

Carnegie Mellon University
Mellon College of Science

Thesis

Submitted in partial fulfillment of the requirements for the degree of Ph.D. in Biological Sciences

Title

Presented by

Accepted by the Department of Biological Sciences

Major Professor

Date

Department Head

Date

Approved by the MCS College Council

Dean

Date

Evolution of DNA Recognition in a Tbrain Transcription Factor and Its Impact on Larval Neurogenesis in Echinoderms

A thesis presented by

Alys M. Cheatle Jarvela
B.S. The Catholic University of America

to

The Department of Biological Sciences
Carnegie Mellon University
Pittsburgh, Pennsylvania

in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in the field of
Biological Sciences

August 19th, 2014

Thesis Advisor: Dr. Veronica Hinman

Abstract

Study of Gene Regulatory Networks (GRNs) is essential for the understanding of developmental processes because GRNs describe the genetic specification mechanisms that instruct an egg to become a complex organism. Additionally, because cell types, organs, tissues, and other morphological features are specified during development, changes to the underlying GRNs lead to differences in such features. Therefore, an understanding of developmental GRNs is required to understand how morphology evolves.

Echinoderms offer an attractive group of model organisms for the study of GRN evolution. A variety of echinoderm species have publically available genomic or transcriptomic information, their development is well characterized, and for some species and cell types extensive GRNs have already been elucidated, all of which greatly facilitate comparative approaches. Furthermore, echinoderms are deuterostomes, just as vertebrates are. Therefore, studies of echinoderm developmental GRNs can enhance our understanding of vertebrate evolution and development.

Here, we take multiple approaches to understanding the processes of developmental GRN evolution. We first survey recent literature with the aim of ascertaining the significance and prevalence of transcription factor changes to GRN evolution and morphological novelty. Next, we present a recent publication from our lab which describes a previously uncharacterized source of modularity within orthologous Tbrain (Tbr) transcription factor DNA-binding abilities. We maintain that this type of modularity could be an important contributor to the evolution of novel features. After this, we characterize a new GRN for the specification of serotonergic neurons in the sea star dorsal ganglia. This structure exhibits versatile morphologies among echinoderms and is important for understanding the origins of the vertebrate forebrain as it is thought to be similar to the ancestral deuterostome central nervous system precursor. Finally, we propose a new line of research intended to determine whether modular DNA-binding among orthologous Tbr transcription factors has impacted the evolution of this interesting neuronal structure.

Acknowledgments

First, I would like to thank my advisor, Dr. Veronica Hinman. From the beginning, Veronica took a chance on me. I came from a chemistry background and developed a project that combined my biochemical knowledge with developmental and evolutionary biology. Veronica not only obligingly let me, an upstart first-year student, go way off the beaten path, but also showed me the ropes of these other disciplines. As a result, I learned an extraordinary amount during my time as a student in Veronica's lab. Additionally, Veronica and I were able to produce a really unique, interdisciplinary body of work that was well worth the initial risks.

I also thank my thesis committee, Dr. Aaron Mitchell, Dr. Charles Joel McManus, and Dr. Mark Rebeiz for their support and discussions that greatly improved my work. Aaron in particular has always motivated me to work even harder and encouraged me to think big when planning the rest of my career. Joel and Mark have also provided thoughtful critiques of my work through frequent group lab meetings. My publications have benefitted from this additional feedback in between committee meetings.

I thank all of the past and current members of the Hinman lab, especially Brenna McCauley and Kristen Yankura who helped me settle into the lab and taught me everything they knew. Both were always generous with their time and emotionally supportive while I got my project off the ground. I am also grateful to Adam Foote and Greg Cary, who have been both entertaining friends and helpful colleagues. Greg has been instrumental in catching me up to the genomics era in the short time since he has joined our lab.

I am in debt to many others throughout the Mellon Institute and beyond who have provided advice and technical assistance throughout my time here. I especially thank Mark MacBeth for teaching me the basics of protein purification and for the advice he provided as a former member of my thesis committee. I am grateful to Bruce Armitage and Anisha Gupta from the Department of Chemistry for training me on their SPR machine, thoughtful troubleshooting discussions, and guidance with data analysis. Finally, I thank Martha Bulyk and Anastasia Vedenko at Harvard Medical School for a fruitful collaboration and for assisting me with Protein Binding Microarray studies.

My family has been amazingly supportive and understanding throughout this process. I thank my sister Nicole, for keeping me balanced and making sure I had fun every once in a while. I also thank my parents, not only for their recent encouragement, but also for instilling

qualities in me throughout my childhood that have been important to my success. I thank my mom for teaching me to stay strong when things are difficult, and my dad for teaching me that hard, careful work always pays off in the end.

Finally, I thank my husband, Tim Jarvela. He has been a huge part of every step of my graduate school experience, from recruiting me to Carnegie Mellon to keeping me sane as I wrote my dissertation. No matter what I have needed, ranging from a buddy for a pre-lab five-mile run, to help troubleshooting the DuoScan confocal, he always makes sure I'm taken care of. Most importantly, he believed in me and encouraged me through the hardest parts of my thesis work. I am immeasurably lucky to have him as a colleague, best friend, and spouse.

List of Figures

Figure 2.1: Phylogeny and Sequence Alignment for Pm and SpTbr Tbox-DNA Binding Domains.....	23
Figure 2.2: Position Weight Matrices Depicting Binding Specificities of Tbr Orthologs.....	28
Figure 2.3: Steady State Affinity Evaluations for Tbr DNA Binding Domains.....	30
Figure 2.4: <i>PmTbr</i> Can Use the Primary and Secondary Sites <i>in vivo</i> to Drive Reporter Gene Expression Interchangeably Except when Tbr Levels are Reduced.....	35
Figure 2.5: Secondary Tbr Reporter has Reduced Expression Compared to OtxG in the Ectoderm when Tbr Levels are Declining	38
Figure 2.6: Modular Binding of Tbr may Allow for Diverse Transcriptional Responses during Development and Allow for Greater Evolvability.....	43
Supplemental Figure 2.1: T-box Alignment used in Phylogenetic Tree Construction.....	48
Supplemental Figure 2.2	51
Supplemental Figure 2.3	51
Supplemental Table 2.2	51
Figure 3.1: Lhx2/9 is required for the specification of serotonergic neurons.....	56
Figure 3.2: Lhx2/9 cells are proliferating neural progenitors.....	58
Figure 3.3: <i>Lhx2/9</i> cells originate from <i>soxc</i> -expressing neural stem cells.....	60
Figure 3.4: Lhx2/9 expression is not dependent on components specific to the ciliary band neuron gene regulatory network.....	62
Figure 3.5: BMP2/4 Signaling is needed for <i>lhx2/9</i> expression and provides dorsal regional identity.....	64
Figure 3.6: <i>Lhx2</i> requires an anterior pole domain defined by <i>foxq2</i> for specification and also to differentiate into neurons.....	66
Figure 3.7: Six3 does not affect the initial specification of the apical pole domain, but instead maintains a proliferative <i>lhx2/9</i> territory free from both <i>foxq2</i> expression Wnt signaling.....	69
Figure 3.8: Model of sea star serotonergic neurogenesis.....	74
Supplemental Figure 3.1: PmLhx2/9 is an ortholog of vertebrate Lhx2 and Lhx9.....	76
Figure 4.1: Preliminary Data.....	82

Table of Contents

List of Figures.....	iv
Chapter 1: Evolution of Transcription Factors Modifies Developmental Gene regulatory Networks to Produce Novelty.....	1
1.1 Preface.....	1
1.2 Abstract.....	1
1.3 The Gene Regulatory Network Evolution Debate	2
1.4 The Structure and Function of Transcription Factors are Inherently Modular	4
1.5 The Rise and Expansion of Metazoan Transcription Factor Families	5
1.6 Evolution of Protein-Protein Interactions	7
1.7 Evolution of DNA-Binding Specificity	9
1.8 Evolution of Post-Translational Modifications.....	12
1.9 Remaining Questions: The FoxP2 Mystery	13
1.10 Concluding Remarks.....	14
Chapter 2: Modular Evolution of DNA Binding Preference of a Tbrain Transcription Factor Provides a Mechanism for Modifying Gene Regulatory Networks.....	16
2.1 Preface.....	16
2.2 Abstract.....	17
2.3 Introduction.....	17
2.4 Results.....	21
2.4.1 Sea Urchin and Sea Star Tbr are orthologous to Mouse Eomes	21
2.4.2 Sea Urchin and Sea star Tbr orthologs have different DNA binding preferences	24
2.4.3 <i>SpTbr</i> and <i>PmTbr</i> Maintain Similar Affinity for the Conserved Primary Site, but Differ Significantly in their Affinity for <i>PmTbr</i> 's Secondary Site.....	29
2.4.4 The Secondary Site Can Substitute for the Primary Site <i>in vivo</i> when Tbr Levels are High, but not when they are Reduced	32
2.4.5 The Secondary Site Responds Faster to Tbr's Endogenous Temporal Gradient	36
2.5 Discussion.....	39
2.6 Methods.....	43
2.7 Acknowledgements.....	47
Chapter 3: The GRN for specification of anterior dorsal ganglia in <i>Patiria miniata</i> sea star embryos.....	52
3.1 Preface.....	52
3.2 Abstract.....	52
3.3 Introduction.....	53

3.4 Results.....	55
3.4.1 <i>PmLhx2/9</i> is required for the specification of serotonergic neurons.....	55
3.4.2 <i>Lhx2/9</i> cells are neural precursors that proliferate by symmetric divisions and generate neurons through asymmetric divisions.....	57
3.4.3 <i>Lhx2/9</i> cells originate from <i>soxc</i> -expressing neural stem cells.....	58
3.4.4 The Dorsal Ganglia Does Not Require the Ciliary Band Specification GRN.....	61
3.4.5 Dorsal/Ventral patterning is needed to specify the Dorsal Ganglia.....	62
3.4.6 Dorsal ganglia require an apical pole domain, defined by <i>foxq2</i> and a lack of Wnt expression.....	65
3.4.7 <i>Six3</i> Regulates the Size and Position of the Dorsal Ganglia through Wnt Signaling, but not its Initial Specification.....	67
3.5 Discussion.....	70
3.6 Methods.....	74
Chapter 4: Conclusions and Future Directions.....	77
4.1 Conclusions.....	77
4.2 Background and Rationale: Tbr Function within Echinoderms and across Deuterostomes.....	77
4.3 Experimental Approach and Preliminary Data: Comparative ChIP-seq.....	79
Appendix 1: ChIP Method.....	84
References.....	88

Chapter 1: Evolution of Transcription Factors Modifies Developmental Gene Regulatory Networks to Produce Novelty

1.1 Preface

I wrote this introduction to my thesis with the intent of also submitting it as a review article to the journal *EvoDevo*. For this reason, the breadth of topics and research covered is considerably broader than the research described in Chapter 2, and does not introduce the research discussed in Chapter 3. However, because I included a brief discussion of my own recent publication, Cheate Jarvela et al., 2014, in Section 1.7, this review serves the important purpose of demonstrating where my work fits into the broader field of Evolution and Development and which existing gaps in knowledge I have been able to fill.

1.2 Abstract

The form that an animal takes on during development is directed by gene regulatory networks (GRNs). GRNs interpret maternally deposited signals to instruct cell-fate decisions and positions, leading to the ultimate arrangements of organs and tissues in the completed organism. Modifications to these networks have allowed for the evolution of multicellular animals as well as the wide range of metazoan diversity that exists today. It is well established that gene regulatory networks primarily evolve through changes to *cis*-regulatory DNA, and it was historically theorized that changes to the transcription factors that bind to these *cis*-regulatory modules contribute to this process only rarely. A growing body of evidence suggests that changes to the coding regions of transcription factors play a much larger role in the evolution of developmental gene regulatory networks than originally imagined. Here, we review the recent works that have led to this unexpected change in the field of Evo-Devo, and consider the implications these studies have had on our understanding of the evolution of developmental processes.

1.3 The Gene Regulatory Network Evolution Debate

Gene regulatory networks (GRNs) explain the gene expression states that allow a cell to take on a particular fate (Davidson 2010). In development, these models describe the mechanisms that take an egg and its localized maternal determinants to an organism with properly placed tissues and fully differentiated cells. GRNs are visually represented as diagrams with nodes and edges, such that each node represents a transcription factor or signaling molecule and edges to or from that node depict regulatory inputs to or from other genes in the network. Biologically, GRNs are composed of transcription factor proteins and the *cis*-regulatory module (CRM) DNA they interact with. This interaction allows the transcription factor to positively or negatively influence the expression of the CRM's downstream gene. Because these networks serve as a blueprint for making a particular cell type or structure, changes to these networks result in the evolution of animal morphology.

There has been much debate surrounding the mechanisms by which GRNs evolve. Changes to *cis* regulatory modules have historically been considered as the dominant source of GRN evolution and this idea continues to be supported by new data in the genomics era (reviewed Rebeiz and Williams, 2011; Rubinstein and de Souza, 2013; Wittkopp and Kalay, 2012; Wray, 2007). While it is difficult to uncover and dissect CRMs and subsequently associate them with a discernable functional divergence, nevertheless numerous examples have been unearthed (e.g Arnoult et al., 2013; Guerreiro et al., 2013; Hinman et al., 2007). In recent years, genome-wide experiments, such as ChIP-seq, and computational approaches have also been instrumental in understanding the contribution of regulatory DNA evolution. Using such methods, Schmidt and colleagues demonstrated that in the livers of five different vertebrate species, binding preferences of the transcription factors CEBPA and HNF4a are conserved as are many specific binding events. Yet there were also many instances of lineage-specific gains and losses of binding events, suggesting rapid turnover in *cis* regulatory sequence (Schmidt et al. 2010). Additionally, functional noncoding sequences as a whole, which frequently have regulatory functions, turn over quickly too (Meader et al. 2010).

Furthermore, much historical evidence suggests that transcription factors are incredibly well conserved over evolutionary time. First, it was demonstrated that the Hox transcription factor cluster is conserved in both sequence and function, patterning the body axis of organisms

as disparate as insects and vertebrates (McGinnis et al. 1984; Duboule and Dollé 1989). This was followed by numerous and particularly compelling functional-equivalence studies, in which transcription factors from widely disparate taxa were shown to rescue knock-out phenotypes and create similar structures e.g (McGinnis et al. 1990; Halder et al. 1995; Wang et al. 2002). In fact, the realization that largely overlapping sets of transcription factors drive the development of essentially all metazoans surveyed lead to the concept of the “toolkit for development” and the birth of Evo-Devo as a discipline (Carroll 2005; Carroll 2008).

Even prior to this breadth of experimental evidence in support of CRM evolution as the primary driver of GRN evolution, many theorized that this would be the case (Britten and Davidson 1971). The logic of this argument is as follows: Transcription factors are pleiotropic, meaning that they are multifunctional, and thus mutations that might result in adaptive changes in one context will almost certainly be detrimental to the organism in others. Meanwhile, CRMs are highly modular. A single gene frequently will be regulated by a separate CRM in each of its temporal and spatial expression domains, and therefore one context can easily be altered without affecting the others. Even individual CRMs are modular. CRMs typically contain multiple binding sites for several different transcription factors, each of which can be mutated individually. Therefore, it is commonly accepted that transcription factors are under much more constraint than CRMs and therefore are less free to evolve changes in sequence and function (Stern 2000; Wray 2007).

More recently, it has been argued that transcription factors also have the capacity to be modular, and therefore may be an underappreciated source of developmental GRN evolution (Lynch and Wagner 2008). These authors maintained that many aspects of protein expression and structure permit their evolution by reducing pleiotropy. For example, use of tissue-specific splice forms and changes to protein-protein interactions, which will only be relevant in tissues where both interacting proteins are expressed, both offer mechanisms to reduce the pleiotropy associated with transcription factor changes. Recent work has provided even more support for these ideas in addition to revealing unpredicted sources of modularity. Just as genomic approaches have allowed for increased understanding of the contributions of CRM mutations to GRN evolution, bioinformatic, genome-wide, and other novel techniques have also been instrumental to gaining a better insight into the ways in which transcription factors evolve. Here, we survey recent experimental findings in support of transcription factor evolution as an

important way in which GRNs can be rewired and the various mechanisms that allow for such evolution. In particular, we focus on modular protein changes that seem to be favored by evolution and therefore could occur in many other systems. We also discuss the implications of these changes on the evolution of development.

1.4 The Structure and Function of Transcription Factors are Inherently Modular

The basic biochemical function of a transcription factor is two-fold: to recognize and bind a short, specific piece of DNA within a regulatory region, and to recruit or bind other proteins relevant to transcriptional regulation, such as other transcription factors, chromatin remodeling proteins, and basic RNA polymerase machinery. The first function, DNA binding, directs the transcription factor to its target genes. The second allows it to elicit changes in transcriptional levels by influencing the stability of the transcriptional apparatus or the chromatin state. Combined, these functions allow transcription factors to influence gene expression. At the structural level, transcription factor proteins contain discrete domains for exerting these functions, known as DNA-binding domains and protein-protein interaction domains. Some have more than one of each, and others may perform both functions via a single domain. Because transcription factors have such functional units, which may individually acquire mutations and be lost or gained over time, they are modular just as CRMs are and thus have opportunities to evolve in ways that minimize pleiotropy.

