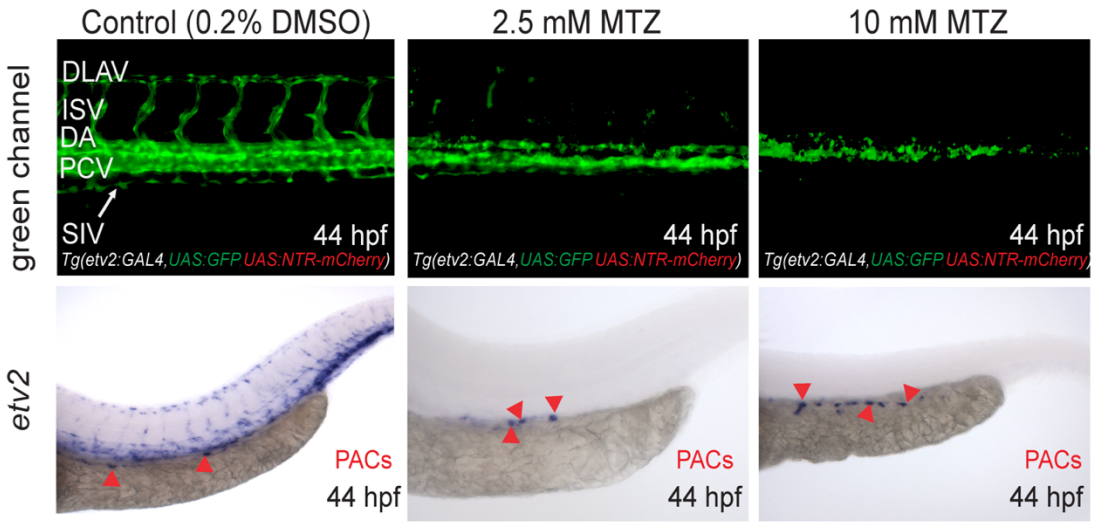

Figure 4.10. 10mM MTZ is sufficient to induce complete vascular ablation. Fluorescent and *ISH* imaging of control, 2.5mM MTZ treated, and 10mM MTZ treated embryos from the *Tg(etv2:GAL4, UAS:gfp, UAS:ntr-mCherry)* transgenic line. Embryos were subjected to treatment from 6hpf (50% epiboly) until 44hpf (roughly 38-40 hours of treatment was necessary to induce complete vascular ablation in the trunk vasculature). While 2.5mM MTZ treatment does not incur complete vascular ablation, 10mM MTZ appears to be sufficiently concentrated. Red arrowheads correspond to example PACs following ISH under the three experimental conditions. SIV: Subintestinal Vein; PCV: Posterior Cardinal Vein; DA: Dorsal Aorta; ISV: Intersegmental Vessels; DLAV: Dorsal Longitudinal Anastomotic Vessel.

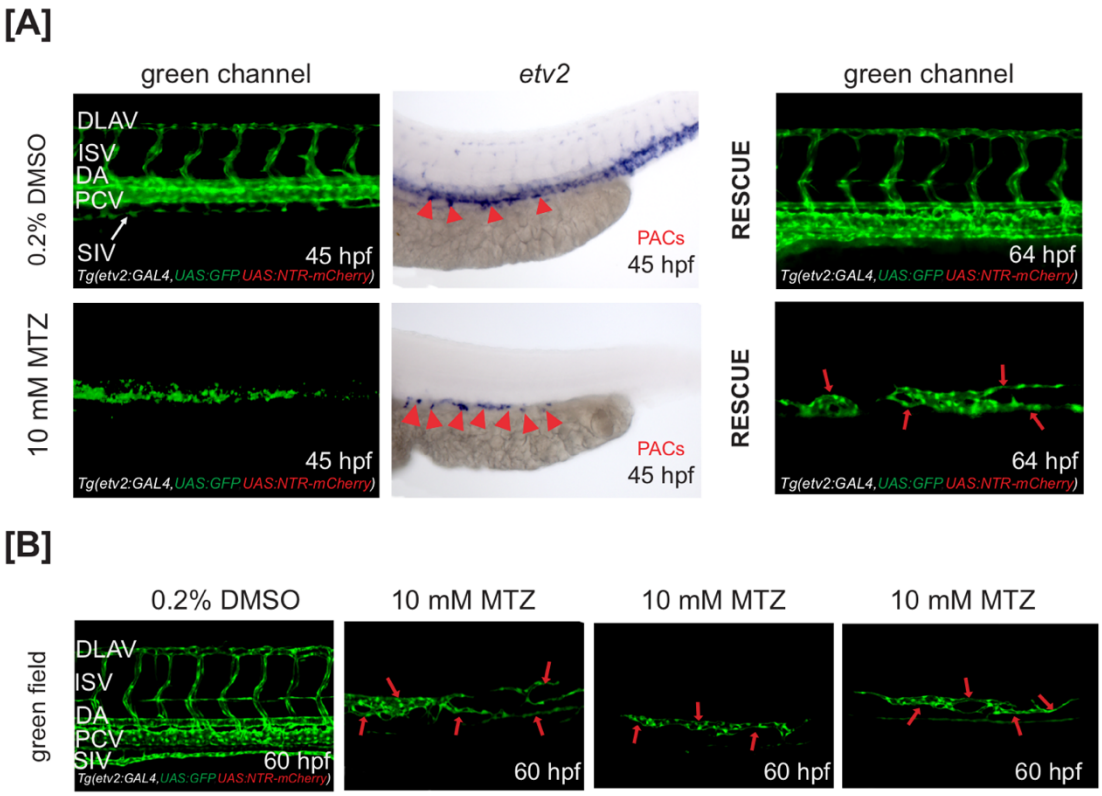

Figure 4.11. Vascular ablation and recovery allow new endothelial cells to emerge. (A) Fluorescent and *ISH* imaging of control and 10 mM MTZ treated embryos from the *Tg(etv2:GAL4, UAS:gfp, UAS:ntr-mCherry)* transgenic line. Rescue was performed via replacement of MTZ-treated fish water with 0.2% DMSO fish water. (A-B) Upon MTZ treatment, PACs (arrowheads) are the only cells in the trunk vasculature because they appear later in development. At 60-64hpf, new endothelial cells (elongated cells, red arrows) appear in the trunk after removal of MTZ treatment. (C) Observed location of new endothelial cells post MTZ-recovery at 66hpf. Cell location (PCV, DA, subintestinal) determined by comparison with control of the same age. SIV: Subintestinal Vein; PCV: Posterior Cardinal Vein; DA: Dorsal Aorta; ISV: Intersegmental Vessels; DLAV: Dorsal Longitudinal Anastomotic Vessel.

4.5: **jam2b,** ***A Potential Regulatory Molecule for PACs- Methodology***

In an effort to identify a molecular regulator for PACs, our laboratory performed single-cell RNA sequencing analysis on vascular progenitor cells in an *etv2* transgenic reporter line. Embryos were collected at the 15-20 somite stage (development time point in which angioblasts migrate bilaterally to the embryonic midline and coalesce) and were subjected to Fluorescence-activated cell sorting for the vascular endothelial progenitor subpopulation (Figure 4.12). Single-cell sequencing results included *jam2b* as an upregulated differentially-expressed gene in this population; ISH using a *jam2b* probe demonstrated that *jam2b* is expressed in the yolk extension at the same developmental time as PACs (30hpf, see Figure 4.7). Due to this spatiotemporal pattern of *jam2b* expression at 30hpf, we hypothesized that *jam2b* is a molecular regulator of PACs. We utilized a morpholino injection methodology to quantitatively vary the amount of *jam2b* expression in another transgenic line.