Exon shuffling allows for the creation of new genes by piecing together existing functional domains. This mechanism has been known to create novel genetic toolkit components, such as the developmentally important signaling molecule, hedgehog (Adamska et al. 2007). The components of this protein exist in sponges and cnidarians, but not as a complete hedgehog protein, suggesting that this gene is the product of exon shuffling. Likewise, a comprehensive study of domain-shuffling in deuterostomes revealed that a handful of transcription factors in the vertebrate lineage acquired new transactivation domains that may have been important for the evolution of vertebrate-specific features (Kawashima et al. 2009). This mechanism also allowed the COE family of transcription factors to diverge through a tandem duplication of part of the helix-loop-helix domain at the base of the vertebrate lineage (Daburon et al. 2008). It is suggested that this change might allow vertebrate COE orthologs to

make a wider variety of heterodimer pairings. Importantly, such rearrangements occur without necessarily altering the existing components, and therefore might take place without disrupting all of the ancestral functions.

Alternative splicing can also evolve to produce lineage-specific variants of transcription factors in a modular way from the existing structural composition. This is thought to be particularly useful in the evolution of developmental GRNs because different variants can be limited to a particular tissue or developmental stage (reviewed Lynch and Wagner, 2008). Alternative splicing has been shown to be able to alter DNA-binding domain architecture and potentially also DNA-binding specificity in a tissue-specific manner (Taneri et al. 2004). More recently, Blekhman and colleagues used RNA-seq to study transcript levels among three primate species and found that the expression of particular splice forms differs between lineages and sexes (Blekhman et al. 2010). In *Drosophila*, sex-specific abdominal pigmentation patterns require gender-specific splice forms of the transcription factor *Dsx*, such that the female form activates gene expression and the male form represses expression from the same CRM (Williams et al. 2008). Interestingly, these splice forms differ in their ability to bind a transcriptional cofactor, *Ix* (Garrett-Engle et al. 2002). These examples demonstrate that both the composition and usage of transcription factor domains are modular and thus has the potential to be evolutionarily labile.

1.5 The Rise and Expansion of Metazoan Transcription Factor Families

The relative lack of novel transcription factors arising in metazoan evolution historically suggested that the evolution of transcription factors must contribute very little to the evolution of animal form. Indeed, many transcription factor families arose at the base of the metazoan lineage (reviewed by Degnan et al., 2009) and many even pre-date metazoans (Sebé-Pedrós et al. 2011). However, these families have each undergone series of duplications and divergence, resulting in numerous homologs, which are an important source of novel material for building GRNs (Teichmann and Babu 2004; Holland 2013; Pérez et al. 2014). Additionally, an increase in the number and types of transcription factors available may have promoted the evolution of multi-cellularity; it has been suggested that even more transcription factors were added to the repertoire before embryonic development could arise (de Mendoza et al. 2013). Several

important developmental transcription factors are not present in the sponge genome, suggesting that the creation of new transcription factors was critical to the evolution of this process (Srivastava et al. 2010).

New transcription factor homologs are created in two ways. When species diverge from a common ancestor, each initially is endowed with the same collection of transcription factors. Following the split, each set will acquire mutations, which generates orthologs. Until a novel regulatory mechanism is devised, in each species the orthologous proteins must execute the same tasks as they did in the common ancestor. This means that orthologous transcription factors are under a great amount of constraint and are therefore thought to remain thoroughly conserved. Conversely, paralogs, transcription factors generated by gene duplication events, are much more able to evolve (reviewed by Hoekstra and Coyne, 2007). Because the new transcription factor is a duplicate, it has several fates to choose from. Some are simply lost. Others take on some of the roles of the original transcription factor, lessening the burden on each copy and giving each more flexibility to mutate, known as subfunctionalization. Finally, if one copy maintains all of the ancestral roles of the transcription factor, the other paralog will have essentially no constraint and can neofunctionalize. Vertebrate A-Myb and C-Myb neofunctionalized after diverging from B-Myb and acquired novel transactivation domains; as a result B-Myb can rescue the single *Drosophila* Myb in functional-equivalence assays, but A and C-Myb cannot (Davidson et al. 2005). In this way, generation of paralogs results in modularity within a transcription factor family, because each paralog endows the others with greater freedom to evolve.

Some transcription factor families lend themselves to lineage-specific expansions, and could be an important source of gene regulatory change (reviewed by Nowick and Stubbs, 2010). Zinc finger transcription factor subfamilies seem to be especially prone to this phenomenon. The zinc-finger associated domain (ZAD) subfamily underwent extensive lineage-specific expansion in the insect lineage, yet there is only one such protein in the vertebrate lineage (Chung et al. 2007). Many of these insect-specific ZAD transcription factors are associated with developmental processes, and have been implicated in the evolution of the merostic ovary.

Conversely, a different zinc-finger subfamily, *Krüppel* type (KZNF), radiated dramatically in tetrapod vertebrate lineages, while only one paralog, PRDM9, exists in invertebrates (Liu et al. 2014). Many of these KZNF proteins are expressed during early development and are crucial for executing epigenetic reprogramming and other early

developmental tasks (Quenneville et al. 2012; Corsinotti et al. 2013; Liu et al. 2014). Overall, KZNFs diverge in sequence and expression rapidly, indicating that they acquire new functions as opposed to being redundant (Nowick et al. 2010). Over one-hundred KZNFs are unique to primates (Huntley et al. 2006), and many of them are expressed in reproductive tissues, such as the placenta, which vary between different mammalian species (Krebs et al. 2005; Liu et al. 2014). A survey of KZNFs in primates demonstrated that even among primates there is rapid gain and loss of new KZNFs (Nowick et al. 2011). When orthologous KZNFs are retained in more than one lineage their sequences diverge quickly, even in positions predicted to impact DNA binding properties (Nowick et al. 2011). KZNFs have been shown to be under positive selection, and have acquired amino acid differences between humans and chimpanzees much faster than other genes (Bustamante et al. 2005). The DNA-binding abilities of KZNFs are thought to be able to diverge by changing the number of zinc-finger domains in the protein, or by changing the DNA-contacting amino acids within a particular zinc-finger (Nowick et al. 2011). These changes have been predicted to have effects on target genes known to be involved in neurogenesis, muscle, and limb development, all of which differ between humans and other primates. Additionally, many KZNFs are differentially expressed in the human brain compared to the chimpanzee brain, suggesting a role in the evolutionary divergence of brain development in these species (Nowick et al. 2009). Thus, expansions within the developmental toolkit are important to the evolution of developmental processes and potentially even the evolution of development *as a* process after multicellular animals emerged.

1.6 Evolution of Protein-Protein Interactions

Transcription factors do not influence gene expression on their own, but do so as regulatory complexes mediated by interactions between the constituent transcription factors and cofactors. These interactions tend to be context-dependent; obviously a particular protein-protein interaction will only be relevant when both interacting partners are present. The composition of a transcription factor complex is also guided by the types of binding sites present in the CRM, and so many transcription factors participate in multiple non-identical complexes and are able to form interactions with more than one other protein. Therefore, changes to such interactions are predicted to be minimally pleiotropic.

A well-known example critical to arthropod evolution is Ftz, which acquired novel cofactor interactions that change the function of this transcription factor from homeotic to pair-rule segmentation factor (Löhr and Pick 2005). This occurred through loss of an ancestral interaction peptide motif and gain of a different one. More recently, it was shown that this is not a simple case of drastic changes in a particular lineage. Rather, the homeotic potential motif evolved into stronger and weaker variants of the ancestral sequence throughout the arthropod clade (Heffer et al. 2010). This suggests an inherent flexibility in this binding motif that could be coopted by GRNs to create novelty at other points in the evolutionary trajectory of these organisms. It also suggests that intermediate forms of an adaptive protein change need not be catastrophic to development, which is a common argument against transcription factor evolution as an important component of GRN evolution.

Newly evolved interaction motifs are also able to change the magnitude of an existing function. Throughout bilaterians, the transcription factor Engrailed (EN) interacts with a co-repressor Groucho (GRO), usually through a well-conserved motif (Smith and Jaynes 1996; Tolkunova et al. 1998). However certain groups of insects, namely dipterans and lepidopterans, have an additional GRO interaction motif. This motif strengthens the interaction between GRO and EN, and as a result augments EN's repressive abilities (Hittinger and Carroll 2008). Although the newer interaction motif does not currently promote novel function in EN, it might relieve constraint on itself and the ancestral motif such that new mutations could continue to be introduced, just as occurs in duplicated paralogs. Additionally, an advantage of changing GRNs through CRMs includes the ability to increase or decrease the quantity of a gene product and thus enhance or tone-down its function. This work suggests that the evolution of protein-protein interaction motifs is capable of producing quantitative vs. qualitative changes as well.

Importantly, changes to protein-protein interactions can occur without major disruptions of the existing protein-protein interaction domain. Brayer and colleagues discovered that an important new interaction evolved between HoxA11 and Foxo1a in placental mammals without actually changing the ancestral binding interface (Brayer et al. 2011). These genes are both crucial to the regulation of gene expression in endometrial stromal cells, and adaptive changes in HoxA11 had already been shown to be a driving force in evolution of pregnancy in mammals (Lynch et al. 2008). Without Foxo1a, HoxA11 represses the expression of pregnancy-related genes instead of activating them, so the advent of the Foxo1a/HoxA11 interaction is key to the

origin of this novelty (Lynch et al. 2009). Interestingly, the binding interface of these proteins did not change; in fact Foxo1a had not evolved much at all as evidenced by the fact that eutherian HoxA11 is able to interact with non-mammalian orthologs of Foxo1a (Brayer et al. 2011). This is critical because HoxA11 interacts with Foxo1a via its homeodomain, which is used in other essential functions of this transcription factor such as DNA-binding. The authors suggest that the causative amino-acid changes most likely produced a conformational difference in the protein in the eutherian lineage that makes a pre-existing binding interface accessible to Fox1a (Brayer et al. 2011).

1.7 Evolution of DNA-Binding Specify

Perhaps the most unexpected source of transcription factor adaptability is modular DNA-binding. This is partly because functional-equivalence studies implied conserved DNA specificity of both orthologous (McGinnis et al. 1990; Wang et al. 2002; Pockock et al. 2008) and paralogous transcription factors (Hoser et al. 2008; Gao et al. 2009). Instances of complete DNA-binding divergence have been uncovered, but they are quite rare (Hanes and Brent 1989; Baker et al. 2011). The inability to assay transcription factor binding preferences in a sensitive and high-throughput way was for a long time a larger roadblock to such studies. PCR-based methods for discovering DNA-binding preference such as SELEX recover only the highest affinity binding sites, and caused the misconception that protein-DNA recognition follows a simple one-to-one code. Only recently have we realized that protein-DNA interactions are extraordinarily complex (reviewed by Siggers and Gordân, 2014). Newer technologies, such as Protein Binding Microarrays (Berger et al. 2006; Berger and Bulyk 2006), are able to universally assess DNA-binding preference and thus have been crucial for recent works that have revealed modularity in transcription factor binding.

Initial studies that made use of Protein Binding Microarrays came across a few surprising findings. First, many transcription factor's binding preferences are best described by multiple rather than one position weight matrix (Badis et al. 2009; Zhu et al. 2009; Gordân et al. 2011). These are commonly called primary and secondary motifs, where the primary motif is the most-preferred. Collapsing these motifs into one can be misleading because it obliterates important nucleotide interdependencies (e.g. AC or TG can be found together in a binding site, but not AG

or TC). Additionally, while closely related paralogs share their primary binding site, they frequently recognize different secondary binding sites (Badis et al. 2009; Zhu et al. 2009; Gordân et al. 2011). These studies suggested an important source of modularity that has only just been characterized in greater detail.

A recent study of yeast C2H2 zinc finger paralogs found modular differences in binding (Siggers et al., 2014). These proteins bind DNA using two adjacent zinc finger domains and can be divided into groups in which a common canonical motif is bound by all members and subgroups that share an additional specific motif. Here, it was found that paralogs from the same group are able to adopt different conformations to recognize alternative binding sites; however the mechanism differs between subgroups. For example, one subgroup has evolved changes within both zinc finger domains that permit an alternate docking geometry, while another makes use of an N-terminal region outside the zinc finger domains to stabilize alternative site binding. In all subgroups, both the canonical and alternative sites are bound with high affinity, indicating that recognition of the common canonical motif is not compromised by this plasticity. Extensive cataloging of the forkhead box (Fox) transcription factor family revealed flexibility in binding over evolutionary time too (Nakagawa et al. 2013). Some Fox proteins bind canonical primary and secondary motifs, some bind a completely different motif, termed FHL, and others are bispecific and therefore can use the primary, secondary, and FHL motifs. Intriguingly, preference for motifs like FHL and dual specificity has arisen multiple times within the Fox family, but never through changes to the DNA-binding helix. Instead, an N-terminal tail that allows for alternative structural configurations appears to be responsible for modular binding changes. These studies describe important mechanisms that allow paralogous transcription factors to evolve while avoiding pleiotropic effects, in many cases by preserving binding to a canonical motif.

Orthologous transcription factors are under greater evolutionary constraint; therefore, until recently it was uncertain whether this type of modularity would extend to them. In addition to the differences in Fox paralog families described above, Nakagawa and colleagues also observed that different orthologs of yeast Fox3 exhibit substantial DNA binding diversity. Some recognize the canonical primary and secondary motifs, others use the aforementioned FHL motif, and yet another subset recognizes a different variant, termed FVH. Fox3 orthologs that bind the FVH motif also have divergent amino acids in their DNA recognition helix (Nakagawa et al.

2013). These observations suggested that orthologs may be able to make use of the same mechanisms as paralogs to diverge in DNA specificity. However these orthologs diverged between single-celled yeast species, and therefore may be under less constraint than the transcription factors used in metazoan development.

Work from the Hinman laboratory demonstrates that while developmental transcription factors may not diverge as dramatically as yeast orthologs, they do seem capable of exploiting modular divergence mechanisms used by paralogs. In this study, it was found that echinoderm Tbrain transcription factors from a sea urchin (*Strongylocentrotus purpuratus*) and a sea star (*Patiria miniata*) have evolved differences in their secondary binding abilities (Cheatle Jarvela et al. 2014). Interestingly this secondary motif is also different compared to what has been reported for the vertebrate ortholog of Tbr, Eomesodermin (Badis et al. 2009). However, all three orthologs maintain the same primary motif in spite of 800 million years of divergence time. The mechanism that allowed this change to evolve is not yet known, but these orthologs have differences in DNA-contacting amino acids which might have caused changes in binding specificity. Interestingly, Tbr is known to have different developmental functions in the sea urchin compared to the sea star. In the sea star, Tbr has roles in the development of the endomesoderm and also in the ectoderm (Hinman et al. 2007; Hinman and Davidson 2007; McCauley et al. 2010). However, in the sea urchin, Tbr's only function is in skeletogenesis (Croce et al. 2001; Oliveri et al. 2002). Changes to Tbr's DNA binding abilities over the course of echinoderm evolution may be responsible for differences in the developmental roles of this protein.

Several studies have demonstrated that these secondary and non-canonical alternative binding sites are not only functional *in vivo*, but in many cases have distinct developmental tasks. Notably, in the case of Hedgehog-responsive genes used during *Drosophila* development, low-affinity, non-consensus Ci sites cannot be replaced by higher affinity sites as this results in a switch from activation to repression (Parker et al. 2011; Ramos and Barolo 2013). As a result, these sites convey important positional information across the anterior-posterior axis during development. In another example, it was found that differences in secondary motif specificity among homeodomain paralogs allows each to execute a particular regulatory program during *Drosophila* muscle development; all have the same primary motif and therefore would not be able to confer different myoblast identities without these unique secondary motifs (Busser et al.

2012). Importantly, since alternative binding sites can be gained and lost without affecting a conserved site (Nakagawa et al. 2013; Cheatle Jarvela et al. 2014; Siggers et al. 2014), these developmental functions can be uncoupled and evolve independently, thus relieving constraint on developmental processes and allowing for more diverse cell types and structures to arise.

It has been suggested that use of high affinity primary and lower affinity secondary sites during development could be important to coordinate the timing of different developmental events through a temporal protein gradient (Cheatle Jarvela et al. 2014). It has been shown that during eye development, proper timing of *pax6* expression is controlled by the affinity of the Prep1 binding sites within its enhancer (Rowan et al. 2010). The endogenous sites are low affinity; replacing these with higher affinity sites causes *pax6* expression to begin too early. Heterochrony, or shifts in the rate or timing of developmental processes, is an important source of morphological differences between species (reviewed by Keyte and Smith, 2014; Smith, 2003). In some cases, heterochrony is the result of altering a known developmental timekeeping mechanism, such as the oscillating segmentation clock used in vertebrate somatogenesis (Gomez et al. 2008; Keyte and Smith 2012), but such elaborate molecular clocks are rare. Less is known about how other mechanisms that coordinate developmental events evolve. Modular evolution of binding site preference and affinity could explain some cases where shifts in relative timing occur because it allows for coupling and decoupling of processes coordinated by the same spatio-temporal protein gradient.

1.8 Evolution of Post-Translational Modifications

Post-translational modifications are a common way to increase protein functional diversity. They are of particular interest to those seeking to understand how transcription factors may evolve while avoiding pleiotropy because they are known to regulate the location, longevity, and activity of proteins. They can also allow for alternate protein structure and enhance or prevent protein-protein interactions and DNA-binding (reviewed by van Loosdregt and Coffey, 2014; Prasad et al., 2012). Thus, as is the case for CRMs, the effects of mutations to post-translational modifications can easily be limited to a particular developmental context. Some types of modification, such as phosphorylation, are reversible, and therefore offer even more flexibility.