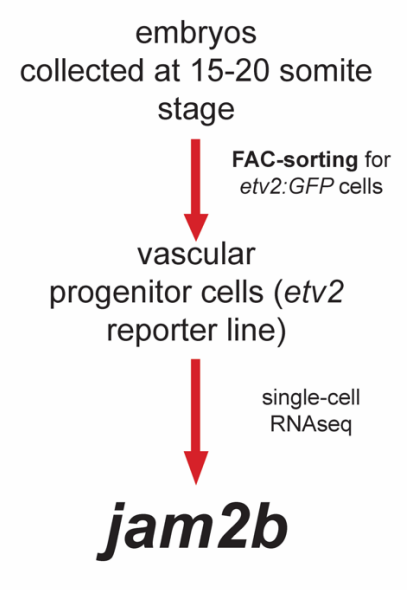

Figure 4.12. RNA-sequencing analysis in vascular progenitors implicates *jam2b* as a potential regulator of PACs. Embryos were collected at the 15-20 somite stage and subjected to FAC sorting for *etv2:gfp* expressing cells.

Contrary to the MTZ methodology, embryos from the *fli1* reporter line *Tg(fli1-UAS:gfp)* (another TF specific to vascular endothelial cells) were subjected to morpholino injection following early-morning collection. A morpholino is an oligomer molecule that is used by molecular biologists to conditionally "knockdown" the expression of a given transcript; through the injection of a morpholino, which is complementary to the messenger RNA for this protein, molecular biologists are able to alter the spatiotemporal expression of *jam2b* in model organisms. Morpholinos have been widely used in zebrafish conditional knockdown studies because these embryos are large in size, transparent, and amenable to injection. Our morpholino included a red fluorescent dye (phenol red) such that successful injection could be observed under a compound microscope (MO injection methodologies adapted from Nasevicius & Ekker, 2000).

Following morpholino injection between 1 and 2hpf, eggs were placed at 28.5 °C. Inspection and fluorescent imaging of injected embryos as compared to controls occurred periodically between 24hpf and 48hpf (time span in which PAC integration into the trunk vasculature is most frequent). Fluorescent imaging and subsequent confirmation via ISH were conducted at 24hpf, 36hpf, and 48hpf with morpholino injection concentrations at 0.5 ng, 0.75 ng, and 1.5 ng.

4.6: **jam2b*, A Potential Regulatory Molecule for PACs- Results***

Much like the MTZ-mediated ablation protocol, higher concentrations of morpholino have the potential to incur debilitating cytotoxic effects for developing zebrafish embryos. For this reason, we considered the side effects of injecting a *jam2b* MO at concentrations of 0.5 ng, 0.75 ng, and 1.5 ng. We noted the anatomical characteristics of each batch of embryos from the same transgenic line and reported any physical deformities and delayed development compared to that of same-age controls. We noticed that *jam2b* knockdown had a more significant effect on both vascular and general development in the later stages; while high concentrations (1.5 ng) incurred tail deformities at 24hpf, these deformities were comparatively more numerous at 36hpf. Furthermore, we found that a high dosage at 1.5 ng led (on average) to a four-hour developmental delay. For these reasons, we optimized the morpholino injection at a low dosage of 0.5 ng for future studies concerning PAC depletion in the trunk vasculature; at this concentration, treated embryos had markedly diminished trunk vasculatures

compared to that of their control counterparts, yet did not feature major tail deformities and significant developmental delays.

In congruence with our hypothesis, *jam2b* loss-of-function negatively affects PAC counts in the trunk vasculature at 24hpf, 36hpf, and 48hpf. At higher dosages of 0.75 ng and 1.5 ng, the trunk vasculature was either delayed or incomplete at these time points; however, at 0.5 ng, angiogenic timing of sprouting from the dorsal aorta and posterior cardinal vein aligned with that of the control embryos. Control embryos at each of the three collection points featured higher PAC counts on average compared to their treated counterparts. Furthermore, we observed more PACs overall in embryos at 36hpf than 24hpf. These results suggest that even a low-dosage knockdown of JAM2B function affects PAC proliferation in the trunk vasculature to a statistically significant extent. *jam2b* thus remains a promising candidate regulator of PACs (Figure 4.13).

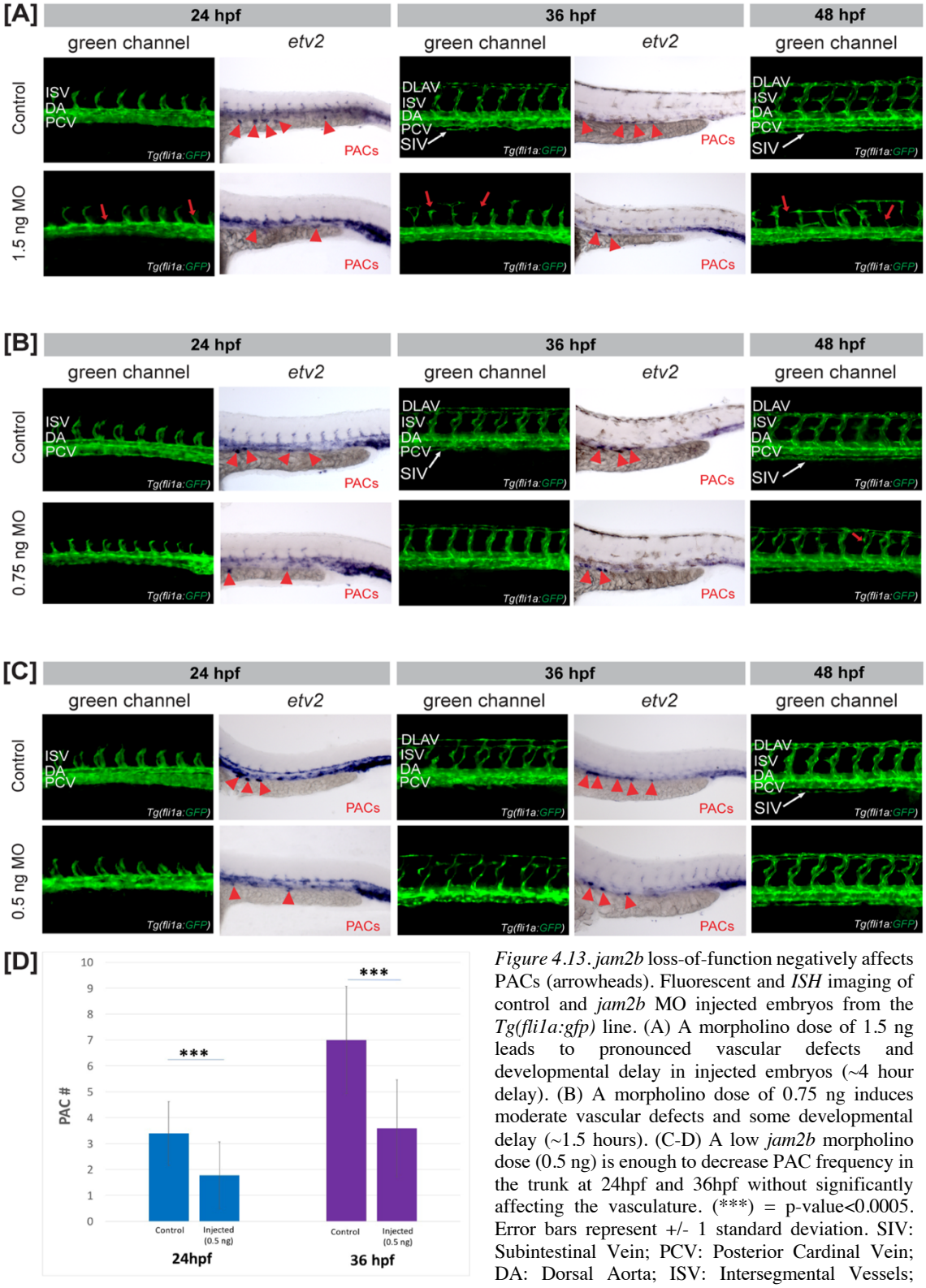

Figure 4.13. jam2b loss-of-function negatively affects PACs (arrowheads). Fluorescent and *ISH* imaging of control and *jam2b* MO injected embryos from the *Tg(fli1a:gfp)* line. (A) A morpholino dose of 1.5 ng leads to pronounced vascular defects and developmental delay in injected embryos (~4 hour delay). (B) A morpholino dose of 0.75 ng induces moderate vascular defects and some developmental delay (~1.5 hours). (C-D) A low *jam2b* morpholino dose (0.5 ng) is enough to decrease PAC frequency in the trunk at 24hpf and 36hpf without significantly affecting the vasculature. (***) = p-value<0.0005. Error bars represent +/- 1 standard deviation. SIV: Subintestinal Vein; PCV: Posterior Cardinal Vein; DA: Dorsal Aorta; ISV: Intersegmental Vessels; DLAV: Dorsal Longitudinal Anastomotic Vessel.