Moreover, new modification sites evolve rapidly. A comprehensive bioinformatics screen identified over two-hundred ubiquitylation sites that arose in the human lineage since it split from other primates (Kim and Hahn 2012). A similar study also found 37 human-specific phosphorylation sites (Kim and Hahn 2011). Interestingly, many of these occurred in chromatin-remodeling proteins that frequently interact with transcription factors.

It is unsurprising then that recent work has found compelling connections between novel post-translational modification sites within transcription factors and the evolution of new features. For example, Ultrabithorax (Ubx), a Hox transcription factor, is expressed in the limb primordia of both insects and crustaceans. Alteration of the Ubx protein explains differences in appendage number between different groups of arthropods rather than CRM level changes (Ronshaugen et al. 2002). Taghli-Lamalle and colleagues found that an important difference in Ubx between crustaceans and insects involves loss of phosphorylation sites (Taghli-Lamalle et al. 2008). They demonstrated that phosphorylation of these sites interferes with the ability of Ubx to repress another transcription factor, Distal-less, and as a result repress appendage formation. The molecular consequence of phosphorylating these sites is unknown, but there are precedents for phosphorylation affecting DNA-binding of Hox proteins and also their protein-protein interactions (Bourbon et al. 1995; Jaffe et al. 1997).

Another interesting example entails evolution of pregnancy in mammals, due in part to changes in phosphorylation of the CCAATT enhancer binding protein beta, or CEBP β (Lynch et al. 2011). This work demonstrated that a mere three amino-acid changes in an internal regulatory domain, resulting in the loss of two ancestral phosphorylation sites and the gain of a new one elsewhere, completely changed how this transcription factor responds to cAMP signaling. Phosphorylation of the novel site by GSK-3 β is required for CEBP β to activate the expression of prolactin, an important pregnancy hormone. Developmental gene regulatory networks integrate both signaling pathways and transcription factors, and so alteration of the post-translational modifications that connect them offers an attractive way of modifying developmental GRNs.

1.9 Remaining Questions: The FoxP2 Mystery

One of the most exciting and enigmatic examples of coding changes in a transcription factor leading to developmental changes and evolutionary novelty is the case of FoxP2. This

gene is thought to be essential to explaining the evolution of speech in humans (Enard et al. 2002). Furthermore, FoxP2 is an important developmental regulator; many of its target genes are other transcription factors (Konopka et al. 2009; Nelson et al. 2013). FoxP2 has also been shown to be crucial for the genesis of neuronal precursors during cortical development (Tsui et al. 2013), tying the evolution of speech as a novel feature to a human-specific change in embryonic neurogenesis.

FoxP2 has acquired two functional amino-acid changes since the human and chimpanzee lineages split, which have been shown to impact neurogenesis and FoxP2's regulatory abilities (Enard et al. 2009; Konopka et al. 2009). However, these mutations lie outside of known functional domains, thus the consequence of these changes has been mysterious. These changes do not affect the ability to bind FoxP1 or FoxP4 to produce known heterodimers (Konopka et al. 2009). Using a combination of microfluidic techniques and ChIP-seq data, Nelson and colleagues found that the DNA binding specificities of the human and chimpanzee FoxP2 are very similar (Nelson et al. 2013), effectively ruling out DNA-binding changes as a possibility as well. Cis-regulatory changes leading to enhanced expression of FoxP2 in humans have also been found (Maricic et al. 2013) and seem to be the source of a selective-sweep in the human lineage. However, the functional consequences of the two amino-acid substitutions remains elusive in spite of numerous studies devoted to the evolution of this transcription factor.

1.10 Concluding Remarks

Transcription factor coding changes are becoming a theoretically more accepted source of GRN evolution, but there are still only a few studies documenting specific changes and tying those to developmental novelties. As these experimental examples continue to increase, we will be able to decipher what impact these changes have on the wiring of their GRNs and how this might differ from CRM mutations. The original logic supporting CRM mutations over transcription factor changes would suggest that the former are ideally suited to alter the expression of a particular gene and potentially also its downstream targets within a tissue or cell-type, while changes to transcription factors will have broader effects, changing the regulation of large sets of target genes across the organism. The experimental evidence described here points to incremental and modular transcription factor mutations being favored by evolution, and latent motifs and abilities becoming more pronounced or reduced over time. While this may look more

like CRM evolution than what was originally believed about transcription factor coding changes, it could be that their effects on the surrounding GRNs are not equal. Each type of mutation may be more ideal for driving different types of GRN changes and developing different types of novelty. More information about both types of change is required to tease out this discrepancy. On the other hand, several recent works demonstrate that CRM and transcription factor mutations may generally operate together (Heffer et al. 2010; Maricic et al. 2013). Additional work will reveal whether such cooperative changes to GRNs are the exception, the rule, or simply another option in creating diverse GRNs, a myriad of developmental processes, and seemingly endless animal forms.

Chapter 2: Modular Evolution of DNA Binding Preference of a Tbrain Transcription Factor Provides a Mechanism for Modifying Gene Regulatory Networks.

2.1 Preface

The following work has been accepted for publication in *Molecular Biology and Evolution* (Oxford University Press) as of July 8, 2014, and is available at doi:10.1093/molbev/msu213. This work is the product of collaboration between members of the Hinman lab, including myself, Veronica Hinman, and a former undergraduate Lisa Brubaker, Martha Bulyk and her technician Anastasisa Vedenko (Brigham and Women's Hospital and Harvard Medical School), and Bruce Armitage and his former graduate student Anisha Gupta (Carnegie Mellon Department of Chemistry). The Bulyk lab performed the Protein-Binding Microarray experiments using proteins purified by the Hinman lab, and analyzed the resulting datasets. The Armitage lab members aided in the design, troubleshooting, and data analysis related to the SPR experiments. All other experiments were designed and executed by myself and Veronica. Lisa assisted with the cloning of some of the constructs used, pilot qPCR experiments, and also with the preliminary validation of the *PmTbr* antibody.

In this work, we extensively examined the DNA-binding abilities of two echinoderm orthologs of the transcription factor, Tbrain, in order to understand whether modular binding abilities discovered among transcription factor paralogs might also facilitate divergence of orthologs. We also demonstrate that the alternate binding site found for the sea star ortholog is able to function *in vivo* and is differentially responsive to temporal changes in protein concentration. We have not yet demonstrated that this difference has impacted the evolution of echinoderm development, which is the ultimate goal of the larger project that encompasses this study. This will be addressed in Chapter 4.

Due to its large size, Supplemental Table 2.1 (called Supplemental Table 1 online and in the original work) has not been included in this thesis, but can be freely accessed online as we have chosen the open-access publishing option for this manuscript.

2.2 Abstract

Gene regulatory networks (GRNs) describe the progression of transcriptional states that take a single-celled zygote to a multicellular organism. It is well documented that GRNs can evolve extensively through mutations to *cis*-regulatory modules. Transcription factor proteins that bind these *cis*-regulatory modules may also evolve to produce novelty. Coding changes are considered to be rarer, however, because transcription factors are multifunctional and hence are more constrained to evolve in ways that will not produce widespread detrimental effects. Recent technological advances have unearthed a surprising variation in DNA binding abilities, such that individual transcription factors may recognize both a preferred primary motif and an additional secondary motif. This provides a source of modularity in function. Here, we demonstrate that orthologous transcription factors can also evolve a changed preference for a secondary binding motif, thereby offering an unexplored mechanism for GRN evolution. Using Protein Binding Microarray, Surface Plasmon Resonance, and *in vivo* reporter assays, we demonstrate an important difference in DNA binding preference between Tbrain protein orthologs in two species of echinoderms, the sea star, *Patiria miniata*, and the sea urchin, *Strongylocentrotus purpuratus*. While both orthologs recognize the same primary motif, only the sea star Tbr also has a secondary binding motif. Our *in vivo* assays demonstrate that this difference may allow for greater evolutionary change in timing of regulatory control. This uncovers a layer of transcription factor binding divergence that could exist for many pairs of orthologs. We hypothesize that this divergence provides modularity that allows orthologous transcription factors to evolve novel roles in gene regulatory networks through modification of binding to secondary sites.

2.3 Introduction

Animal morphology arises under the control of interacting networks of regulatory genes that operate during embryonic development. A central pursuit for understanding evolution of animal form is therefore to determine how these gene regulatory networks (GRNs) evolve. Several influential papers, published almost 50 years ago, set forth the hypothesis that non-coding DNA, i.e. the *cis* regulatory DNA, would be the predominant source of evolutionary change. This idea was first predicted in 1961 by Monod and Jacob, who emphasized the

important distinction between biochemical protein function and context of the action of that protein. Britten and Davidson established the hypothesis that regulatory mutations, which control this context, would be the prominent source of evolutionary variation. In 1975, King and Wilson suggested that the stark differences in morphology and behavior between chimpanzees and humans, despite their overall high similarity in DNA sequence, could be the result of differences in their regulatory DNA. These, and other papers of this era, firmly established the notion that changes to the deployment of genes, rather than the biochemical function of genes would be the main driver in morphological diversity. The rationale for this is theoretically straightforward. A single gene is usually regulated by multiple *cis* regulatory modules (CRMs; and also referred to as enhancers), so that its expression in distinct spatial and temporal domains is governed independently. By comparison, the transcription factors that utilize these CRMs must remain evolutionarily dormant because they often are needed to orchestrate a variety of crucial tasks. This tends to be especially evident during development where transcription factors are used in multiple contexts. It stands to reason that mutations to CRMs have fewer pleiotropic effects and are therefore more likely to pass the filter of selection and thus these become the source of novelty and change (Carroll 2005; Prud'homme et al. 2007; Wray 2007).

Many early discoveries in evolutionary developmental biology supported this hypothesis. A wealth of data demonstrates that all animals share highly similar sets of regulatory genes, which have been dubbed the toolkit for development (Carroll 2005). Regulatory genes comprise a relatively small portion of the transcriptome and hence must be used in many tissues and times in the developing embryo. Elegant xeno-transfer experiments further cemented the idea that regulatory proteins were evolutionarily dormant (McGinnis et al. 1990; Wang et al. 2002; Wang et al. 2004). One of the most exciting of these was the demonstration that the mouse *pax6* gene could rescue the mutant phenotypes of the *eyes absent* ortholog in *Drosophila*, and had therefore presumably changed very little in the 900 million years (Hedges et al. 2006) since insects and vertebrates last shared a common *pax6* gene (Halder et al. 1995).

More recently, a growing body of evidence suggests that while transcription factors may be a less common source of GRN evolutionary change, they are certainly not unchanging (Galant and Carroll 2002; Ronshaugen et al. 2002; Lynch and Wagner 2008; Chen et al. 2010; Nakagawa et al. 2013b). In fact, the transcription factors that specify chemosensory neurons in *Caenorhabditis* acquired more nonsynonymous mutations than the chemosensory structural

genes that they regulate in the same evolutionary distance (Jovelin 2009). Evolutionary changes occur in protein-protein interactions (Löhr and Pick 2005; Brayer et al. 2011) and post-translational modifications (Lynch et al. 2011). The aforementioned examples explain how Ftz switched from a homeotic to a segmentation gene in insects and events contributing to the evolution of pregnancy as a novel feature in mammals, respectively. In very rare instances, evolutionary changes are also found within DNA consensus motif recognition (Hanes and Brent 1989; Baker et al. 2011). In the case of Bicoid, this new specificity is crucial for its function in directing anterior patterning in the *Drosophila* embryo (Hanes et al. 1994). Changes to DNA binding appear to be the rarest because unlike changes to the transcription factor's cohort of protein-binding partners and post-translational regulation, these presumably affect all instances of their function.

New technologies can determine DNA binding motifs with greater sensitivities, particularly Protein Binding Microarrays (Berger et al. 2006). These arrays are designed with double-stranded DNA oligonucleotides of all possible k -mers, usually 44,000 oligonucleotides of 60bp (with a 35 bp variable region). This provides 32-fold coverage of all possible 8-mer sequences. Protein binding to all oligonucleotides is measured and position weight matrices that best represent binding sequence preferences are compiled. This type of data demonstrates that transcription factor-DNA interactions are more complex than originally imagined. In a survey of mouse transcription factor binding preferences, nearly half of the proteins display binding preference for two distinct motifs; these have been termed their primary and secondary motifs (Badis et al. 2009). Secondary motifs are built when a single position weight matrix is unable to explain all of the highly bound sequences from the array data. Equally intriguing was the realization that these secondary motifs frequently differ for closely related paralogs. Presumably this provides a mechanism through which paralogs may evolve. Upon duplication, one gene paralog can acquire new functions while the other maintains original functions. The *in vivo* functional significance of this additional component of binding specificity is still largely unknown, although a number of studies demonstrate that the binding motifs that do not match the primary consensus motif are not only present in endogenous CRMs but are often functionally distinct from the primary motif (Rowan et al. 2010; Parker et al. 2011; Busser et al. 2012; Zhu et al. 2012). Orthologs, which arise when species diverge instead of through gene duplication, experience greater evolutionary constraint, as they must maintain original functional roles while

acquiring changes. Little is known about whether such flexibility in secondary binding also applies to orthologous transcription factors.

Recently, protein-binding microarray technology has revealed that the forkhead family of transcription factors can acquire novel binding specificity among both orthologs and paralogs (Nakagawa et al. 2013b). Importantly, this acquisition seems to have a modular component to it. Some forkhead families can bind both the primary and secondary motif as well as an additional novel motif, while others bind to either the primary and secondary or only to novel motifs. It is unknown whether this phenomenon extends to other transcription factor families and the functional consequences of this change.

Here, we investigate orthologous Tbrain (Tbr) transcription factors from the sea star, *Patiria miniata* (*Pm*), and sea urchin, *Strongylocentrotus purpuratus* (*Sp*), to question whether these proteins evolved biochemical changes in their DNA binding preferences. These proteins were selected as they have well characterized and critical roles in early echinoderm development (Ryan et al. 1998; Shoguchi et al. 2000; Croce et al. 2001; Tagawa et al. 2001; Fuchikami et al. 2002; Horton and Gibson-Brown 2002; Oliveri et al. 2002; Hinman, Nguyen, Cameron, et al. 2003). During sea star embryogenesis, Tbr is highly pleiotropic and required for specification of cell types within the mesoderm, endoderm, and ectoderm (Hinman and Davidson 2007; McCauley et al. 2010). In sea urchins, intriguingly, Tbr appears to have lost these roles and is instead only required for the specification of one type of mesoderm, the skeletogenic mesoderm. These genes are members of the T-box family of transcription factors which are characterized by having a single T-box DNA binding domain. The DNA binding properties of these proteins are relatively well-studied. There is a particular interest in understanding how groups of T-boxes with the same primary binding motif, expressed in the same tissue, are capable of exerting distinct functions. Many studies show that these transcription factors are characteristically dose dependent and others suggest that differences in binding site affinities may be crucial for allowing them to operate in a competitive and hierarchical fashion (Macindoe et al. 2009; Sakabe et al. 2012). Therefore, there is a great interest in understanding the binding properties of these transcription factors.

The echinoderm Tbr proteins are orthologous to vertebrate Eomesodermin (Eomes) (also known as Tbr2), Tbr1, and Tbx21 (Papaioannou and Silver 1998; Croce et al. 2001). As is the case for many vertebrate transcription factors, these paralogs presumably arose as a result of the

vertebrate lineage-specific duplication from a single deuterostome ortholog. We show that these three deuterostome orthologs (sea urchin Tbr, sea star Tbr and mouse Eomes) have a highly similar primary binding motif, which we think has therefore been maintained in the approximately 800 million years (Hedges et al. 2006) since these taxa last shared a common ancestor. Here we show that, the sea star Tbr and mouse Eomes each have a preference for an additional, unique secondary motif, while the sea urchin Tbr protein has no preference for a secondary motif. This demonstrates that these orthologs evolved biochemical changes in function of their DNA binding domains. We show that at saturating levels of Tbr, the primary and secondary motifs are functionally interchangeable in sea stars. The motifs, however, provide different transcriptional responses as Tbr protein levels change. The use of primary and secondary motifs represents a modular component to transcriptional regulation; subsets of target genes under control of secondary motifs can evolve while those regulated by primary motifs remain conserved. Our data indicate that this evolvable function can manifest as differences in relative timing in response to transcriptional state changes. Given the pervasiveness of secondary binding ability among transcription factors, such changes in secondary binding may prove to be an important source of gene regulatory evolutionary change.