4.7: Discussion & Future Directions

While time-lapse imaging is needed to confirm if the migratory endothelial cells following MTZ ablation are derived from PACs, the presence of these cells in the vasculature between 38 and 60hfp suggests that vasculogenesis remains an important mechanism even after angiogenic sprouting has commenced in the trunk. As per the MTZ methodology, the only cells that could possibly escape complete ablation between these developmental time points would be the PACs. Due to their late migration into the trunk and potential for regeneration as a progenitor population, it is possible that the observed endothelial cells at 60hpf are nothing more than differentiated daughter cells of the PAC population. Previously, it was thought that the completion of vasculogenesis linearly gives rise to the onset of angiogenic sprouting in the trunk vasculature; however, our data suggest that endothelial integration into the posterior cardinal vein occurs well after blood circulation has commenced via a distinctly vasculogenic pattern of migration (Casie Chetty et al., 2018).

ISH analysis using vessel-specific markers is needed to confirm whether these migratory endothelial cells favor integration into the dorsal aorta, posterior cardinal vein, or subintestinal lymphatic vasculature following MTZ-mediated ablation. *flt4* (fms-related tyrosine kinase), a venous marker, *cldn5* (claudin 5), an arterial marker, and *lyve1* (lymphatic vessel endothelial hyaluronic receptor 1), a lymphatic marker, are all potential candidate probes for use in an ISH protocol following MTZ ablation. Future experimentation is also needed to consider if complete vascular ablation following 10 mM

MTZ treatment is complete for all *etv2*-expressing endothelial cells by 24, 30, and 36hpf, rather than as late as 44hpf.

The *etv2* expression profile and late migratory patterns of PACs have allowed for the isolation of PAC cells from other vascular cell populations in the zebrafish trunk; however, future studies concerning these cells would benefit from the knowledge of a potential regulatory molecule specific to their unique transcriptome. The confirmation of *jam2b* as a PAC-specific regulatory molecule after 30hpf, as well as the optimization of *jam2b* morpholino injection concentration, will allow for easier identification of the unique PAC identity. Even at the low dosage of 0.5 ng, *jam2b* morpholino injection reduces the frequency of PACs in the trunk vasculature by half; this optimized model allows for the isolation of PAC functionality without a significant inhibition of normal vascular development.

With this optimized morpholino injection methodology in mind, we aim to determine if *jam2b* knockdown affects the number of endothelial cells that integrate into the vasculature at later stages in development. If PAC counts are significantly diminished in the trunk vasculature even after low-concentration *jam2b* MO injections, it is possible that injected embryos will experience vessel defects in the later stages of development, such as through diminished blood flow capacity and vessel leakage. Furthermore, we hope to use our morpholino methodology to determine if *jam2b* knockdown significantly affects the comparative diameters of the dorsal aorta and posterior cardinal vein to that of wild-type embryos in the same transgenic line.

Since PAC cells have previously remained as an unidentified population of progenitor cells in the zebrafish vasculature, there is no research to date on whether these cells appear in a murine model; however, considering the basic homology of vascular development across the animal kingdom, PACs certainly have the potential to become therapeutic targets should they prove to be important late-stage migratory cells in the mammalian vasculature. The high expression of *etv2* during the earliest stages of vasculogenesis and in PACs at the beginning of angiogenesis also further implicates the ETS family of transcription factors as molecular bridges between basic and translational hematopoietic research.

4.8: Chapter Four References

Amsterdam, A., Lin, S., Moss, L. G., & Hopkins, N. (1996). Requirements for green fluorescent protein detection in transgenic zebrafish embryos. *Gene*, *173*(1), 99-103.

Baltrunaite K, Craig MP, Palencia-Desai S, Chaturvedi P, Pandey RN, Hegde RS, Sumanas S. ETS transcription factors Etv2 and Fli1b are required for tumor angiogenesis. Angiogenesis. 2017 Aug:20(3):307-323.

Bolcome, R. E., Sullivan, S. E., Zeller, R., Barker, A. P., Collier, R. J., & Chan, J. (2008). Anthrax lethal toxin induces cell death- independent permeability in zebrafish vasculature. *Proceedings of the National Academy of Sciences*, *105*(7), 2439-2444.

Browder, L. & Iten, L. (1998). Why study Zebrafish? *University of Calgary.* Retrieved from https://people.ucalgary.ca/~browder/why_fish.html.

Chetty S.C., Choi K. and Sumanas S. (2018) Vasculogenesis in Development. In: Vasan, R., Sawyer, D., The Encyclopedia of Cardiovascular Research and Medicine. vol. 4, p. 640-653. Oxford: Elsevier.

Craig, M. P., & Sumanas, S. (2016). ETS transcription factors in embryonic vascular development. Angiogenesis, 19(3), 275-285.

Curado, S., Anderson, R. M., Jungblut, B., Mumm, J., Schroeter, E., & Stainier, D. Y. (2007). Conditional targeted cell ablation in zebrafish: a new tool for regeneration studies. *Developmental dynamics: an official publication of the American Association of Anatomists*, *236*(4), 1025-1035.

Ellertsdóttir, E., Lenard, A., Blum, Y., Krudewig, A., Herwig, L., Affolter, M., & Belting, H. G. (2010). Vascular morphogenesis in the zebrafish embryo. *Developmental biology*, *341*(1), 56-65.

Figg, W. D., & Folkman, J. (Eds.). (2008). Angiogenesis: an integrative approach from science to medicine. Springer Science & Business Media.

Graves, B. J., & Petersen, J. M. (1998). Specificity within the ets family of transcription factors. In Advances in cancer research (Vol. 75, pp. 1-57). Academic Press.

Howe DG, Bradford YM, Conlin T, Eagle AE, Fashena D, Frazer K, Knight J, Mani P, Martin R, Moxon SA, Paddock H, Pich C, Ramachandran S, Ruef BJ, Ruzicka L, Schaper K, Shao X, Singer A, Sprunger B, Van Slyke CE, Westerfield M.

(2013). ZFIN, the Zebrafish Model Organism Database: increased support for mutants and transgenics. Nucleic Acids Res. Jan;41(Database issue):D854-60.

Kohli, V., Schumacher, J. A., Desai, S. P., Rehn, K., & Sumanas, S. (2013). Arterial and venous progenitors of the major axial vessels originate at distinct locations. *Developmental cell*, *25*(2), 196-206.

Nasevicius, A., & Ekker, S. C. (2000). Effective targeted gene 'knockdown'in zebrafish. *Nature genetics*, *26*(2), 216.

Park, C., Lee, T. J., Bhang, S. H., Liu, F., Nakamura, R., Oladipupo, S. S., ... & Urao, N. (2015). Injury-mediated vascular regeneration requires endothelial ER71/ETV2. Arteriosclerosis, thrombosis, and vascular biology, ATVBAHA-115.

Sato, Y. (2001). Role of ETS family transcription factors in vascular development and angiogenesis. Cell structure and function, 26(1), 19-24.

Sumanas, S., & Lin, S. (2005). Ets1-related protein is a key regulator of vasculogenesis in zebrafish. *PLoS biology*, *4*(1), e10.

Thisse, C., & Thisse, B. (2008). High-resolution in situ hybridization to whole-mount zebrafish embryos. *Nature protocols*, *3*(1), 59.

Washington University Zebrafish Facility. Retrieved from http://zebrafishfacility.wustl.edu/.

Wasylyk, B., Hahn, S. L., & Giovane, A. (1994). The Ets family of transcription factors. In Ejb Reviews 1993 (pp. 7-18). Springer, Berlin, Heidelberg.