2.4 Results

2.4.1 Sea Urchin and Sea Star Tbr are orthologous to Mouse Eomes

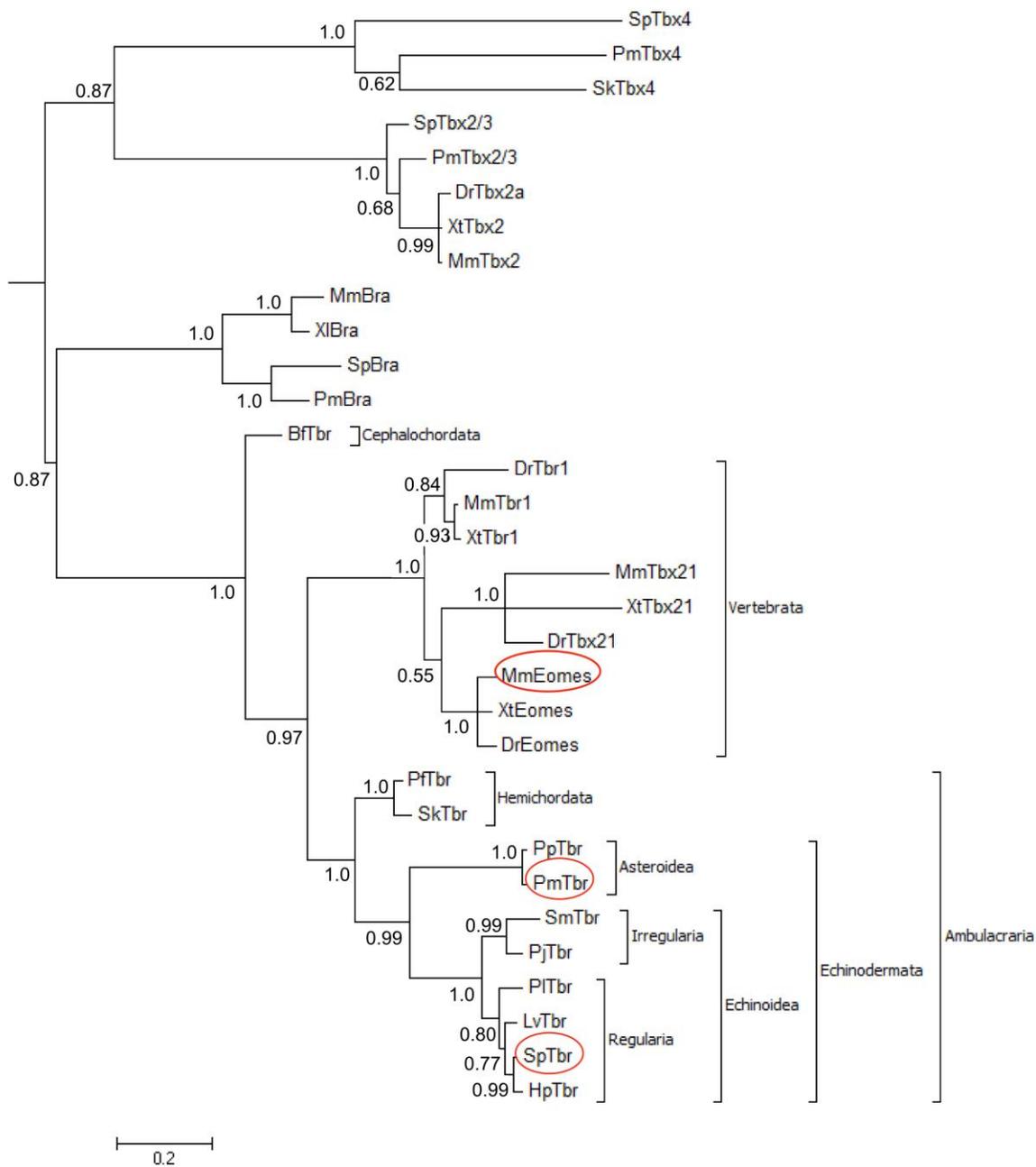
In the sea star, *P. miniata*, *tbrain* (*PmTbr*) was originally isolated from a cDNA library probed with a cDNA clone corresponding to another T-box factor, *PmBrachyury* (*PmBra*) (Hinman, Nguyen, Cameron, et al. 2003). Only *bra* and a single *tbr* ortholog were identified in this screen. To determine whether any other *tbr* orthologs were present within the genome, we bioinformatically queried the *P. miniata* genome sequence (contigs 1.0; Echinobase.org) (Cameron et al. 2009) by performing a tblastn identity search to the translated *MmEomes* T-box domain (Accession: AK089817.1). We collated the *P. miniata* sequences that matched with an e-value less than $1e-12$. These sequences in turn were used to query the National Center for Biotechnology Information non-redundant protein database using blastx (Altschul et al. 1990). Four T-box family members were identified in this comprehensive search. These correspond to a subset of the six T-box family members identified previously in the sea urchin, *S. purpuratus*,

genome (Howard-Ashby et al. 2006). We next determined the orthology of these four T-box factors by constructing a gene tree (see Methods) of these T-boxes and their homologs from other deuterostome animals (Figure 2.1A).

PmTbr clusters with a *tbr* gene isolated from another species of sea star (*Patiria pectinifera*; *PpTbr*), while the *SpTbr* clusters with *tbr* orthologs from five other species of sea urchins, including two species of sand dollars, which form a distinct group (Irregularia) within the sea urchins. Importantly, the sea urchin and sea star genes form a single grouping supported by a posterior probability of 0.99. Thus, there is a strong correspondence between the topology of this gene tree and the echinoderm species tree (Pisani et al. 2012). In vertebrates there are three *tbr* paralogs, viz., *eomes*, *tbx21*, and *tbr1*, which also form a single grouping. These three paralogs form a single cluster with the echinoderm orthologs with the node connecting them supported by a posterior probability of 0.97.

Meanwhile, the other T-box proteins isolated in the screen are orthologous to *bra*, *tbx2/3*, and *tbx4*. Only a single *tbr* ortholog is identified from eight species of echinoderms, including two with sequenced genomes. Therefore, we are confident as reasonably possible that there is a single *tbr* ortholog among these echinoderms, and that it is the only echinoderm ortholog of the vertebrate *eomes*, *tbx21*, and *tbr1* paralogs.

A



B

PmTbr	KASVFLCNSELWRKFHEHRTEMIITKQGRRMFPQLVFRLSGLNPAAHYNVFVDMVIADPNSWKLFQ
SpTbr	KASVYLCNRDLWRKFHQHKTEMIITKQGRRMFPQLVFKLTGLNPTSQYNVFVDMVLCDPNQWKFQ
MmEomes	RAHVYLCNRPLWLKFHRHQTEMIITKQGRRMFPFLSFNINGLNPTAHYNVFVEVVLADPNHWRFQ
PmTbr	SGKWVATGKSDGVPRATGIYKHPDSPNTGEHWMRQDIAFSKCLKLTNNRGKDS---GYLMINSMHI
SpTbr	CGKWIPCGQAENIPKVSNIYLHPDPSNGLHWMHQDIVFSKCLKLTNHRADN---GFVILNSMHQ
MmEomes	GGKWVTCGKADNNMQGNKMYVHPESPNTGSHWMRQEISFGKCLKLTNNKGANNNTQMIVLQSLHK
PmTbr	YQPRIHVLDLTGA-----RVLQTHSFPETQFIGVTAYQNTDITQLKIDHNPFAKGFERNYDS
SpTbr	YQPRIHVLELSES-----RSIQTHSFPETQFFGVYAYQNTDVTQLKIDYNPFAKGFERNYDN
MmEomes	YQPRLHIVEVTEDGVEDLNEPSKTQTFTFSETQFIAVTAYQNTDITQLKIDHNPFAKGFERNYDS

Figure 2.1: Phylogeny and Sequence Alignment for Pm and SpTbr Tbox-DNA Binding Domains.

A. Tree topology was determined using a MrBayes model (TOPALI v2.5), and is based on a character alignment that includes the T-box sequences depicted in Supplemental Figure 2.1. Lengths of branches are drawn to the scale indicated (0.2 expected substitutions per site) and the numbers indicate support by posterior probability. *Bf*- *Branchiostoma floridae*, *Dr*-*Danio rerio*, *Hp*- *Hemicentrotus pulcherrimus*, *Lv*- *Lytechinus variegatus*, *Mm*-*Mus musculus*, *Pf*-*Ptychodera flava*, *Pj*- *Peronella japonica*, *Pl*- *Paracentrotus lividus*, *Pm*- *Patiria miniata*, *Pp*- *Patiria pectinifera*, *Sk*-*Saccoglossus kowalevskii*, *Sm*- *Scaphechinus mirabilis*, *Sp*-*Strongylocentrotus purpuratus*, *Xl*-*Xenopus laevis*, *Xt*- *Xenopus tropicalis* B. Conceptual translation of *PmTbr*, *SpTbr* and *MmEomes* T-box domains. Highlighted amino acids indicate residues involved in interaction with DNA according to alignment with *XlBra* crystal (Protein Data Bank ID 1XBR) (Müller & Herrmann 1997). Yellow amino acids indicate identical amino acids, while blue denotes nonconserved interactions within the echinoderms. Sequence alignments to *XlBra* are provided in Supplemental Figure 2.1.

2.4.2 Sea Urchin and Sea star Tbr orthologs have different DNA binding preferences

The structure and function of transcription factors, especially the DNA binding domains, are often highly conserved across even widely divergent species. The 180 amino acid T-box domain is particularly well-conserved (Macindoe et al. 2009). An alignment of the *SpTbr* and *PmTbr* DNA binding domains demonstrate they are 73% identical and 89% similar (Figure 2.1B). This indicates that these orthologs share high degree of conservation, yet there is variation that could permit functional divergence. We wanted to determine if any of these differences could indeed have a functional consequence. As a first approach, we used the known crystal structure of a closely related T-box protein, *Xenopus laevis* brachyury (*XlBra*) (Protein Data Bank ID 1XBR)(Müller and Herrmann 1997) to map the likely DNA contacts within the sea star and sea urchin Tbr amino acid sequences. We also used these sequences to predict the structures of *PmTbr* and *SpTbr* using the Phyre server (Kelley and Sternberg 2009). The overall structure of the DNA binding domain is not predicted to be perturbed by the non-identical amino acids (Supplemental Figure 2.2A). Nineteen amino acids are predicted to contact the DNA (highlighted in yellow in Figure 2.1B), and of these, two are not identical between the sea urchin and sea star (blue highlight, Figure 2.1B). At residue 338/428 the *SpTbr* protein has a glutamine

where *PmTbr* has a serine. This appears to be unique for each species as neither is conserved with the residue in *XlBra* nor *MmEomes* (Figure 2.1B, Supplemental Figure 2.1). However, at residue 389/479, *PmTbr* has an asparagine that is also present in vertebrate proteins, while *SpTbr* has a histidine at this position. Both of these changes occur in residues known to interact with the DNA backbone as opposed to the bases themselves (Supplemental Figure 2.2B and 2C). However, in the case of the homeodomain protein, Bicoid, a change in DNA binding specificity compared to its Antp paralog is correlated with a single backbone-contacting amino acid difference (Hanes and Brent 1989), and so these two changes to Tbr may also impact DNA binding specificity.

While suggestive of a potential for a functional difference, protein-DNA interactions are not well understood enough to predict binding preferences. Therefore, it is unclear how these changes and others that do not occur in amino acids that contact DNA might affect specificity for DNA sequences. We therefore sought to determine experimentally if any differences in DNA specificity exist for these orthologs. We bacterially expressed and purified *PmTbr* and *SpTbr* DNA binding domains as GST-fusion proteins and used Protein Binding Microarrays to universally assess their binding preferences (Berger et al. 2006; Berger and Bulyk 2009). It is important to note that these experiments cannot account for the effects that cofactors normally encountered *in vivo* might have on Tbr DNA binding specificity. We chose to test only DNA binding domains because full-length proteins prove to be extremely unstable. In a previously reported study, no difference in DNA binding was observed when full-length and DNA binding domain versions of *MmTbx5* were compared (Macindoe et al. 2009). Moreover, T-box protein specificity for several homologs, including *MmEomes*, has previously been shown to reside in the T-box domain itself, while other regions of the protein account for nuclear localization signals and transactivation domains (Conlon et al. 2001). This work suggested that the Tbr DNA binding domains would be sufficient to capture the full DNA binding capabilities of these proteins.

PCR-based methods, such as SELEX, have been used to identify consensus sites for other T-box transcription factors (Conlon et al. 2001; Macindoe et al. 2009). However, these experiments, based on technologies available at the time, were limited to identifying only the highest affinity binding motifs. Protein Binding Microarrays uncover additional layers of

binding specificity, particularly differences in secondary sequence preferences (Badis et al. 2009).

The DNA binding specificity of each Tbr was assayed by Protein Binding Microarray in duplicate with strong agreement between replicates (*PmTbr* Pearson's $r = 0.915$ and *SpTbr* Pearson's $r = 0.917$). Datasets depicting the E-score calculated for each 8-mer are available in Supplemental Table 2.1. The Protein Binding Microarray experiments demonstrate that *PmTbr* and *SpTbr* orthologs recognize the same primary position weight matrix, or motif, which represents the probability of the transcription factor binding to all potential binding sites (Figure 2.2A and C). This motif can explain Tbr binding to a large number of 8-mer binding sites, but for simplicity, it can be represented by the following consensus sequence, 5'-AGGTGTGA-3'. This single binding site was selected for use in subsequent experiments because each position contains the most highly preferred nucleotide predicted by the position weight matrix. Both Tbr orthologs recognize this 8-mer binding site with a very high E-score (*PmTbr*, $E = 0.499$, *SpTbr*, $E = 0.498$). The E-score (enrichment score) is a non-parametric, modified Wilcoxon-Mann-Whitney statistic developed especially to measure relative binding preference for simple and robust comparison of Protein Binding Microarray data across datasets (Berger et al. 2006). E-scores range from -0.5 to 0.5, but scores of 0.45 and greater indicate a stringent binding threshold (Berger et al. 2008; Badis et al. 2009). This motif closely matches previously published T-box consensus sites (Conlon et al. 2001; Macindoe et al. 2009), and in particular, the primary binding site for the mouse ortholog of Tbr, *MmEomes* ($E = 0.497$, UniProbe Database), which was also obtained by universal Protein Binding Microarrays (Badis et al. 2009).

Previous studies using these sensitive protein binding arrays have shown that approximately 40% of transcription factors that have been tested can bind two distinct motifs (Badis et al. 2009; Gordân et al. 2011). By convention, the motif with the higher seed E-score is called the primary motif and the next preferred, high confidence motif, the secondary motif. Of our two echinoderm Tbr orthologs, only *PmTbr*, however, consistently recognized an additional high E-score position weight matrix, best represented by the 8-mer, 5'-AGGTGACA-3' ($E = 0.483$) (Figure 2.2B, Supplemental Table 2.1). While very similar to the initial motif, it differs in positions 13 and 14, where AC replaces the primary site's TG. Therefore, here we call the position weight matrix represented by the 8-mer 5'-AGGTGTGA-3' site the primary motif, and that represented by 5'-AGGTGACA-3', the secondary. These two motifs are not condensed

into one more degenerate position weight matrix, because the two distinct motifs better explain the Protein Binding Microarray data than can a single motif (Badis et al. 2009). This secondary motif was found consistently in replicate experiments. In contrast, *SpTbr* never demonstrated strong preference for a particular additional motif (Supplemental Table 2.1) over replicate experiments. When we performed a similar analysis using the data from *SpTbr* binding in order to find a secondary motif, the result was simply a more degenerate version of the primary motif. Additionally, we show that *SpTbr* and *PmTbr* have similar E-scores for 8-mers that match the primary position weight matrix, but 8-mers corresponding to the *PmTbr* secondary motif are preferred by *PmTbr* (Figure 2.2D).

The mouse Eomes ortholog also was previously shown to also have two high E-score motifs. While both species of echinoderm and the *MmEomes* have highly similar primary position weight matrices, the secondary motifs are dissimilar. The *MmEomes* secondary motif is represented as 5'-AGGTGTCG-3' (E= 0.493, UniProbe Database) (Badis et al. 2009). Both *PmTbr* and *MmEomes* secondary motifs are not the same as the primary motif or each other, particularly in positions 13, 14, and 15 (Figure 2.2). These data suggest that the primary motif has most likely remained the same over the extensive time scale since these deuterostomes have last shared a common ancestor while the preference for a secondary site has evolved, either through single or multiple losses and gains, over the same time scale. This study is the first demonstration of such an evolutionary change in orthologous transcription factor function.

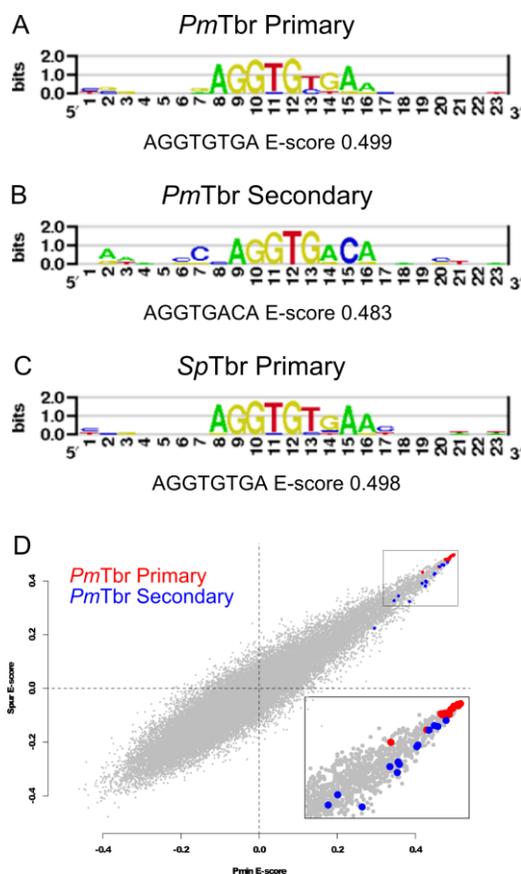


Figure 2.2: Position Weight Matrices Depicting Binding Specificities of Tbr Orthologs.