Chapter 5

THE BIOMEDICAL IMPLICATIONS OF VASCULAR HOMOLOGY

5.1: Introduction

Perhaps one of the most difficult tasks for biomedical researchers today is the challenge of linking basic scientific discoveries with potential therapeutics. While the use of homologous animal models for studying complex human diseases continues to serve us well, we still have much work to do in connecting bench discoveries to biomedical advances. For example, it is certainly relevant to study the nuanced molecular mechanisms that drive zebrafish vasculogenesis and angiogenesis simply for the sake of understanding this model organism system; however, it is not until we have taken these findings and compared them with that of an *in vitro* human model that our discoveries can begin to answer complex questions concerning human health. This is perhaps the most exciting part of studying the homology of the animal kingdom.

5.2: Vascular Regeneration Throughout the Life Cycle

The animal vascular system is a complex network that, much like other systems in the body, must remain constantly responsive to environmental cues and perturbations. Larger vessels, such as that of the aorta and superior vena cava, are comprised of

endothelial cells and mural cells (smooth muscle cells) that sustain necessary structural integrity. In contrast, smaller vessels, such as those contained in dense capillary networks within tissues, are solely comprised of endothelial cells. During development, arterial growth precedes capillary angiogenesis and reigns as the principle method of vascularization; however, the adult life cycle is almost entirely characterized by small instances of capillary angiogenesis that give rise to new vessels upon tissue damage or regeneration (Luttun, Carmeliet, & Carmeliet, 2003; Carmeliet, 2003).

Normal vascular alterations that occur during the adult life cycle tend to be placed in one of three categories—that of immune-driven angiogenesis (in response to inflammation), coagulation (wound healing/clotting), and vessel regression. Hematopoietic and angiogenic factors that are featured in these three mechanisms are produced in the bone marrow during this stage of the life cycle (the job of hematopoietic stem cells) and are mainly controlled by the master angiogenic regulator "vascular endothelial growth factor" (VEGF) (Luttun, Carmeliet, & Carmeliet, 2003; Carmeliet, 2003). Other important molecular players in adult angiogenesis include placental growth factor (P1GF, a homolog of VEGF), angiopoietin 1 (ANG-1), angiopoietin 2 (ANG-2), and various signaling cytokines that respond to a wide variety of environmental cues (Carmeliet, 2003). In contrast to the ETF family of transcription factors, there has been substantial research concerning how these homologous factors play a role in both normal animal development and adult disease etiology.

Adult endothelial cells do not exhibit identical gene expression patterns across tissue systems. Expression gradients of VEGF and ANG-1 between vessels contribute to

these tissue and site-specific expression profiles (Carmeliet, 2003). A great example of this variability is found in a comparison of low-permeability and high-permeability tumors in cancer—while low-permeability tumors generally overexpress ANG-1 and dramatically under express VEGF, high-permeability tumors lack ANG-1 expression almost altogether. This pattern of differential expression among subtypes of many diseases implicates the need for researchers to identify organ-specific molecules and site-specific treatment methodologies (Jain & Munn, 2000).

5.3: Tumorigenesis—Hijacking the Vascular Machinery

The vascular system provides one of the most important functions for the support of multicellular organization—that of consistent nutrient and oxygen access. Due to the diffusion limit of oxygen, all mammalian cells must be situated within 100 to 200 microns of blood vessels, a very small window of access. Actively dividing groups of cells within the body are even more reliant on nearby vessels, as a constant influx of key cellular building blocks and nutrients is absolutely required for successful proliferation. To support growth, the body must employ angiogenesis in novel tissue areas that do not have a robust, pre-existing vascular network (Carmeliet & Jain, 2000). But what about those instances in which angiogenesis and nutrient acquisition go haywire?

The first documented model of tumor angiogenesis in cancers was proposed in 1971, and since then has become a topic of immense research within the biomedical community. Most research to date that has been conducted on the molecular mechanisms of tumor angiogenesis has utilized xenograft and tumor implantation models. While these

models certainly have helped us elucidate why pathological angiogenesis is so important to tumor growth and stability, these models are inevitably different than naturally occurring, *in vivo* tumors found in patients. A spatiotemporal model of tumor angiogenesis still remains to be explored (Liao & Johnson, 2007).

While we may not yet have a complete picture of the relationship between tumor pathology and angiogenesis, considerable research since 1971 has found that tumor vascularization is heavily reliant on microenvironmental characteristics. Before a tumor is capable of growing in diameter above a 1-2-millimeter threshold, it must acquire more blood vessels than can be readily supplied by its surrounding normal tissue (Carmeliet, 2005). Termed the "angiogenic switch," tumors are capable of moving from a non-angiogenic to angiogenic state through the activation of hypoxia-inducible factors (HIFs), a group of transcription factors that respond to low oxygen conditions in a tissue's microenvironment (Liao & Johnson, 2007). Hypoxia is the condition in which cells extend out of the geographic reach of the oxygen diffusion rate; in tumorous cancers, aberrant cells are capable of hijacking the body's natural angiogenic machinery and bypassing many cellular checkpoints that inhibit uninterrupted cell growth (Carmeliet & Jain, 2000). Many well-known and biomedically relevant HIFs are induced by VEGF and can experience upwards of a 30-fold induction within the span of only a few minutes, indicating that timing is of critical importance to this signaling cascade. HIFs are notoriously implicated in cancer, especially in the instance of tumor growth; as cancerous masses grow in diameter and density, their need for readily-available nutrients and

oxygen becomes larger and forces the hyper-activation of capillary angiogenesis (Carmeliet, 2003).

In an actively angiogenic tumor, vessels develop via intussusception (sprouting from pre-existing vessels) and are influenced by the presence of circulating endothelial progenitor cells. Some tumors are capable of growing around existing vessels, creating what is called a "perivascular cuff." Unlike normal tissue, tumorous vessels tend to appear sporadic and disorganized, often feature uneven diameters, are highly dilated, and have far too many branches to be stable over time. Perhaps unsurprisingly, these vessels often lack a functional basement membrane, a type of extracellular matrix (ECM), and are remarkably leaky, which likely relates to problems with the spatiotemporal expression of VEGF and the angiopoietins (Figure 5.1) (Carmeliet & Jain, 2000; Tonini, Rossi, & Claudio, 2003). *de novo* vasculature is capable of forming both in and around the tumor, yet in both instances will be undeniably abnormal; because these vessels are poorly constructed, the tumor continues to produce pro-angiogenic factors for repair, leading to a perpetual cycle of shabby vessel production. Due to its intriguing expression profile in numerous tumor subtypes, VEGF remains an intriguing (albeit difficult) target for anticancer therapies (Carmeliet, 2005).

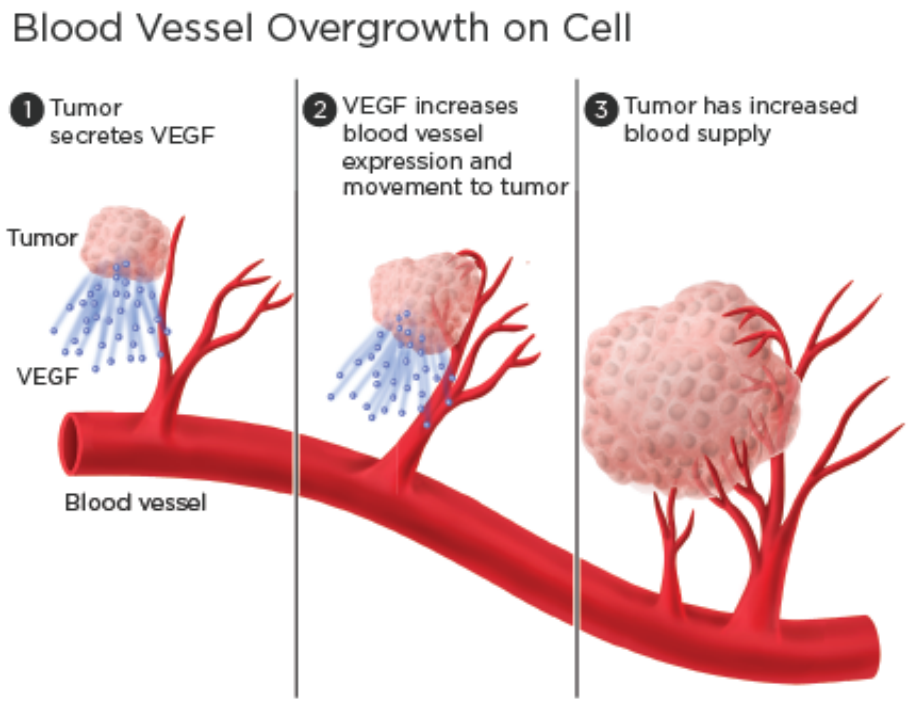

Figure 5.1. Schematic of blood vessel overgrowth in response to VEGF secretion by an actively growing tumor. An increase in blood vessel density leads to increased oxygen and nutrient access, which supports cancer proliferation and possible metastasis. Adapted from "Angiogenesis Inhibitors," 2018.