Position weight matrices represent the top motifs obtained from PBM data using the Seed-and-Wobble algorithm (Berger *et al.*, 2006, Berger and Bulyk 2009) representing *SpTbr* and *PmTbr* dataset 1 (Supplemental Table 2.1). Secondary motifs represent high-scoring oligomers whose specificity is not captured by the primary motif. Representative 8-mers and their E-scores are provided underneath each motif. A. *PmTbr* primary binding motif. B. *PmTbr* secondary binding motif C. *SpTbr* primary motif. D. Scatterplot of E-scores for each 8-mer in the *PmTbr* vs. the *SpTbr* datasets. The top 14 8-mer matches to the shared primary position weight matrix are indicated in red, while the top 14 matches to the *PmTbr* secondary motif are blue. All 8-mers and their reverse compliments (Supplemental Table 2.1) were assigned sum probability scores based on how well they matched any 8 base pair stretch of *PmTbr* primary position weight matrix (from positions 6–17 shown in A) and *PmTbr* secondary position weight matrix (from positions 7–18 shown in B). The 14 matches to each site are the top 0.02% of 8-mer matches ranked by sum probability score. E-score values indicate the statistical confidence in the seed 8-

mer used in position weight matrix construction, where $E > 0.45$ is considered to be a high-confidence binding event (Berger et al. 2008).

2.4.3 *SpTbr* and *PmTbr* Maintain Similar Affinity for the Conserved Primary Site, but Differ Significantly in their Affinity for *PmTbr*'s Secondary Site

Given that the functional amino acids that differ between *PmTbr* and *SpTbr* involve backbone contacts, we next used surface plasmon resonance (SPR) to determine the affinities that *PmTbr* and *SpTbr* had for each of the identified motifs. Biotin-labeled oligonucleotides were designed to fold into a hairpin containing either the primary site, the *PmTbr* secondary site, the *MmEomes* secondary site, or a nonspecific site that was found to be poorly bound by both Tbr orthologs in the Protein Binding Microarray data (*Pm*, $E = -0.03$, *Sp*, $E = -0.04$) (Figure 2.3A).

Protein association and dissociation, which occur when each protein flows across the sensor chip and when wash buffer removes bound protein respectively, are depicted as sensorgrams (Figure 2.3B). A comparison of this binding response at 100nM Tbr DNA binding domain on each oligomer reveals that neither protein binds the nonspecific site (Figure 2.3B). Additionally, the shape of the sensorgrams indicates that stable equilibrium is reached quickly and, therefore, equilibrium response can be ascertained and used to calculate affinity.

To determine affinities, equilibrium response units (RUs) were taken at 95s into the association phase, where equilibrium is established, as indicated by the slope = 0 in the sensorgrams (Figure 2.3B). Such measurements were taken from sensorgrams corresponding to at least five, but as many as ten, concentrations. Samples of Tbr from each species were applied to the same SPR chip alternately so both proteins were assayed with equal binding conditions. The equilibrium RU values were plotted versus protein concentration and fit to a 1:1 binding model (Adjusted $R^2 > 0.99$) (Figure 2.3C and 3D). Averaged affinity results from four or more experiments across these protein concentrations are shown in Figure 2.3E. *PmTbr* recognizes the primary motif slightly better than does *SpTbr*, with affinities of 107 ± 8 nM for *PmTbr* and 137 ± 7 nM for *SpTbr*. By comparison, *PmTbr* binds the secondary site with significantly greater affinity than does *SpTbr*. *PmTbr* binds the secondary site with an affinity of 446 ± 17 nM and *SpTbr* binds with an affinity of 989 ± 49 nM (Two-tailed t-test, $t=11.612$, $df=6$, $p=0.0007$) (Figure 2.3C and E). Neither echinoderm Tbr ortholog binds particularly well to the *MmEomes*

secondary site; PmTbr binds with an affinity of 732 ± 10 nM and SpTbr with an affinity of 882 ± 153 nM (Figure 2.3D and E).

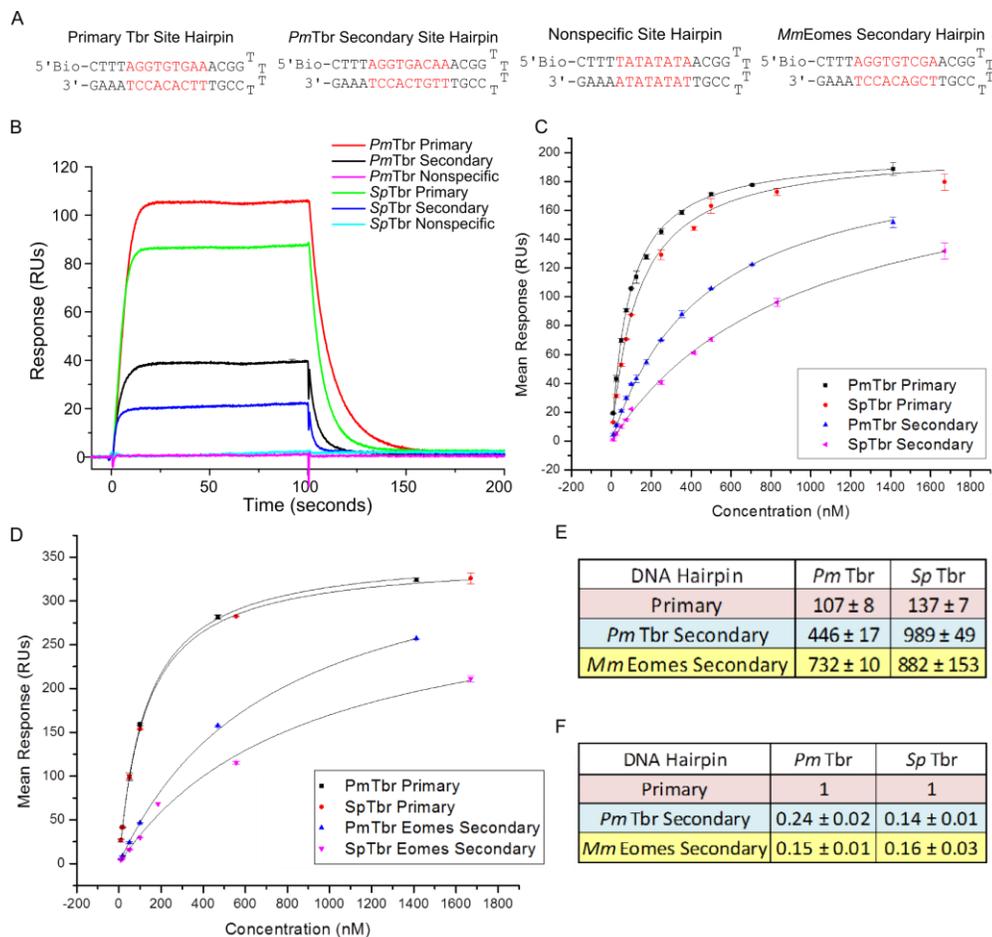


Figure 2.3: Steady State Affinity Evaluations for Tbr DNA Binding Domains. A. DNA sequences of oligonucleotide hairpins used in SPR experiments. Nucleotides depicted in red are the predicted protein binding site. B. Sensorgrams depicting real-time binding of 100nM PmTbr and SpTbr DBD to each biotinylated oligonucleotide. Nonspecific binding was determined using a blank flow cell, which had streptavidin but no DNA bound, and was subtracted from all curves. Equilibrium response (R_{eq}) was taken from these and curves corresponding to all other protein concentrations at 95s. Response curves are also buffer subtracted and represent the average of duplicate samples with corresponding error. Results are representative of typical findings from replicate experiments. C. R_{eq} versus concentration plus 1:1 binding fits for Pm and SpTbr's steady state affinity for primary and PmTbr secondary binding motifs. Data points indicate the

average of duplicate samples plus error from two different concentration series experiments. Errors shown represent standard deviation of data points. D. R_{eq} versus concentration plus 1:1 binding fits to determine *Pm* and *SpTbr*'s steady state affinity for *MmEomes* secondary binding motif. Primary site binding is also shown because this analysis was performed on a different sensor chip than in C. E. Dissociation constants of each Tbr for each oligonucleotide plus standard error of the mean. F. Relative Affinity for each ortholog for each DNA Hairpin plus standard error of the mean. All values are relative to the ortholog's affinity for the primary site. K_{DS} indicate average for two experimental runs, both of which were performed with duplicate scrambled concentration series, with the exception of primary binding site values, which come from data depicted in C and D, and therefore include more experiments.

We also compared relative affinity of *PmTbr* and *SpTbr* for each secondary site vs. affinity for the primary site (Figure 2.3F) by dividing their respective primary site K_D by K_{DS} for all other binding sites. This allowed us to ascertain whether *SpTbr*'s lower affinity for the secondary site could be due to an overall reduction in binding affinity because even *SpTbr*'s affinity for the primary site is slightly lower than *PmTbr*'s. The relative affinity of the secondary site versus the primary site is 0.24 for *PmTbr*, while for *SpTbr* it is significantly lower at 0.14 (Two-tailed t-test, $t = 8.944$, $df = 6$, $p = 0.00022$, Bonferroni corrected). *SpTbr*'s relative affinity for *PmTbr*'s secondary site is comparable to the relative affinity both Tbrs have for *MmEomes*'s secondary site (0.15 and 0.16). *PmTbr* clearly binds its own secondary site better than it binds the *MmEomes* secondary site (Two-tailed t-test, $t = 8.165$, $df = 4$, $p = 0.0024$, Bonferroni corrected). It also has a stronger relative affinity for this site than *SpTbr* has for the secondary site from either *PmTbr* or *MmEomes*.

The data shown in Figure 2.3 provide an independent confirmation of the Protein Binding Microarray data (Figure 2.2 and Supplemental Table 2.1) with an additional quantification of sequence affinity. They show that *PmTbr* has a stronger preference for its secondary motif than does *SpTbr* in spite of the similar affinities these echinoderm proteins have for their primary motif and for the *MmEomes* secondary motif. Although *SpTbr* tends to bind all tested sites with slightly less affinity than does *PmTbr*, it is notable that this is not enough to explain the larger difference in binding observed for the *PmTbr* secondary site, as demonstrated by comparisons of relative affinity.

2.4.4 The Secondary Site Can Substitute for the Primary Site *in vivo* when Tbr Levels are High, but not when they are Reduced

We next wanted to determine how the primary and secondary sites function *in vivo* to regulate transcription in order to understand whether these differences are biologically relevant. We had previously characterized a *cis*-regulatory module (*OtxG*) that controls the expression of the sea star *otx* gene (Hinman et al. 2007) and contains a single endogenous Tbr site that is a perfect match to the Protein Binding Microarray-derived primary motif (Figure 2.4A). We first confirmed that Tbr binds directly to this CRM *in vivo* using Chromatin Immunoprecipitation (ChIP) PCR. ChIP was performed in embryos at 30 hours post-fertilization (h), a time point during which *OtxG* is known to be active (Hinman et al. 2007). We show that the genomic region containing *OtxG* is greatly enriched in chromatin pulled-down by the anti-*PmTbr* antibody compared to input chromatin and mock ChIP chromatin (Figure 2.4B). Importantly, genomic regions 1kb up or downstream of *OtxG* are not enriched in *PmTbr* ChIP DNA (Figure 2.4B).

We next produced a series of constructs to determine how the primary and secondary motifs would behave *in vivo* (Figure 2.4A). “*Basal Promoter GFP*” is a previously existing construct that contains only a basal promoter in a *GFP* expression vector (Hinman et al. 2007). This imparts very low levels of ubiquitous *GFP* expression. The “*OtxG GFP*” construct has the endogenous *OtxG cis*-regulatory module added upstream of the basal promoter. “*2° Tbr GFP*” has a two base pair mutation which changes the endogenous primary motif to a secondary motif. “*Tbr Deletion GFP*” ablates the Tbr binding site by changing the same bases mutated in “*2° Tbr GFP*” but so that the resulting site is one that had an average E-score of -0.058 in the Protein Binding Microarray dataset. By comparison, our motifs selected to represent the primary and secondary position weight matrices had average E-scores of 0.499 and 0.483 respectively. *PmTbr* should, therefore, be unable to bind this site.

These constructs are injected into embryos where they express the reporter gene in clones of cells. In each experiment, our various *GFP* constructs are co-injected with *OtxG mCherry*, which is identical to *OtxG GFP* except that coding sequence for the *mCherry* gene replaces that of the *GFP* reporter. The *OtxG mCherry* construct is used to normalize each sample for differences in injection volume, mosaicism of reporter incorporation, and embryo collection and processing. We used *mCherry* rather than an endogenous housekeeping gene to normalize *GFP*

expression levels as this reporter will also account for injection variation. We do expect that there may be some differences in overall *GFP* versus *mCherry* transcript levels driven by identical CRMs because these mRNA transcripts may have different stability *in vivo*. It is important to note, however, that none of our assays directly compares *GFP* to *mCherry* levels but instead compare *GFP* levels across assays at a single time point that have been normalized to *mCherry*. Therefore, absolute differences in co-injected reporter levels themselves will not affect our analyses.

We assayed the expression of these reporter genes using a combination of approaches. Quantitative reverse transcription PCR (qRT-PCR) was used to determine the abundance of the reporters relative to each other (Figure 2.4). Fluorescent whole-mount *in situ* hybridization (FISH) was used to examine the spatial localization of these reporters (Figure 2.5). We use FISH rather than assays for fluorescent protein localization, as RNA localization is a more direct measure of transcript regulation and should coincide with qRT-PCR. *GFP* and *mCherry* proteins are relatively stable and can persist within the embryo after gene expression is extinguished. We also quantified fluorescent signal strength in whole-mount FISH embryos using ImageJ (Schneider et al. 2012). This last approach allows us to specifically estimate the abundance of each reporter within a particular spatial location (Figure 2.5).

We first performed a series of controls to verify the utility of this reporter system. We confirmed that the *Basal Promoter GFP* construct does not drive significant expression on its own when co-injected with other constructs. *Basal Promoter GFP* drives expression at a roughly ten-fold lower level than *OtxG GFP* in sibling embryos of the same stage (28 hours post fertilization (h)). This indicates that there is no cross-regulation between the *OtxG mCherry* construct used for normalization and the *Basal Promoter GFP* co-injected constructs (Two-tailed t-test, $t=9.082$, $df=12$, $p=0.0002$, Bonferroni corrected) (Figure 2.4C). *Tbr Deletion GFP* expression is also significantly reduced compared to *OtxG GFP*, indicating that the Tbr binding site within *OtxG* is crucial for normal expression levels (Two-tailed t-test, $t=3.305$, $df=12$, $p=0.011$, Bonferroni corrected). Combined, these experiments establish that the validity of this reporter system for assaying primary and secondary site usage *in vivo*. They demonstrate that the basal promoter does not drive any significant expression when co-injected with other constructs and that the Tbr site is a functional *in vivo* binding site.

We then compared the expression driven by our primary and secondary sites using this reporter system. Tbr levels are very high maternally and throughout early development as shown by western blot (Supplemental Figure 2.3A). Using qRT-PCR, we show that 2° *Tbr GFP* and *OtxG GFP* drive expression at roughly the same levels *in vivo* at three early developmental time points; 21h (Two-tailed t-test, $t=0.404$, $df=4$, $p=0.650$), 25h (Two-tailed t-test, $t=1.505$, $df=6$, $p=0.148$), and 28h (Two-tailed t-test, $t= .296$, $df=12$, $p= 1$, Bonferroni corrected) (Figure 2.4C). These data, therefore, convincingly show that Tbr is able to use the secondary site in place of the naturally occurring primary site *in vivo* and with no significant change in transcription of the reporter. This suggests that at these time points, there are sufficient levels of Tbr present to overcome the differential affinity for these sites, and therefore Tbr binds either the primary or secondary site interchangeably to drive gene expression.

We next sought to determine whether the Tbr protein could differentiate between these sites when protein levels are reduced. To this aim, we co-injected each construct with either 400 μ M control morpholino antisense oligonucleotide (MASO) or *PmTbr*-specific translation blocking MASO. These modified oligonucleotides bind in a sequence specific manner to the translation start site of the transcript to block translation and have been used successfully in previous work from our lab (Hinman et al. 2007; McCauley et al. 2010). At this concentration, the Tbr MASO drastically reduces, but does not eliminate, Tbr protein. Knock-down efficiency of all samples was confirmed by assaying for changes in expression of known Tbr target genes, *otx β b* and *delta*, by qRT-PCR (Supplemental Figure 2.3B) (Hinman and Davidson 2007). Therefore, we are confident that our Tbr MASO is reducing levels of Tbr protein. In a Tbr knockdown, 2° *Tbr GFP* drives expression at 40% the level of its expression in sibling Control MASO embryos at 28h (Two-tailed t-test, $t=6.360$, $df=4$, $p=0.0067$, Bonferroni corrected) (Figure 2.4E). To control for any effects that might be associated with the different reporters in this experiment, we show that at 28h, normalized expression of *OtxG GFP* is not significantly different between Tbr MASO and sibling Control MASO embryos (Two-tailed t-test, $t= 1.410$, $df=4$, $p= 0.334$, Bonferroni corrected). Furthermore, when we consider the expression of *OtxG GFP* compared to 2° *Tbr GFP* when they are expressed in Tbr MASO embryos (Figure 2.4E, comparison between red bars), 2° *Tbr GFP* is expressed at significantly lower levels (Two-tailed t-test, $t=3.880$, $df=4$, $p=0.022$, Bonferroni corrected). This demonstrates that even though the 2° *Tbr GFP* construct differs from *OtxG GFP* by only two base pairs, it is significantly more

sensitive to Tbr knockdown than is *OtxG GFP*. This indicates that the secondary binding site is more sensitive to *in vivo* protein levels, as predicted from the *in vitro* affinity data.

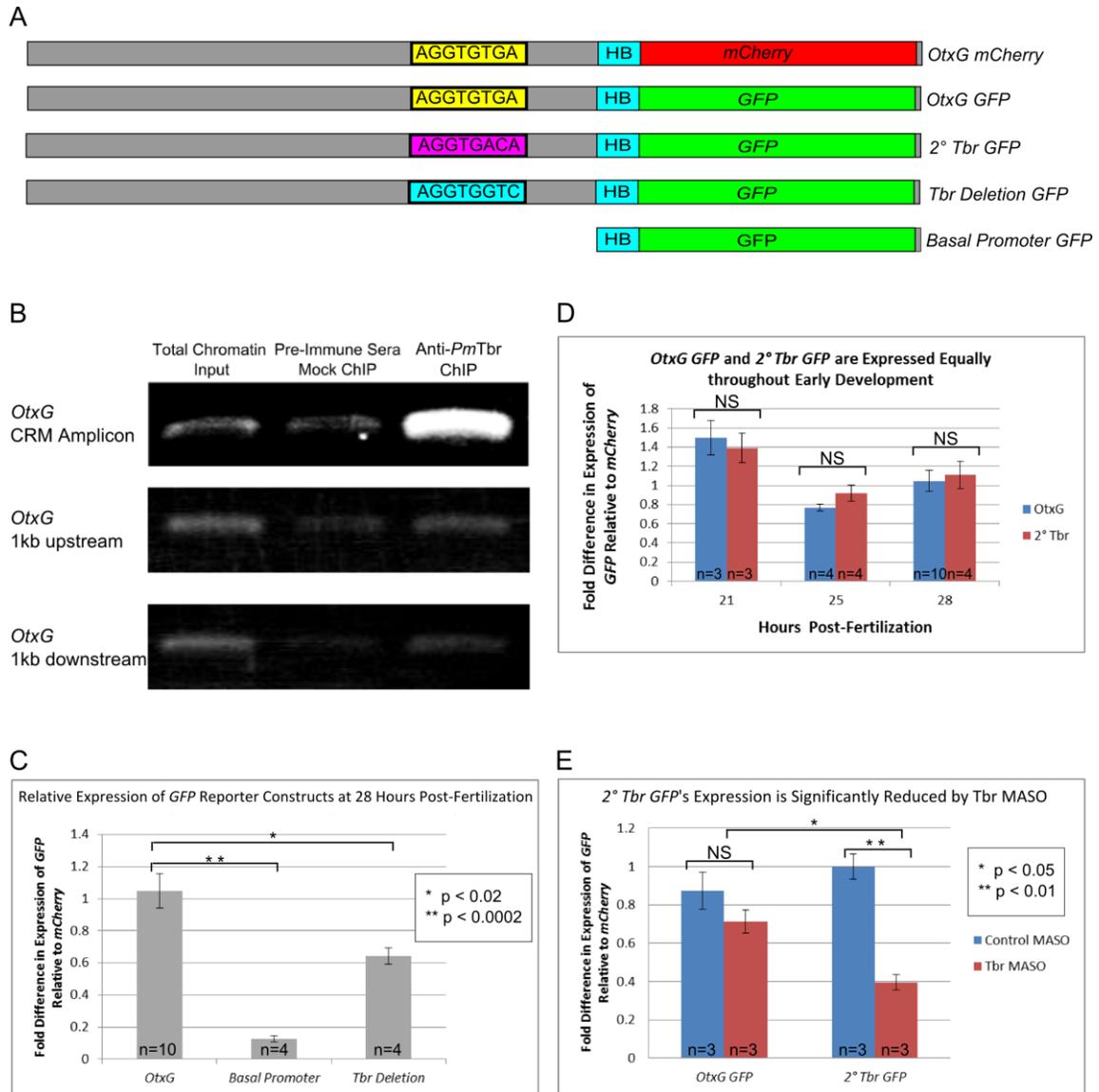


Figure 2.4: *PmTbr* Can Use the Primary and Secondary Sites *in vivo* to Drive Reporter Gene Expression Interchangeably Except when Tbr Levels are Reduced.

A. Schematics depicting *OtxG mCherry*, *OtxG GFP*, *2° Tbr GFP*, *Tbr Deletion GFP*, and *Basal promoter GFP* reporter gene constructs including the endogenous and mutated Tbr binding motifs of interest. B. ChIP PCR using primers pairs surrounding *OtxG* (*OtxG* CRM Amplicon) or primers pairs 1kb up or downstream of *OtxG*. EtBr stained gel shows amplicons obtained from total chromatin, pre-immune sera mock ChIP, and Anti-*PmTbr* ChIP. C-E. QPCR analysis of

GFP expression levels driven by constructs indicated. All *GFP* expression levels have been normalized to *mCherry* levels that were driven by the co-injected *OtxG mCherry* construct. C. Normalized GFP expression levels of *OtxG GFP*, *Basal Promoter GFP*, and *Tbr Deletion GFP* at 28h. D. At developmental time points 21h, 25h and 28 h, Tbr is equally able to drive expression from *OtxG* reporters containing an endogenous primary site and introduced secondary site. The normalized expression level of *GFP* in *OtxG GFP* (blue bars) compared to 2° *Tbr GFP* (red bars) is not significantly different. E. Normalized *GFP* expression levels resulting from 2° *Tbr GFP* or *OtxG GFP* co-injected with control MASO (blue bars) or Tbr (red bars) MASOs. In panels n indicates the number of replicate samples, each consisting of 50 sibling embryos. All error bars indicate Standard Error of the Mean. P-values indicate the results of a Two-tailed t-test. Details of these tests are provided in the main text. NS indicates not significant by Two-tailed t-test.

2.4.5 The Secondary Site Responds Faster to Tbr's Endogenous Temporal Gradient

We wanted to determine whether the secondary and primary binding sites would respond differently to endogenously changing levels of Tbr. To test how the primary and secondary sites might differ in their response to a temporal decline in Tbr levels, we first determined when Tbr decreases endogenously. Tbr levels are high maternally, which makes it difficult to determine how genes respond to zygotic Tbr levels as the gene's transcription is initiated (Supplemental Figure 2.3A). However, we see that during the later gastrula stages, between 54h and 65h, Tbr goes from being localized broadly throughout the ectoderm (31h and 52h embryos) to being specifically localized within the ciliary band territory within the ectoderm (Figure 2.5A). We also see an overall reduction in Tbr levels between 48h and 70h by western blot (Supplemental Figure 2.3A). The *otx* gene, regulated by Tbr through the *OtxG* CRM, has a similar progression of its expression domain and time course (Hinman, Nguyen, and Davidson 2003).

We therefore determined whether expression driven by the 2° *Tbr GFP* reporter extinguishes more rapidly in the ectoderm between 54h and 65h than that driven by *OtxG*. We examined the expression of GFP and mCherry reporters using FISH. In all of these stages, endoderm expression of Tbr is high (Hinman, Nguyen, Cameron, et al. 2003), which necessitates spatial comparison of transcripts localized to the ectoderm as opposed to qRT-PCR, which can only determine global transcriptional levels. We examined the spatial co-expression of *GFP* and

mCherry in appropriately staged embryos and then quantified levels of expression in these cells. As in our qRT-PCR experiments, we normalize the level of *GFP* expression driven by *OtxG GFP* and 2° *Tbr GFP* to *mCherry* levels driven by *OtxG mCherry*. We first confirmed that *OtxG GFP* and *OtxG mCherry* co-express in the same cells in early (28h, Figure 2.5B-B'') and late development (56h, Figure 2.5D-D'') so that *mCherry* expression can be used for normalization of fluorescent intensity. We next show that 2° *Tbr GFP* and *OtxG mCherry* also co-express in the same set of cells at these time points (Figure 2.5C-C'' and 5E-E''). Finally, we quantify and compare the normalized *GFP* expression driven by primary and secondary motifs in early development (28h) when *Tbr* levels are high, and in late development (56h) when *Tbr* levels are low.

At 28h, we show that *OtxG GFP* does not drive significantly different expression in the ectoderm compared to 2° *Tbr GFP* (Two-tailed t-test, $t=0.663$, $df=18$, $p=0.987$, Bonferroni corrected). Thus, at this stage, as predicted by our earlier quantitative assays, there is no effect of primary versus secondary binding site on the abundance of reporter gene expression, and we also show here on spatial localization. When we compare the expression of 2° *Tbr GFP* to *OtxG GFP* at 56 h, however (compare ratio of E'/E'' to D'/D''; Figure 2.5E), we find that 2° *Tbr GFP* reporter is expressed in reduced patches and at visually lower levels. Quantification of fluorescent intensities of normalized *GFP* signals demonstrates significant reduction of 2° *Tbr GFP* expression relative to *OtxG GFP* (Two-tailed t-test, $t=6.109$, $df=28$, $p=.0000019$, Bonferroni corrected). These data (Figure 2.4 and 5) show that a two base pair change from the higher-affinity primary to the lower-affinity secondary *Tbr* binding site is sufficient to elicit a response to reduced *Tbr* levels that is more pronounced than the wild type response.

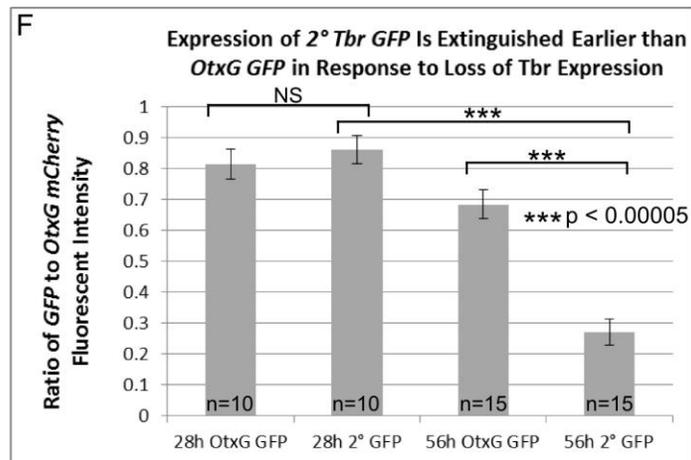
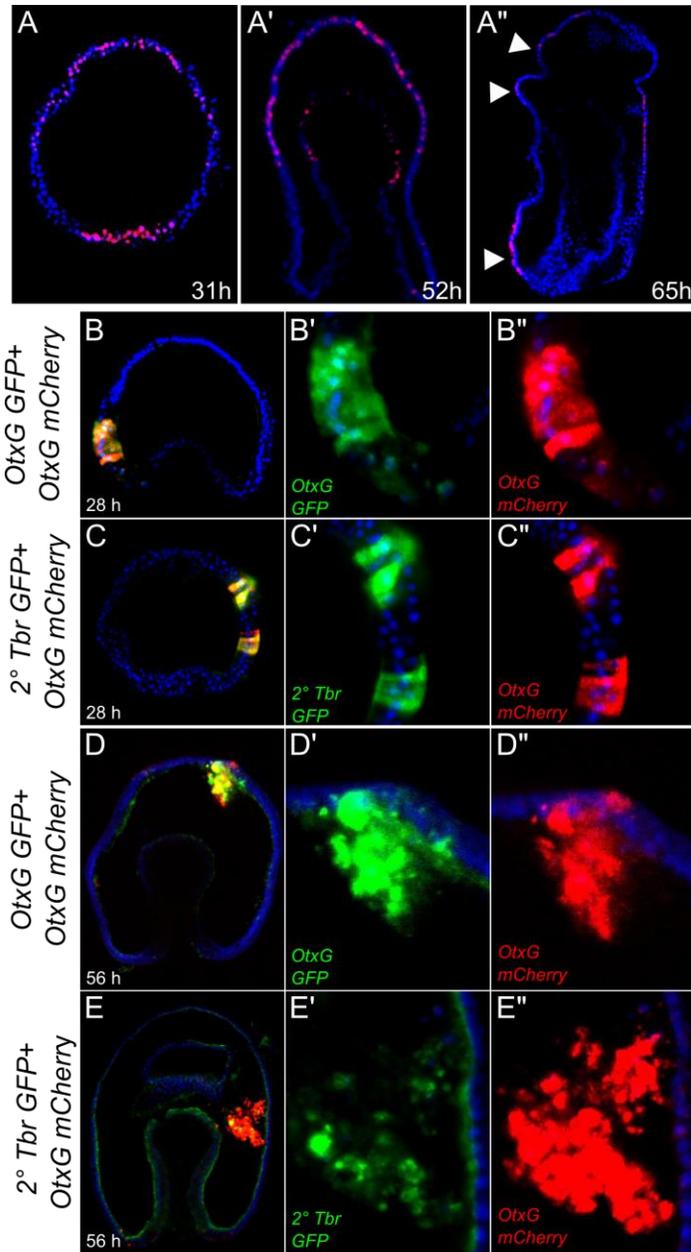


Figure 2.5: Secondary Tbr Reporter has Reduced Expression Compared to OtxG in the Ectoderm when Tbr Levels are Declining. A-A". In all panels, blue indicates DAPI nuclear stain and red indicates Tbr localization. A. 31 h blastula stage *P. miniata* embryo. A'. 52 h gastrula stage embryo A". 65 h late gastrula stage embryo. Arrow heads indicate localization which is present in only the ciliary band ectoderm by 65h. B-E". In all panels, blue indicates DAPI nuclear stain, red indicates *mCherry* transcripts labeled by CyIII, green indicates *GFP* transcripts labeled by fluorescein. B, C, D and E depict the entire embryo with merged expression, while B'-B", C'-C", D'-D" and E'-E" are insets of the region of interest for each probe. B-C". *OtxG GFP* and 2° *Tbr GFP* both co-express spatially with *OtxG mCherry* at 28h D-D". *OtxG GFP* reporter co-injected with *OtxG mCherry* at 56h. The reporters are still spatially co-expressed at this stage. E-E" 2° *Tbr GFP* reporter co-injected with *OtxG mCherry* at 56h. *GFP* expression is reduced compared to *OtxG GFP* while *mCherry* levels remain more consistent. F. Quantification of fluorescent intensities of fluorescein (*GFP*) relative to CyIII (*mCherry*) at 28h and 56h. N indicates the number of embryos imaged. Error bars indicate Standard Error of the Mean. P-values indicate the result of Two-tailed t-tests, which are described in the Results.

2.5 Discussion

There has been a great deal of interest and controversy surrounding theories of how developmental GRNs might evolve. Debate has centered on the effects that protein versus *cis*-regulatory mutations may have on the capacity for change in a GRN. Much work suggests that CRM variation is the prominent source of change to GRNs and evolution of novel phenotypes (reviewed in Wray 2007; Rebeiz and Williams 2011; Wittkopp and Kalay 2012; Rubinstein and de Souza 2013). There are many explanations for why CRMs are so equipped to evolve, but a crucial source of their evolutionary flexibility is their modularity. A single gene is frequently regulated by many CRMs, each CRM orchestrating expression of that gene in a specific spatio-temporal context (Arnone and Davidson 1997). So then, a particular CRM for a given gene can be lost, gained, or altered independently from all of the other CRMs, and likewise, binding sites within a CRM can be lost, gained, or altered independently from the rest of the sites within the CRM. These properties create a scenario with very little pleiotropy, and as a result, a great deal of evolutionary freedom.

A key to understanding how protein changes can affect GRNs therefore is to understand the ways that proteins can themselves evolve in ways that reduce pleiotropy. In actuality, proteins are often composed of multiple domains which may be gained, lost, and changed independently of each other to create diverse proteins (Levitt 2009; Wang and Caetano-Anollés 2009; Kersting et al. 2012). Each domain has the capacity to be modified individually and some of these modifications may limit the activity of the protein to a specific time and place. A novel protein-protein interaction, for example, might limit the activity of a protein to contexts where it is co-expressed with its new cofactor. It is unsurprising then that changes in protein-protein interactions (Löhr and Pick 2005; Tuch et al. 2008) and post-translational modifications (Lynch et al. 2011) also allow for the evolution of novel features and rewiring of gene regulatory networks.

Understanding of how transcription factors might directly evolve changes in DNA binding properties has been less clear. Outside of a few striking examples (Hanes and Brent 1989; Baker et al. 2011; Nakagawa et al. 2013b), it has been considered that this feature of transcription factor function will remain highly conserved and will not represent a substantial source of evolutionary novelty. Recent work, however, demonstrates that DNA binding properties also have a capacity to be modular as they can have secondary or alternative binding preferences in addition to their primary or most preferred binding site (Badis et al. 2009; Gordân et al. 2011; Busser et al. 2012; Nakagawa et al. 2013b). Other work reveals that transcription factors need multiple binding sites that differ in affinity because they are crucial for executing unique developmental functions (Rowan et al. 2010; Peterson et al. 2012). In the *Drosophila* mesoderm, many homeodomain transcription factors are co-expressed and share a primary binding motif. Use of secondary binding sites, which are unique to a particular paralog, allows different homeodomain paralogs to bind appropriate CRMs and execute discrete developmental functions (Busser et al. 2012). The ability to use multiple binding site sequences imparts flexibility in gene regulation and is crucial for developmental functions of these transcription factors. Several surveys of transcription factors indicate that secondary binding preferences are common and frequently differ between paralogous transcription factors (Badis et al. 2009; Gordân et al. 2011). Paralog diversity, however, represents an evolutionary scenario particular to gene duplication events. A pair of paralogs originates from a single protein and, therefore, they are often able to divide the responsibilities of the original protein between them. In some

cases, one paralog maintains all the functions of the original protein and the other is free to neofunctionalize (Plaitakis et al. 2003; Zhang et al. 2004; Lee and Irish 2011). In either case, this division of labor relieves evolutionary constraint on one or both paralogs and may allow new secondary binding preferences to evolve.