Originally, it was thought that tumors were capable of producing a unique "angiogenic molecule" that could induce other pro-angiogenic factors to become heavily expressed. We now know, however, that many of the issues in vessel number and quality in tumors are linked to the relative balance in expression of pro and anti-angiogenic factors, suggesting that tumors may grow to remarkable sizes simply because the expression of anti-angiogenic factors cannot keep up with that of pro-angiogenic factors. In most tumors, VEGF and the angiopoietins have poorly coordinated temporal expression patterns. VEGF seems to respond based on gene dosage, implying that

duplication of this growth factor may also activate tumorigenesis via neovascularization. Interestingly, many of the leaky features associated with tumor vessels are linked to the lack of functional perivascular cells surrounding these vessels; without a homogenous layer of endothelial cells, which are functional based on proper VEGF expression, vessels become amenable to plasma leakage and significant ECM damage. It has even been postulated that these endothelial cells are not endothelial at all—instead, they are simply tumorous cancer cells that mimic the expression profiles of endothelial cells via a mechanism of "vasculogenic mimicry." It is important to note that dormant tumors (tumors that are not actively growing in size) tend to express pro-angiogenic and anti-angiogenic factors at relatively equivalent levels, further implicating this precarious balance as a potential therapeutic target (Carmeliet, 2005; Carmeliet & Jain, 2000).

Besides an imbalance between the expression of pro-angiogenic and anti-angiogenic factors in tumor cells, hematopoietic cells and nitric oxide synthases have been implicated in the vascularization of cancerous tissues. Hematopoietic cells are canonically involved in the downregulation of angiogenesis following the completion of neovascularization; in cancerous tissues, however, it is possible that hematopoietic cells found in new vessels do not release enough (or the "right") factors to inhibit further angiogenesis (Carmeliet & Jain, 2000). In addition, nitric oxide synthases, a class of enzymes that are activated by HIFs and help modulate vascular toning through the production of nitric oxide, are capable of acting as endothelial cell survival factors. These factors support apoptosis inhibition and excessive proliferation. While NO synthases certainly play a role in normal angiogenesis, cancerous tissues appear to upregulate and

modulate their expression such that nitric oxide is produced in remarkable excess (Liao & Johnson, 2007). The relatively new discovery of these factors in contributing to pathological angiogenesis in cancerous tumors opens up the possibility for even more therapeutic targets.

VEGF remains as a promising target for molecular cancer therapeutics in patients with actively growing tumors; like other pathological phenotypes associated with adult angiogenesis and neovascularization, however, VEGF plays many roles outside of vascular regeneration and remains a convoluted target. For example, besides its role in activating macrophage recruitment and supporting vessel development, VEGF also has been shown to help "protect" against stroke through mediating neurogenesis and vascular renewal throughout the adult brain. VEGF even plays a role in bone health throughout the adult life cycle, with its conditional deletion relating to severe age-related bone loss. Unsurprisingly, VEGF expression is also necessary for normal alveolar repair in the lungs (Liao & Johnson, 2007). Until we have devised a therapeutic technology in which conditional downregulation of a target transcript may be spatiotemporally controlled in an *in vivo* model, VEGF and other "high-impact," master regulatory transcription factors remain dubious as therapeutic targets.

5.4: Inflammation and Wound Healing—Responding to a Nick

Inflammation, wound healing, and vessel regression are all reactions of the body to diverse environmental cues and stressors. Inflammation, a localized process in which the immune system becomes activated in response to injury or infection, initiates

angiogenic sprouting through the recruitment of mast cells (a type of white blood cell) and the release of vasoactive/angiogenic factors. An inflamed section of tissue must support capillary growth simply because increased quantities of oxygen and nutrients are required for cellular growth and repair. At the molecular level, angiogenic factors function to amplify inflammatory signals by recruiting increasing numbers of lymphocytes to the site of injury. Many of these mechanisms are mimicked in coagulation (wound healing); upon injury, fibrin clots are formed through platelet aggregation before angiogenesis-driven capillary repair may begin. Like white blood cells, platelets store and release large quantities of angiogenic factors that contribute to the regulation of microvessel density, a process that often becomes pathogenic in cancer (Carmeliet, 2003; Luttun, Carmeliet, & Carmeliet, 2003).

Traditionally, it was thought that angiogenesis was the only possible mechanism in which new vascularization could arise in the adult organism; however, it has been demonstrated that EPCs (endothelial progenitor cells) circulate in the peripheral blood stream postnatally and are capable of incorporating into sites of active neovascularization through vasculogenic mechanisms (Kalka et al., 2003). Both larger arterial/venous growth and capillary angiogenesis seem to play distinct roles in adult vascular homeostasis, and as such have been targeted as potential therapeutic topics for numerous biomedical diseases. Some of the most well-known disorders associated with excessive vessel growth include cancer, psoriasis, arthritis, obesity, and asthma, while diseases implicated in insufficient vascular growth and abnormal vessel regression include

ischemia, neurodegeneration, hypertension, pre-eclampsia, osteoporosis, and various respiratory disorders (Carmeliet, 2003).

Chronic wounds, a class of injuries in which healing does not occur properly and during a predictable amount of time, are common in the general population (Tonnesen, Feng, & Clark, 2000). As a remarkably heterogeneous condition, the impaired ability to heal chronic wounds is often associated with other diagnosed conditions, such as that of diabetes and autoimmune dysfunctions. People who suffer from the inability to heal such wounds often produce macrophages that are incapable of producing sufficient levels of cytokines, a class of inflammatory molecules that trigger countless signal cascades for cellular damage repair and infection control. This lack of cytokines also indirectly contributes to decreased rates of angiogenesis at the site of injury; without angiogenesis and the formation of new blood vessels, a damaged tissue is unable to acquire the necessary nutrients and oxygen for successful, timely healing (Wu, Chen, Scott, & Tredget, 2007).

Upon cytokine release and inflammatory induction at a site of injury, normal adult wound healing factors trigger two primary phases of blood vessel integration. The first, which lasts roughly 24 hours in adults, is characterized by increased rates of dilatation (dilation of blood vessels), vessel hyperpermeability, endothelium activation, and diapedesis (a recruitment of macrophages to the site of injury). The second phase, which conversely lasts until the wound is fully healed, features dramatic vascular remodeling and functional changes needed to support new tissue growth. Either of these phases may be disrupted and lead to a wide array of diseased phenotypes; however, many problems

can be traced back to macrophage ineptitude and abnormally low cytokine activity (Majno, 1998).

For external wounds such as cuts or gashes, macrophage recruitment is generally the first step of a localized inflammatory response. Macrophages at the site of injury are responsible for producing cytokines that stimulate the onset of fibroplasia (new tissue formation) and angiogenesis. Fibroblasts, a group of cells that produce collagen and other fibers necessary for tissue regeneration, are responsible for producing new materials for the underlying extracellular matrix in the wound, while new blood vessels supply the oxygen and nutrients required for the fibroblasts to initiate repair. Angiogenic factors that are released by macrophages during wound healing include fibroblast growth factors (FGFs), VEGF, TGF-β, ANG-1, and ANG-2. Interestingly, the initial production of VEGF upon wound healing is initiated by the injured epidermal cells at the site of perturbation (not site-recruited macrophages). These epidermal cells begin to produce VEGF upon induction by HIFs, which subsequently activates all resulting signal cascades (Figure 5.2) (Tonneson, Feng, & Clark, 2000).

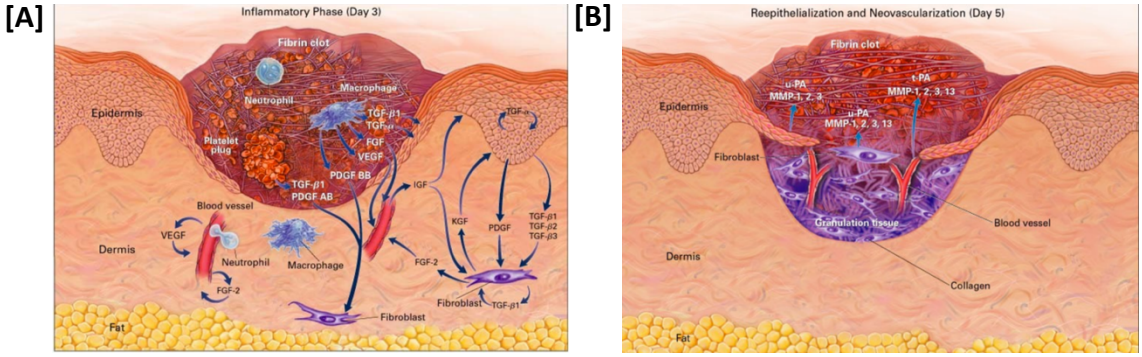

Figure 5.2. Schematic of the molecular signaling cascades associated with traditional wound healing. (A) VEGF, among other angiogenic factors, is secreted by neutrophils both in and underneath the site of injury, which promotes blood vessel growth via activation of angiogenesis. (B) New blood vessels appear directly beneath the site of perturbation as soon as five days after injury. Adapted from Singer & Clark, 1999.

FGF-1 and FGF-2 are released by resident macrophages immediately following tissue disruption. This increase in FGF expression activates the upregulation of VEGF by macrophages within the newly hypoxic microenvironment of the injury and recruits proteolytic enzymes. Among other tasks, the principle role of proteolytic enzymes in a wound is to degrade the underlying, damaged ECM. The deposition of collagen, fibronectin, and elastin by fibroblasts in this region also recruits neighboring blood monocytes and supports their differentiation into macrophages. This rapid increase in macrophage recruitment amplifies the angiogenic signals of ANG-1 and ANG-2 and induces capillary sprouts to form and migrate in response to pre-established FGF/VEGF gradients. This highly regulated spatiotemporal mechanism allows for new blood vessels to establish a provisional matrix, a step necessary for collagen-rich scar tissue to repopulate the tissue (Figure 5.2) (Tonneson, Feng, & Clark, 2000).

The complex interactions between new cells, cytokines, and the underlying provisional matrix (ECM) in a tissue during wound healing is termed “dynamic reciprocity.” Surprisingly, it takes upwards of four days for post-injury capillary sprouts to begin migration out of “mother” vessels surrounding the site of injury. Collagen deposition on top of the provision matrix is inversely associated with capillary density during this time and is also slow to accumulate (Tonnesen, Feng, & Clark, 2000). Even though many of the vessel regeneration mechanisms in the wound healing cascade are reliant on angiogenesis, some endothelial cells in sprouting vessels are derived from circulating stem cells in the blood, suggesting that vasculogenic mechanisms may not be entirely silenced during the adult life cycle (Majno, 1998). Mesenchymal stem cells in the vicinity of active wounds also allow for increased capillary density and increased rates of angiogenesis in wounded tissue, but function through a paracrine (hormonal) mechanism. Like macrophages, fibroblasts, and wounded epithelial cells, mesenchymal stem cells are capable of secreting high levels of VEGF-α and ANG-1, indicating that the adult body has numerous mechanisms in which angiogenesis may become activated upon external injury (Wu, Chen, Scott, & Tredget, 2007).

5.5: Inflammation in the Asthmatic Airway

At first glance, there is no reason to suspect that vascular remodeling could possibly play a role in the symptomology of a common pulmonary disorder like asthma; however, along with mucous metaplasia, two common features of the asthma phenotype are smooth muscle hypercontraction and increased airway thickness, both of which are in

part the result of abnormal tissue vascularization. The acquisition of abnormally large networks of vessels in asthmatic airways, regardless of phenotypic severity, allows for increased blood flow and microvascular permeability; as capillaries become hyperpermeable and "leaky," plasma accumulates on the underlying ECM of the lungs and contributes to the smooth muscle contractions characteristic of an asthma attack (Ribatti et al., 2009). Remodeling of the airway vasculature in asthmatic patients also contributes to the overall increased airway wall thickness, luminal narrowing, and hyperresponsiveness to inhaled environmental stimuli characteristic of the clinical asthma subtypes (Feltis et al., 2006).

At the molecular level, the abnormal release of fibrogenic growth factors like TGF-β, tumor necrosis factor-α, and basic FGF all contribute to the high rates of angiogenesis found in the lungs (Feltis et al., 2006). Unlike in other tissue systems, VEGF plays many distinct roles in the lungs, including that of regulating capillary hyperpermeability *in vivo.* VEGF is produced and secreted by alveolar endothelial cells, bronchial endothelial cells, smooth muscle cells, fibroblasts, and macrophages, and it affects pulmonary immune responses to environmental stimuli (such as pollen or pathogens) by amplifying the inflammatory hyperactivity characteristic of the TH-2 asthma subtype. Cytokines also secreted by these cells, such as IL-4, IL-5, and IL-13, are also secreted by smooth muscle cells that line the airways and enhance VEGF production (Figure 5.3). Due to its remarkable expression in almost every cell type in the adult airway and elsewhere in the body, VEGF has proven to be a difficult target for biomedical inquiry (Ribatti et al., 2009).

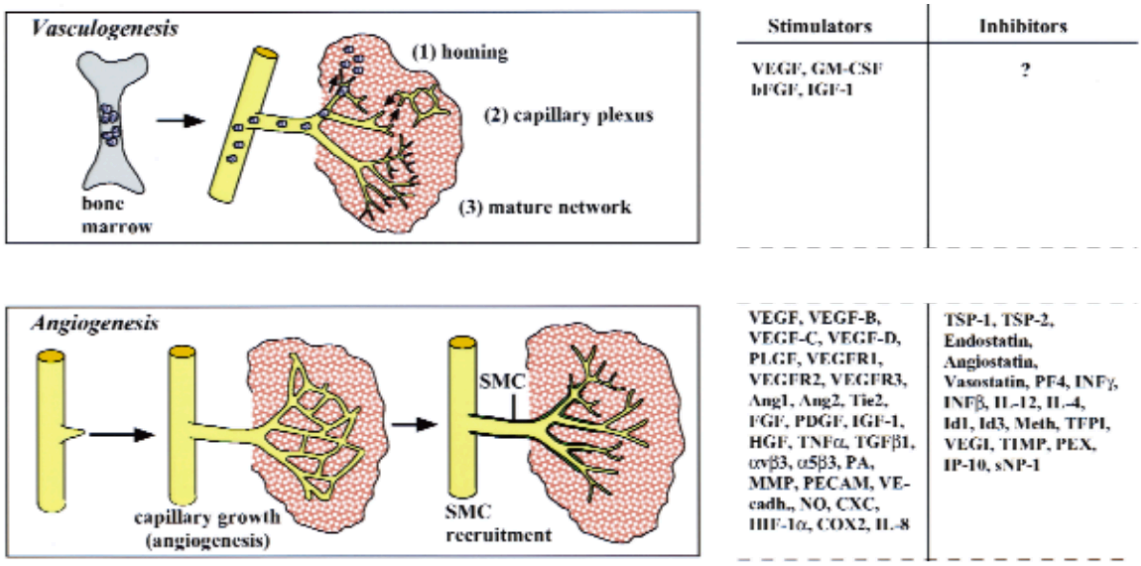

Figure 5.3. Schematic of the general vasculogenic and angiogenic mechanisms promoted by the airway epithelium of the lungs. Angiogenic factors like the angiopoietins (ANG) and the vascular endothelial growth factors (VEGF) play a crucial role in the initiation and propagation of neovascularization in the normal epithelium. Adapted from Carmeliet, 2000.