Here, we demonstrate for the first time that orthologous transcription factors also diversify by evolving differences in secondary motif binding. We show that the two echinoderm Tbr orthologs, *SpTbr* and *PmTbr*, bind a highly similar primary motif. This motif also matches the previously published primary motif of *MmEomes* (Badis et al. 2009). *SpTbr* and *PmTbr* recognize that motif with similar affinity. Importantly, we determine that there is a greater evolutionary variation in secondary binding motif preference since echinoderms and vertebrates last shared an ancestor. We find that *PmTbr* and *MmEomes* recognize distinct secondary motifs, while, the sea urchin *SpTbr* does not have any significant secondary motif preference and has a significantly reduced ability to bind *PmTbr* and *MmEomes*'s secondary motifs.

The fold-changes in binding site affinity that we determine here between preferences for the sea star primary and secondary motifs are the same order of magnitude as observed between different classes T-box transcription factors for a consensus primary site. For example, Macindoe *et al.* (2009) determined the affinities that three divergent T-box proteins, human Tbx5 (*HsTbx5*), Mouse Tbx20 (*MmTbx20*) and human Tbx2 (*HsTbx2*), had for their consensus primary sequence, AGGTGTGA. This work demonstrated that *MmTbx20*, *HsTbx5*, and *MmTbx2* bound to this site with affinities of 913 nM, 232 nM, and 1511 nM respectively. It was suggested that this difference in affinity, which is less than two-fold between *MmTbx20* and *MmTbx2*, could be functionally significant and permit the competitive, hierarchical gene regulation known to occur when these transcription factors are co-expressed in the developing heart (Macindoe et al. 2009).

This study is the first demonstration of this type of evolutionary change in orthologous transcription factor function. This finding points to a previously overlooked source of modularity for evolution to exploit and, therefore, to a mechanism for allowing a transcription factor to evolve a new function. We speculate that *PmTbr* may be able to carry out multiple developmental functions simultaneously by dividing them among its two binding motifs. *PmTbr* is needed for the correct specification of endoderm, mesoderm, and ectoderm during sea star embryogenesis (Hinman et al. 2007; Hinman and Davidson 2007; McCauley et al. 2010).

Meanwhile, *SpTbr* has a single role in the sea urchin embryo, which is to specify skeletogenic mesenchyme (Croce et al. 2001; Oliveri et al. 2002). Even within the skeletogenic network, *SpTbr* has relatively few inputs into skeletogenic genes (Rafiq et al. 2012) suggesting that it is a much less pleiotropic gene than *PmTbr*. In hemichordates and cephalochordates, the *Tbr* ortholog is also expressed in multiple embryonic tissue types, including endoderm and ectoderm (Tagawa et al. 2001; Horton and Gibson-Brown 2002), suggesting that these orthologs and *PmTbr* may share an ancestral function in the endoderm and ectoderm that must have been lost in sea urchins.

The ability to divide functions between different binding motifs has potential to be very useful during development because a limited number of regulatory molecules must orchestrate the specification of an increasingly complex embryo. Ideally, such regulatory molecules will be as multifunctional as possible to allow development to progress rapidly and create diverse cell types. Yet, this pleiotropy is what causes transcription factors to be evolutionarily constrained. Our finding, that these functions can be uncoupled and evolve independently through separate binding sites offers a mechanism by which new features can arise.

We also demonstrate that the secondary binding site is more responsive to changes in *Tbr* protein levels during development. This quality is particularly important for functions that require rapid transcriptional responses and may be especially important during early development where the timing of developmental events must be precisely coordinated. We predict such affinity differences are also advantageous when a rapid transcriptional response is required during development for some, but not all target genes (Figure 2.6). Such targets can make use of more sensitive, lower affinity secondary sites.

It is often assumed that transcription factors are under an enormous amount of evolutionary constraint because they regulate large numbers of target genes. Presumably these targets are essential to the organism and must be maintained by all orthologs that arise by speciation. However, if these target genes are subdivided into groups based on the binding sites they are regulated by, then there are fewer genes affected by changes in binding preference. This reduces pleiotropy, because a loss of ability to use a secondary site would affect only a subset of target genes while others would be regulated normally (Figure 2.6). *SpTbr* should be able to maintain developmental functions associated with the primary site, yet its reduced ability to utilize a secondary site may have led to evolutionary differences in cell patterning and

specification between these species. This modification in function between orthologs will not only lead to a dramatic loss or gain of target genes, but also offers a mechanism to affect timing control of gene regulation. Change in relative order or timing of developmental events can be acquired by evolving higher or lower affinity for a secondary binding site. We hypothesize that this newfound source of modularity in orthologous transcription factors offers a previously overlooked source of gene regulatory network evolutionary change.

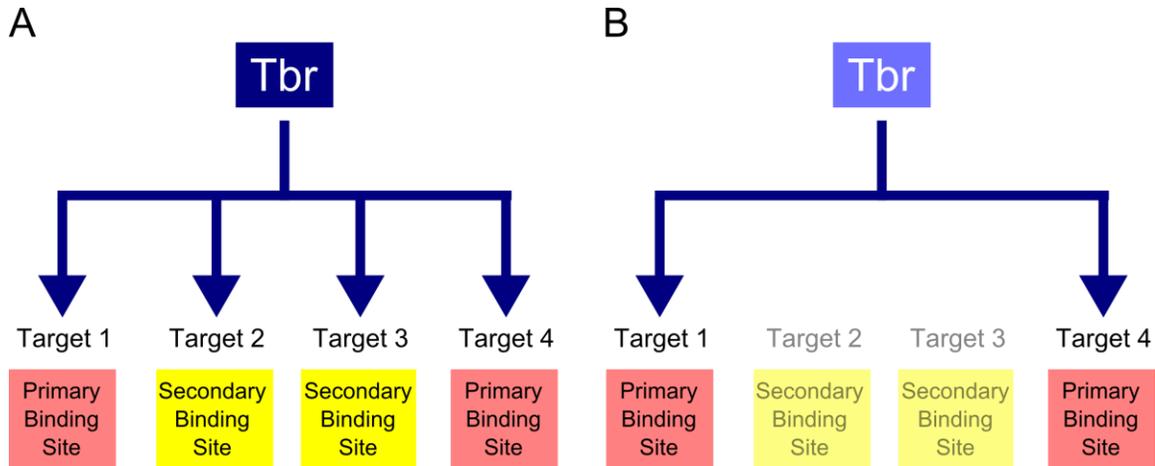


Figure 2.6: Modular Binding of Tbr may Allow for Diverse Transcriptional Responses during Development and Allow for Greater Evolvability. A. When *PmTbr* levels are high, transcription of target genes can be activated via primary and secondary sites. Activated targets are denoted by arrow inputs. However, when *PmTbr* levels are low (B), only genes regulated via primary sites are activated, while those that use secondary sites will have no or reduced transcription, which are shown with no arrows. Because *SpTbr* has reduced affinity for the secondary site, it will encounter the later scenario, shown in B, more frequently and may never have an opportunity to activate target genes that are dependent on secondary sites.

2.6 Methods

Phylogenetics: Tbr orthology was established using a MrBayes model (JTT plus Gamma), 5 runs, 100,000 generations, sampling frequency of 10, in TOPALi v2.5 (Milne et al. 2004). Branches are supported by posterior probability. The T-box domain alignment of all represented proteins was generated by Clustal Omega (Sievers et al. 2014) and is shown in Supplemental Figure 2.1. Accession numbers are listed in Supplemental Figure 2.1.

Protein Expression and Purification of DNA Binding Domains: GST fusion protein constructs for protein binding microarray and SPR were made by cloning T-box sequences into pKM vector and were purified from BL21 *E. coli*. The T-box domain constructs consisted of residues 272-466 of *PmTbr* and residues 362-554 of *SpTbr* to include the whole T-box plus five amino acids flanking each side. Cultures were grown at 20°C and protein expression was induced by addition of 0.2 mM IPTG at OD600 0.5 and growth was continued overnight. Cell pellets were resuspended in PBS Triton x-100 (0.1% v/v) (pH 7.5) for protein binding microarrays or 20 mM Mops (pH 7.5), 150 mM NaCl, 1 mM DTT, 0.005% Surfactant P20 (v/v) for SPR. In both cases, Complete Protease Inhibitors (Roche Diagnostics, Indianapolis, IL, USA) were added just prior to use and cells were lysed by sonication. All fusion proteins were purified by GSH affinity chromatography (Thermo Scientific Pierce, Rockford, IL, USA). For protein binding microarray experiments, glycerol was added to eluted proteins to 10% (v/v) and single use aliquots were flash-frozen and stored at -80°C. For SPR protein samples, T-box DNA Binding Domains were cleaved from GST-His on beads by treatment with TEV protease (Eton Bioscience, San Diego, CA, USA). DNA binding domains were then flash-frozen and stored at -80°C in single-use aliquots.

Protein Binding Microarrays : Custom-designed, ‘universal’ oligonucleotide arrays (Agilent Technologies, AMADID #016060 (Zhu et al. 2009)) were converted to double-stranded DNA arrays by primer extension and used in Protein Binding Microarray experiments essentially as described previously (Berger et al. 2006). 200 nM samples of *PmTbr* and *SpTbr* were assayed in PBS (pH 7.5). Two replicate datasets for each protein are reported in Supplemental Table 2.1. Microarrays were scanned and quantified, and then analyzed using the Universal PBM Analysis Suite and the Seed-and-Wobble motif derivation algorithm as described previously (Berger et al. 2006; Berger and Bulyk 2009).

Surface Plasmon Resonance: The sequences of 5’ Biotin labeled hairpin DNA oligomers are depicted in Figure 2.3A. 25 nM stocks of hairpin oligomers were diluted in HBS-EP buffer (0.01 M HEPES, pH 7.4, 0.15 M NaCl, 3 mM EDTA, 0.005% Surfactant P20). These were applied to a streptavidin-coated CM5 chip, prepared according to (Nguyen et al. 2006), with minor modifications for a Biacore T100 SPR instrument. The first flow cell was left blank for

reference subtraction, while primary, secondary, and nonspecific DNA hairpins were immobilized to flow cells 2-4 respectively such that each had 150 response units of DNA. Separate chips were made to assess affinity for *PmTbr* secondary and *MmEomes* secondary sites (both on flow cell 3 of their respective chips). Both chips were designed with the primary site hairpin on flow cell 2 and nonspecific hairpin on flow cell 4. Because the maximal binding capacity of each chip was not equivalent, this necessitated that the data shown in Figure 2.3C and 3D be split into separate graphs. The sensor chip was washed several times in running buffer prior to use (50mM Mops, 150 mM NaCl, 1 mM DTT, and 0.01% (vol/vol) P20 surfactant). Kinetic measurements were performed at 20°C with a flow rate of 30 µl/min. Tbr DNA binding domain protein samples were run alternately across the same chip, and all four flow cells were exposed to a sample simultaneously. The concentration series was scrambled for each protein. Immediately following protein injection, buffer was injected to monitor dissociation. Zero concentration (buffer only) samples were included and used to subtract background from protein samples. Data was analyzed first using the BIAevaluation software to determine steady-state response levels for each concentration 95 seconds after injection start. This data was then evaluated using Origin and a 1:1 binding model to determine K_{DS} .

Embryo culture and injection: *P. miniata* embryos were obtained and injected as described in (Hinman, Nguyen, Cameron, et al. 2003) and Cheatle Jarvela and Hinman 2014 (In Press).

Reporter Expression Constructs: *OtxG GFP* and *Basal promoter GFP* reporter constructs were developed by Hinman *et al.*, 2007. *2° Tbr GFP*, *Tbr Deletion GFP*, and *OtxG mCherry* were developed from these existing constructs using the methods described in Hinman *et al.*, 2007. Primer sequences are provided in Supplemental Table 2.2.

Fluorescent *In Situ* Hybridization (FISH): FISH was performed as previously described (Yankura et al. 2010) using digoxigenin-or dinitrophenol labeled antisense RNA probes targeted to *GFP* and *mCherry* respectively. Samples consisted of cohorts of sibling embryos injected with either *OtxG GFP* plus *OtxG mCherry*, or *2° Tbr GFP* plus *OtxG mCherry*. Embryos were reared at 15°C until 28h or 56h.

Image Analysis: FISH embryos were imaged with a Carl Zeiss LSM-510 Meta DuoScan Inverted Confocal Microscope. Laser power, gain, and digital offset settings were optimized for embryos injected with *OtxG GFP* plus *OtxG mCherry*, and then left unchanged for subsequent imaging of sibling embryos injected with 2° *Tbr GFP* plus *OtxG mCherry*. The relative fluorescence of *mCherry* transcripts (CyIII) to *GFP* transcripts (fluorescein) was quantified using ImageJ (National Institutes of Health, Bethesda, MD). All images were background subtracted using 'BG subtraction from ROI' plugin prior to analysis. The 'Measure' function was used to determine the mean fluorescence value of a region in interest for both channels.

Quantitative RT-PCR (qPCR): Total RNA from injected embryos was obtained using GenElute Mammalian Total RNA kit (Sigma, St. Louis, MO, USA). The total RNA was used to make cDNA using iSCRIPT™ Select cDNA synthesis kit (Bio-Rad, Hercules, CA, USA). QPCR was performed according to Hinman et al. (2003b) using an Applied Biosystems 7300 Real-Time PCR system along with SYBR® green PCR master mix. The threshold cycle number (Ct) was normalized to nuclear pore protein, *lamin2β receptor* (Accession: KJ868807) (Supplemental Figure 2.3B) for endogenous gene expression or *mCherry* mRNA for reporter gene expression (Figure 2.4C-E). Primer sequences are provided in Supplemental Table 2.2.

Immunofluorescence: *P. miniata* embryos were fixed in 4% paraformaldehyde/PBS for 20 min at RT, followed by permeabilization in 1% Triton X-100/PBS for 10 min. Embryos were then washed four times in PBS/0.1% Triton X-100 and post-fixed in ice cold methanol for 20 min. After another four washes, embryos were blocked in 3% BSA/PBS for 30 min and incubated with anti-*PmTbr* (1:500) overnight at 4°C. Affinity purified polyclonal anti-*PmTbr* was produced in rabbits by Piece Custom Antibody Services. Embryos were washed four times and incubated in 1:100 FITC anti-rabbit (Sigma) overnight. Embryos were incubated in 1:10,000 DAPI (Life Technologies) for 30 min, washed four times in PBS/0.1% Triton X-100. Embryos were imaged in Slowfade mounting media (Life Technologies) by confocal microscopy.

ChIP-PCR: ChIP was carried out as described by (Mortazavi et al. 2006), with several modifications for sea star embryo samples. Chromatin extraction was performed as follows.

Roughly 10^5 *P. miniata* embryos ($\sim 10^8$ cells) were collected at 30 hours post fertilization. These were cross-linked in 1% formaldehyde in artificial sea water for 10 minutes, stopped with 0.125M glycine, collected by centrifugation, and washed 3x in cold PBS. Embryos were resuspended in lysis buffer (5 mM 1,4-piperazine-bis-[ethanesulphonic acid] (pH8.0), 85 mM KCl, 0.5% NP-40, Complete Protease Inhibitors (Roche Diagnostics, Indianapolis, IL, USA). After 10 minutes of lysis on ice, the embryos were passed through a 25 gauge needle 5-10 times and centrifuged to collect the crude nuclear preparation. Chromatin was digested to 500-100 bp pieces by micrococcal nuclease (New England Biolabs, Ipswich, MA, USA) according to the SimpleChIP® Enzymatic Chromatin IP Kit protocol (Cell Signaling Technology, Danvers, MA, USA). The nuclear pellet was collected by centrifugation and lysed on ice for 10 minutes in 50 mM Tris (pH 8), 10 mM EDTA, 1% SDS (w/vol), protease inhibitors. After the lysate was clarified by centrifugation, small aliquots were flash-frozen for immunoprecipitation, which was performed as described (Mortazavi et al. 2006).

Enrichment of the *PmOtxG* regulatory region was examined by PCR. A primer set was designed for an amplicon within the 850 bp CRM. Amplicons corresponding to regions 1kb upstream and 1kb downstream of *OtxG* were used as negative controls. Primer sequences are available in Supplemental Table 2.2. PCR was performed for 30 cycles to achieve a linear range with the following conditions: : 94 °C for 30 s, 58 °C for 30 s, and 72 °C for 20 s. All reactions contained 1ng template (Total chromatin, mock ChIP, or Tbr ChIP). Products were analyzed by 1% agarose gel.

2.7 Acknowledgements

The authors would like to thank Adam Foote and Greg Cary for their thoughtful critiques of this manuscript. Dr. Cary also provided assistance with statistical analysis and presentation. We also thank Mark MacBeth for assistance with protein purification and Timothy Jarvela for support with imaging and image analysis. We are especially thankful for three anonymous reviewers whose thoughtful comments and suggestions much improved earlier versions of this manuscript. This work was supported in part by NSF IOS 0844948 (V.F.H.) and NIH/NHGRI grant # R01 HG003985 (M.L.B.). L.B. received additional funding through HHMI 52006917. SPR instrumentation was purchased with support from NSF Major Research Instrumentation Award 0821296 (B.A.A.).