Another family of angiogenic factors implicated in the general asthma phenotype is that of the angiopoietins. Coupled with high levels of VEGF, ANG-2 supports a rapid increase in the diameter of lung capillaries and induces angiogenic endothelial cells to proliferate and migrate (Ribatti et al., 2009). ANG-1, an angiopoietin that functions alongside VEGF expression, is responsible for maintaining quiescence and stability of the mature adult vasculature throughout the lungs. Disruption of ANG-1 expression has been found to lead to aberrant VEGF reactivation and subsequent hyperactivity of angiogenic sprouting, suggesting that ANG-1 may act as an indirect regulatory molecule for dampening the extreme effects of high VEGF levels in the lungs (Feltis et al., 2006). Vessel sustainment, a process that is aberrantly regulated in asthmatics, is also controlled

by relative concentrations of VEGF and ANG-1, as asthmatic airways have increased vessel numbers overall, larger numbers of small vessels (indicating high angiogenic activity), and abnormally "leaky" vessels that promote extreme tissue swelling. Interestingly, vascular sprouts are not unique to asthmatics in the airway; rather, it is the incorrect timing and intensity of angiogenic signaling in patient airways that contribute to the phenotype of smooth muscle hypercontraction (Feltis et al., 2006).

A major area of biomedical research for the development of asthma therapeutics has focused on how cytokine release and inflammatory stimulation relate to the accumulation of angiogenic factors in the asthmatic airway (Ribatti et al., 2009). Pulmonary eosinophil infiltration (macrophage recruitment) has been successfully reduced in a murine model through the injection of anti-VEGF receptor antibodies, more commonly called by its trade name, Avastin (Lee, Kwak, & Song, 2002). Given the molecular homology of adult angiogenesis throughout the body, it is unsurprising that Avastin is also a potent chemotherapeutic. While direct inhibition of VEGF activity would in theory allow for an overall dampening of angiogenic hyperactivity in the lungs, VEGF is important for too many other normal functions in the airway (as well as throughout other organ systems in the adult body) and would likely lead to dramatic side effects *in vivo*. It is important to note that besides dampening the effects of smooth muscle hypercontraction, inhaled corticosteroids also suppress VEGF transcription *in vitro,* suggesting that some of their efficacy in asthmatic patients may be related to the added effects of VEGF downregulation and smooth muscle relaxation (Ribatti et al.,

2009). Overall, much remains to be discovered concerning the potential targeting of angiogenic factors in the treatment of asthma.

5.6: Menses & Endometriosis

Perhaps the most obvious example of the role of angiogenesis in the adult life cycle can be found in the female menstrual cycle. Physiological angiogenesis within the female reproductive system occurs in the ovaries, endometrium, and placenta during pregnancy. VEGF is expressed at varying levels throughout the estrus cycle, and placental growth factor (P1GF) is so named because of its specific role in supporting angiogenesis between mother and child during gestation. Like mesenchymal stem cells in wound healing, *VEGF* mRNA expression in the endometrium (mucous membrane lining the uterus) is activated through hormonal regulation cascades involving estradiol and progesterone. Much like in the human airway, cytokines are also implicated in the upregulation of angiogenic factors; uterine natural killer (NK) T cells, small granular cells that reside in the proliferative endometrium, produce high levels of cytokines during the secretory phase of the estrus cycle that consequently activate angiogenic activity in the uterus. While the normal endometrium suppresses *VEGF-C*, *ANG-1*, and *ANG-2* expression, NK T cells express multiple pro-angiogenic factors that may be principally implicated in the cyclical endometrial regeneration of menses. Furthermore, VEGF promotes NK T cell adhesion to underlying endothelial cells in the female reproductive tract, suggesting that NK cells have a unique role in supporting endometrial angiogenesis (Li et al., 2001).

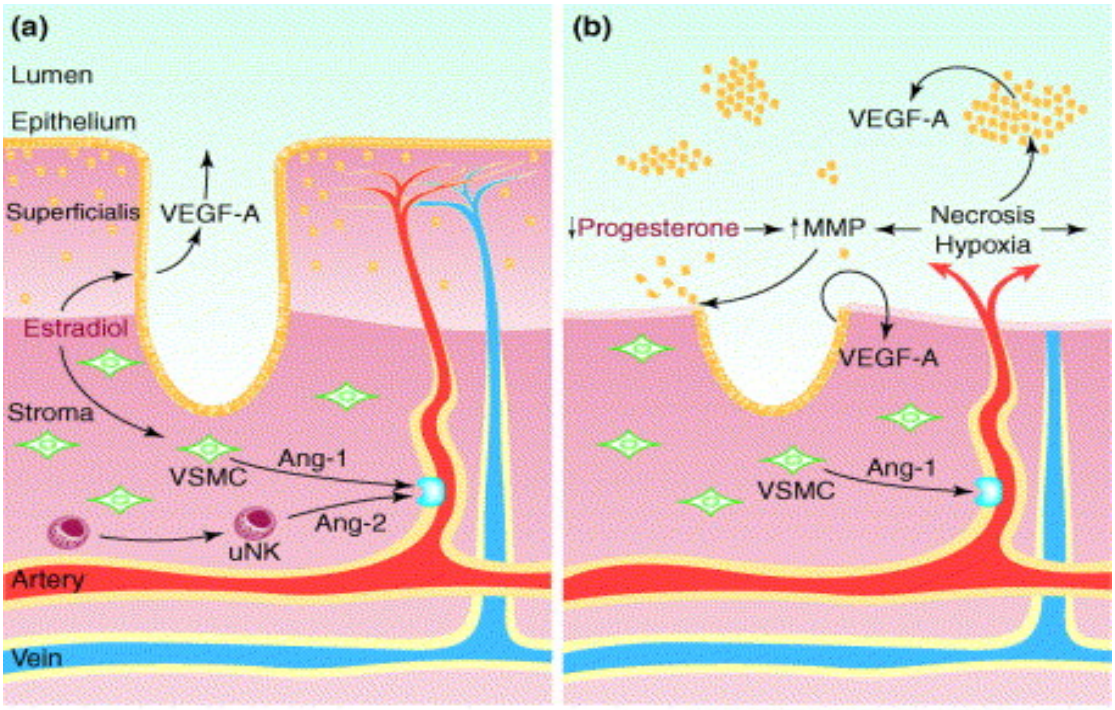

Figure 5.4. Schematic of molecular signaling cascades associated with (a) the proliferative phase and (b) the end of the secretory phase of menstruation. VEGF-α and ANG-1/ANG-2, like other parts of the body reliant on neovascularization, play a major role in response to cyclical hormonal cues, such as that of estradiol and progesterone. Adapted from Smith, 2001.

Endometriosis, a reproductive disorder characterized by excessive endothelial growth outside of the uterus, is associated with dysfunctional inflammatory reactions within the endothelium during menses. Characterized by chronic cystic lesions throughout the reproductive tract, endometriosis occurs when exfoliated endometrium becomes aberrantly deposited in the peritoneal cavity following the completion of menstruation. This ectopic deposition of tissue invokes a local inflammatory response in which macrophages release excessively high concentrations of cytokines and subsequently invoke angiogenic neovascularization. VEGF is found in high

concentrations in the peritoneal fluid of patients with endometriosis, even though cyclic variations occur based on menstruation phase (higher in proliferative phase compared to secretory phase) (McLarne, Prentice, Charnock-Jones, & Smith, 1996). Much like other adult physiological mechanisms that rely on neovascularization, it seems as if angiogenesis is not the only process implicated in endometriosis—post-natal vasculogenesis and circulating endothelial progenitor cells have been implicated in endometriosis pathogenesis and high expression of numerous angiogenic factors in the uterus (Laschke, Giebels, & Menger, 2011).

5.7: Summary

What could cancer, wound healing, airway hypervascularization, and endometriosis possibly have to do with the homology of development? While pathological adult conditions such as these may not seem to relate to the homologous pathways that have defined the evolution of the animal kingdom, it appears as if many of the master regulators that play a major role in development are often reactivated and misused in adult diseases. Transcription factors like VEGF and the angiopoietins are not meant to be silenced during adulthood; however, because of their diverse roles during development and beyond, their aberrant expression often leads to large-scale phenotypic effects that relate to clinically relevant diseased phenotypes. It is foolish to think that developmental proteins become irrelevant following birth; because of the masterful tinkering of evolution, mammalian molecular circuitry remains both beautifully intricate and devastatingly connected.

5.8: Chapter Five References

Angiogenesis Inhibitors. (2018). *LUNGevity Foundation.* Retrieved from https://lungevity.org/ for-patients-caregivers/lung-cancer-101/treatment-options/angiogenesis-inhibitors

Carmeliet, P. (2003). Angiogenesis in health and disease. *Nature medicine*, *9*(6), 653.

Carmeliet, P. (2000). Mechanisms of angiogenesis and arteriogenesis. *Nature medicine*, *6*(4), 389.

Carmeliet, P. (2005). VEGF as a key mediator of angiogenesis in cancer. *Oncology*, *69*(Suppl. 3), 4-10.

Carmeliet, P., & Jain, R. K. (2000). Angiogenesis in cancer and other diseases. *Nature*, *407*(6801), 249.

Feltis, B. N., Wignarajah, D., Zheng, L., Ward, C., Reid, D., Harding, R., & Walters, E. H. (2006). Increased vascular endothelial growth factor and receptors: relationship to angiogenesis in asthma. *American journal of respiratory and critical care medicine*, *173*(11), 1201-1207.

Jain, R.K. & Munn, L.L. (2000). Leaky vessels? Call Ang1! Nat. Med. 6, 131–132

Kalka, C., Tehrani, H., Laudenberg, B., Vale, P. R., Isner, J. M., Asahara, T., & Symes, J. F. (2000). VEGF gene transfer mobilizes endothelial progenitor cells in patients with inoperable coronary disease. *The Annals of thoracic surgery*, *70*(3), 829-834.

Laschke, M. W., Giebels, C., & Menger, M. D. (2011). Vasculogenesis: a new piece of the endometriosis puzzle. *Human reproduction update*, *17*(5), 628-636.

Li, X. F., Charnock-Jones, D. S., Zhang, E., Hiby, S., Malik, S., Day, K., ... & Loke, Y. W. (2001). Angiogenic growth factor messenger ribonucleic acids in uterine natural killer cells. *The Journal of Clinical Endocrinology & Metabolism*, *86*(4), 1823-1834.

Liao, D., & Johnson, R. S. (2007). Hypoxia: a key regulator of angiogenesis in cancer. *Cancer and Metastasis Reviews*, *26*(2), 281-290.

Luttun, A., Carmeliet, G., & Carmeliet, P. (2002). Vascular progenitors: from biology to treatment. *Trends in cardiovascular medicine*, *12*(2), 88-96.

Majno, G. (1998). Chronic inflammation: links with angiogenesis and wound healing. *The American journal of pathology*, *153*(4), 1035.

McLaren, J., Prentice, A., Charnock-Jones, D. S., & Smith, S. K. (1996). Vascular endothelial growth factor (VEGF) concentrations are elevated in peritoneal fluid of women with endometriosis. *Human Reproduction*, *11*(1), 220-223.

Ribatti, D., Puxeddu, I., Crivellato, E., Nico, B., Vacca, A., & Levi-Schaffer, F. (2009). Angiogenesis in asthma. *Clinical & Experimental Allergy*, *39*(12), 1815-1821.

Singer, A. J., & Clark, R. A. (1999). Cutaneous wound healing. *New England journal of medicine*, *341*(10), 738-746.

Smith, S. K. (2001). Regulation of angiogenesis in the endometrium. *Trends in Endocrinology & Metabolism*, *12*(4), 147-151.

Tonini, T., Rossi, F., & Claudio, P. P. (2003). Molecular basis of angiogenesis and cancer. *Oncogene*, *22*(42), 6549.

Tonnesen, M. G., Feng, X., & Clark, R. A. (2000, December). Angiogenesis in wound healing. In *Journal of Investigative Dermatology Symposium Proceedings* (Vol. 5, No. 1, pp. 40-46). Elsevier.

Tonnesen, M. G., Feng, X., & Clark, R. A. (2000, December). Angiogenesis in wound healing. In *Journal of Investigative Dermatology Symposium Proceedings* (Vol. 5, No. 1, pp. 40-46). Elsevier.

Wu, Y., Chen, L., Scott, P. G., & Tredget, E. E. (2007). Mesenchymal stem cells enhance wound healing through differentiation and angiogenesis. *Stem cells*, *25*(10), 2648-2659.

Chapter 6

CONCLUSION

ETS TFs, VEGF, and the angiopoietins are certainly not the only examples of how homologous molecules impact development and disease etiology. Homology colors the very core of what it means to understand multicellularity—without it, the transition from "bench to bedside," i.e. the connectivity between basic and translational research, would be inaccessible and unattainable. The study of evolution need not be solely practical for the sake of organismic biology and ecological conservation; a doctor's office seems to be just as amenable to evolutionary discovery as a paleontological dig.

If we are to continue bridging the gap between human evolution and medicine through the scope of homology, then what is the complete narrative with which we must operate? And how do we decide which lens is the right fit for "viewing" homology? Countless years of utilizing model organism systems has taught us that homology is simply the stamp left behind from "our lowly origins." At the time of the Cambrian explosion, the genomic framework for the animal kingdom was established in a simple bilateral organism with enormous evolutionary potential. Amid constraints from an ever-changing environment, the animal genome diversified through mutational events coupled

with the beautiful simplicity of natural selection. Gene duplications and "trial-and-error" alterations of CREs likely allowed for incremental changes in gene regulatory networks that continued to support rapid speciation and diversification. This model, which Darwin aptly titled "Descent with Modification," stipulates that the fundamental building blocks of the animal kingdom have been contained in an "ancient genetic toolkit" that is found within each extant species today. The beauty of this model of evolution is that it largely explains why model organism studies remain so effective. We still know relatively little about the regulatory capacity of the human genome; however, it is through the holistic lens of developmental homology that these patterns of evolutionary novelty are most likely to be elucidated.

At first glance, my research project at Cincinnati Children's Hospital would seem to be unrelated to evo-devo and homology. What could the identification of a new progenitor population implicated in zebrafish vasculogenesis and angiogenesis have to do with the larger patterns of animal evolution? If homology tells us anything, it is that novel discoveries in one organism are often found in others throughout the phylogenetic tree. When I presented my research at a poster symposium at the conclusion of the summer, I met a mouse biologist who expressed interest in determining if PAC cells are found in a murine model of circulatory development—little did she know that her interest in applying my findings to research in her own lab was grounded in the fundamental principle of animal homology!

My project continues to function as a preliminary investigation into how PAC cells shape zebrafish vascular development; however, should our laboratory successfully

demonstrate that these cells are an evolutionarily conserved population, our findings would have the potential to impact future studies on biomedical vascular regeneration. Given what we know about the pervasive nature of molecular homology, it would be unsurprising to find that PAC cells have human analogs. The simple knowledge that so many deeply homologous transcription factors shape vascular development throughout the animal kingdom indicates that such discoveries are both entirely feasible and significantly important.

When I began formulating a trajectory for this thesis, I had very little foresight into the awe-inspiring discoveries that I would make about animal homology. I was well aware that homology is an overarching facet of the evolution of life itself; however, I was completely unaware of how much biomedical inquiry has been reliant on the simple homology of eukaryotic protein networks. I was dumbfounded to learn that the vastly disparate topics of cancer, wound healing, asthma, and endometriosis are all connected via a few homologous transcription factors—how humbling it is to witness the beautiful design of creation through such simple molecules. The writing of this thesis has been a profoundly religious experience for me. Who would have guessed that a narrative of homology would end up being my most poetic prayer.

Moment of peace like brief arctic bloom
Red-gold ripple of the sun going down
Line of black hills makes my bed
Sky full of love pulled over my head
In this world of wonders

"World of Wonders"
--Bruce Cockburn