SpTbx4	GITATLAQSSMWRKFHECETEMIINRSGRRMFPCFAVSLSGLQPDALYRISVTITSDNRS	60
PmTbx4	GIQVTLNENSKLWKMFNKCGTEMIIVNRIGRRMFPCVITMLGMDPTTLYRVQMELDASDRR	60
SkTbx4	GVTVSLDEPDLWRFHKKHGTTEMIINRTGRRMFPCIGVQISGLEPAALYSVEMEMVMSDR	60
DrTbx2a	DPKVTLEAKELWDQFHKIGTEMVITKSGRRMFPFPKVRVNLGDKKAKYIILMDIVAADD	60
XtTbx2	DPKVTLEAKELWDQFHKLGTTEMIITKSGRRMFPFPKVRVNLGDKKAKYIILMDIVAADD	58
MmTbx2	DPKVTLEAKELWDQFHKLGTTEMIITKSGRRMFPFPKVRVNLGDKKAKYIILMDIVAADD	60
SpTbx2/3	DPQVTLSEKELWEKFFHRRGTTEMIITKSGRRMFPFPKVRVNLGDKKAKYIILMDIVAADD	60
PmTbx2/3	DPQVTLSEKELWDQFHRRGTTEMIITKSGRRMFPFPKVRVNLGDKKAKYIILMDIVAADD	60
MmBra	ELRVGLEESELWLRFKELTNEMIVTKNGRRMFPVLKVNVSGLDPNAMYSLDFVADNH	60
XlBra	ELKVSLEERDLWTRFKELTNEMIVTKNGRRMFPVLKVSMSGLDPNAMYTVLDFVAADNH	60
SpBra	GLKVRLLDDVELWKKFHKLTNEMIVTKSGRRMFPVLSASVGLDPNSMYSVLLDFSAADDH	60
PmBra	GLKVTLEDRLWRRFSLKLTNEMIVTKGRRMFPVLSASVGLNPNAMYSLDFTPADEH	60
MmTbx21	KLRVALSNHLLWSKFNQHQTEMIITKQGRRMFPFLSFTVAGLEPTSHYRMFVDDVVLVDQH	60
XtTbx21	KVQITLTLNYSLWKKFHKHQTEMIITKQGRRMFPFLSFRVAGLDPVAQYNLHVDDVVLADQN	60
DrTbx21	KTQVLLNNYPLWAKFHKKYQTEMIITKQGRRMFPFLSFNITSLDPSAHYNIYVDVVLADQH	60
MmEomes (Tbr2)	RAHVYLCNRPLWLKFFHRHQTEMIITKQGRRMFPFLSFNINGLNPTAHYNVFVEVVLADPN	60
XtEomes	RAQVFLCNRPLWLKFFHRHQTEMIITKQGRRMFPFLSFNITGLNPTAHYNVFVEVVLADPN	60
DrEomes	RAQVYLCNRPLWLKFFHRHQTEMIITKQGRRMFPFLSFNITGLNPTAHYNVFVEIVLADPN	60
DrTbr1	KAQVYLCNRALWKFHRRHQTEMIITKQGRRMFPCLTFNVSGLDPAGHYNIIVDVIADPN	60
MmTbr1	KAQVYLCNRPLWLKFFHRHQTEMIITKQGRRMFPFLSFNISGLDPTAHYNIIVDVIADPN	60
XtTbr1	KAQVYLCNRPLWLKFFHRHQTEMIITKQGRRMFPFLSFNISGLDPTAHYNIIVDVIADPN	60
PpTbr	KASVFLCNSELWKRKFHEHRTEMIITKQGRRMFPQLVFRSLNPAAHYNVFVDMVIADPN	60
PmTbr	KASVFLCNSELWKRKFHEHRTEMIITKQGRRMFPQLVFRSLNPAAHYNVFVDMVIADPN	60
PlTbr	KAVVYLCNRDLWRKFHQHTEMIITKQGRRMFPQLVYKLSGLNPTSQYNVFDVDMVLCDPN	60
LvTbr	KASVYLCNRDLWRKFHQHTEMIITKQGRRMFPQLVFKLTGLNPTSQYNVFDVDMVLCDPN	60
SpTbr	KASVYLCNRDLWRKFHQHTEMIITKQGRRMFPQLVFKLTGLNPTSQYNVFDVDMVLCDPN	60
HpTbr	KASVYLCNRDLWRKFHQHTEMIITKQGRRMFPQLVFKLTGLNPTSQYNVFDVDMVLCDPN	60
SmTbr	RAAVYLCNRDLWRKFHQHTEMIITKQGRRMFPQLVYKLSGLDPTQYNVFDVDMVLCDPN	60
PjTbr	RASAYLCNRQLWRKFHHKTEMIITKQGRRMFPQLVFKLTGLDPTQYNVFDVDMVLCDPN	60
BfTbr	KLSVFLTNRLDLWVKFHQHTEMIITKQGRRMFPVLQFAISGLDPAQYNVFDVDMVLADVN	60
PfTbr	KACVYLCNRDLWLKFFHQHTEMIITKQGRRMFPPLSFRFTGLDPSAHYNVFVDMVLSDPN	60
SkTbr	KACVYLCNRDLWLKFFHQHTEMIITKQGRRMFPPLSFRFTGLDPTAHYNVFVDMVLSDPN	60

* : * * . . ** : . : * * * * . . . : * . . . :

SpTbx4	RYKFINGKWLAVGKADPEMP-NEPYEHPLSPNHGLFWESNVVSFAKLKITNNKDTKAK--	117
PmTbx4	RYKFINGKWPVPGKADAEPP-NKLFHHPDPSPLGAFWMQDRVSFAKLKITNNQETGG---	116
SkTbx4	RYKFIHNKWLPIGKADSDIN-NTPFHHPDSTARGSFWMNSKVSFAKVKITNNKENLG---	116
DrTbx2a	RYKFHNSRWVAVGKADPEMP-KRMYIHPDPSPATGEQWMAKPVAFHKLKLTNNISDKHG--	117
XtTbx2	RYKFHNSRWVAVGKADPEMP-KRMYIHPDPSPATGEQWMAKPVAFHKLKLTNNISDKHG--	115
MmTbx2	RYKFHNSRWVAVGKADPEMP-KRMYIHPDPSPATGEQWMAKPVAFHKLKLTNNISDKHG--	117
SpTbx2/3	RYKFHNSRWVAVGKADPEMP-KRMYIHPDPSPTGEGWQKCVSFHKLKLTNNISDKHGF-	118
PmTbx2/3	RYKFHNSRWVAVGKADPEMP-KRMYIHPDPSPTGEGWQKTVSFHKLKLTNNISDKHGFV	119
MmBra	RWKYVNGEWWPGGKPEPQAP-SCVYIHPDPSPNFGAHWMKAPVVSFVKVLTNKLNGGG---	116
XlBra	RWKYVNGEWWPGGKPEPQAP-SCVYIHPDPSPNFGAHWMKDPVVSFVKVLTNKMNGGG---	116
SpBra	RWKYVNGEWIPGGKPDGSP-PTAYIHPDPSPNFGAHWMKQAVNFSKVKLSNKLNGSG---	116
PmBra	RWKYVNGEWWPGGKPDSP-PTAYIHPDPSPNFGAHWMKQSVVSFVKVLSNKLNGTG---	116
MmTbx21	HWRYQGGKWWQCGKAEGSMGNRNYVHPDPSNTGAHWMRQEVSFGKLLTNNKGASNNVT	120
XtTbx21	HWRYQGGKWWQCGKAEGSNRNYVHPDPSNTGAHWMRQEVSFGKLLTNNKGASNNVS	120
DrTbx21	HWRYQGGKWWQCGKAEGSNRMYHPDPSNTGTHWMRQEVSFGKLLTNNKGSNNVA	120
MmEomes (Tbr2)	HWRFGGKWWTCGKADNNMQGNKMYVHPESNTGSHWMRQEISFGKLLTNNKGANNNT	120
XtEomes	HWRFGGKWWTCGKADNNMQGNKMYVHPESNTGSHWMRQEISFGKLLTNNKGANNNT	120
DrEomes	HWRFGGKWWTCGKADNNMQGNKMYVHPESNTGSHWMRQEISFGKLLTNNKGANNNT	120
DrTbr1	HWRFGGKWWPCGKADTNVGNRVYTHPDSNTGAHWMRQEISFGKLLTNNKGASNNNT	120
MmTbr1	HWRFGGKWWPCGKADTNVGNRVYTHPDSNTGAHWMRQEISFGKLLTNNKGASNNNG	120
XtTbr1	HWRFGGKWWPCGKADTNVGNRVYTHPDSNTGAHWMRQEISFGKLLTNNKGASNNNG	120
PpTbr	SWKFQSGKWWATGKSDGVPRATGIYKHPDPSNTGEHWMRQDIAFSKLLTNNRKGKDSG--	118
PmTbr	SWKFQSGKWWATGKSDGVPRATGIYKHPDPSNTGEHWMRQDIAFSKLLTNNRKGKDSG--	118
PlTbr	QWKFQCGKWI PCGQAENI PKVSNTYI LHPDPS NGLHWMHQDIVFSKLLTNNHRGKDNG--	118
LvTbr	QWKFQCGKWI PCGQAENI PKVSNTYI LHPDPS NGLHWMHQDIVFSKLLTNNHRGKDNG--	118
SpTbr	QWKFQCGKWI PCGQAENI PKVSNTYI LHPDPS NGLHWMHQDIVFSKLLTNNHRGKDNG--	118
HpTbr	QWKFQCGKWI PCGQAENI PKVSNTYI LHPDPS NGLHWMHQDIVFSKLLTNNHRGKDNG--	118
SmTbr	QWKFQCGKWI PCGQAENI PKVSNTYI LHPDPS NGLHWMHQDIVFSKLLTNNHRGKDNG--	118
PjTbr	QWKFQCGKWI PCGQAENI PKVSNTYI LHPDPS NGLHWMHQDIVFSKLLTNNHRGKDNG--	118
BfTbr	HWKFGNGKWWPCGRADTNPQGSRVYVHPESNSGAHWMKQEVVFSKLLTNNKGADNG--	118
PfTbr	HWKFGNGKWWPCGQAEHVHPGSIYIHPDPSNTGNHWMKQEVVFSKLLTNNKGADNG--	118
SkTbr	HWKFGNGKWWPCGQAEHVHPGSIYIHPDPSNTGNHWMKQEVVFSKLLTNNKGADNG--	118

::: . * * : : . : * * * * * : * * : * :

SpTbx4	-----NQIVLHSMHEYTPRLFIERLISNKSPHGDIKVEKDSMSEPSPTSNSMST	166
PmTbx4	-----TNTVLHSMHRYTPRIITQLRCCSSANVRGIAA-----	148
SkTbx4	-----THTVLHSMHKYTPVIKIKHGSRT-----	141

DrTbx2a	-----FTILNSMHKYQPRFHIVRANDILKLP-----	143
XtTbx2	-----FTILNSMHKYQPRFHIVRANDILKLP-----	141
MmTbx2	-----FTILNSMHKYQPRFHIVRANDILKLP-----	143
SpTbx2/3	-----QTILNSMHKYQPRFHIVKANDILKLP-----	144
PmTbx2/3	SSPVSCYFGFLNWTLWNAMHKYQPRFHIVKANDILKLP-----	157
MmBra	-----QIMLNSLHKYEPRIHIVRVGGPQ-----	139
XlBra	-----QIMLNSLHKYEPRIHIVRVGGTQ-----	139
SpBra	-----QVMLNSLHKYEPRIHIVRVGGREK-----	140
PmBra	-----QIMLNSLHKYEPRIHIVRVGGPEK-----	140
MmTbx21	Q-----MIVLQSLHKYQPRLHIVEVNDGEPEA-----	147
XtTbx21	Q-----MVVLQSLHKYQPRFHVTRVEDPGGPE-----	147
DrTbx21	Q-----MIVLQSLHKYQPRLHIVEVKEDGTED-----	147
MmEomes (Tbr2)	Q-----MIVLQSLHKYQPRLHIVEVTEDEGVED-----	147
XtEomes	Q-----MIVLQSLHKYQPRLHIVEVSEDEGVED-----	147
DrEomes	Q-----MIVLQSLHKYQPRLHIVEVTEDEGVED-----	147
DrTbr1	Q-----MIVLQSLHKYQPRHVHVEISKNEDED-----	147
MmTbr1	Q-----MVVLQSLHKYQPRLHVVEVNEDEGTED-----	147
XtTbr1	Q-----MVVLQSLHKYQPRLHVVEVNEDEGTED-----	147
PpTbr	-----YLVINSMHIYQPRIHVLDLTG-----	139
PmTbr	-----YLMINSMHIYQPRIHVLDLTG-----	139
PlTbr	-----FVVLNSMHKYQPRIHVLELGE-----	139
LvTbr	-----FVILNSMHKYQPRIHVVELSE-----	139
SpTbr	-----FVILNSMHQYQPRIHVLELSE-----	139
HpTbr	-----FVILNSMHQYQPRIHVLELTE-----	139
SmTbr	-----FVILNSMHQYQPRIHVLELND-----	139
PjTbr	-----FVILNSMHKYQPRIHVLELND-----	139
BfTbr	-----HVVLNSMHKYQPRLHIEVSNRAGGG-----	144
PfTbr	-----HIVLNSMHKYQPRIHVIEVSPNRPPD-----	144
SkTbr	-----HIVLNSMHKYQPRIHVIEVSPNRPPD-----	144
	: .::* * * . :	

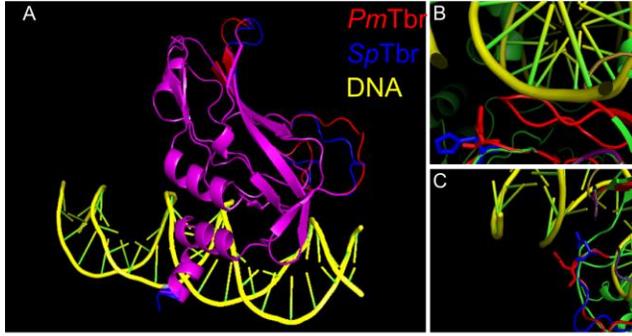
SpTbx4	SSSPSTWMMHQDHRSESPSSDAASTSTGLRHSGGSTSSTMTSSSPDPRVDQMTSISFS	226
PmTbx4	-----PCSDPNLPTEGGVPHYRQ-----LPAAVGGSRPSCFEFE	182
SkTbx4	-----DNGTGMLQFSFQ-----	153
DrTbx2a	-----YSTFRTYVFP-----	153
XtTbx2	-----YSTFRTYVFP-----	151
MmTbx2	-----YSTFRTYVFP-----	153
SpTbx2/3	-----WSQFRTFVFP-----	154
PmTbx2/3	-----WSHFRTFVFP-----	167
MmBra	-----RMITSHCFP-----	148
XlBra	-----RMITSHSFP-----	148
SpBra	-----QRLVGSYSFT-----	150
PmBra	-----QRLIRSFSP-----	150
MmTbx21	-----ACASANTHVFTFQ-----	160
XtTbx21	-----SQSHSFIFP-----	156
DrTbx21	-----PFLTSKTQTFVFP-----	160
MmEomes (Tbr2)	-----LNEPSKTQTFTFP-----	160
XtEomes	-----LNDSAKSQTFTFP-----	160
DrEomes	-----MSSEAKTQTFTFP-----	160
DrTbr1	-----TSDPDGVQTFTFP-----	160
MmTbr1	-----TSQPGRVQTFTFP-----	160
XtTbr1	-----TSQPGRVQTFTFP-----	160
PpTbr	-----ARVLQTHSFP-----	149
PmTbr	-----ARVLQTHSFP-----	149
PlTbr	-----SRSLQTHSFP-----	149
LvTbr	-----SRYIQTHSFP-----	149
SpTbr	-----SRSIQTHSFP-----	149
HpTbr	-----SRSIQTHSFP-----	149
SmTbr	-----RRSLRTFNFP-----	149
PjTbr	-----RRSLQTHSFP-----	149
BfTbr	-----ERVLQSHSFP-----	154
PfTbr	-----QRTLQTHSFP-----	154
SkTbr	-----QRTLQTHSFP-----	154
	*	

SpTbx4	ETSFVAVTAYQNDHITQLKIQNNPFAKAFRDAEVA	261
PmTbx4	ETAFAVAVTAYHSEQITQLKIQNNPFAKAFRDADIA	218
SkTbx4	QTSFIAVTAYQNEHVTQLKIQNNPFAKAFRDADVA	187
DrTbx2a	ETDFIAVTAYQNDKITQLKIDNNPFAKGFRTGNG	186
XtTbx2	ETDFIAVTAYQNDKITQLKIDNNPFAKGFRTGNG	184
MmTbx2	ETDFIAVTAYQNDKITQLKIDNNPFAKGFRTGNG	186
SpTbx2/3	ETVFIGVTAYQNEKITQLKIDYNPFAKGFRTGAG	188
PmTbx2/3	ETDFIAVTAYQNEKVTQLKIDNNPFAKGFRTQGTG	201
MmBra	ETQFIAVTAYQNEEITALKIKYNPFAKAFDLDAKER	183
XlBra	ETQFIAVTAYQNEEITALKIKHNPFAKAFDLDAKER	183
SpBra	ETRFIAVTAYQNEDITQLKIKYNPFAKAFDLIDKDK	185
PmBra	ETQFIAVTAYQNEDITQLKIKYNPFAKAFDLIDKEK	185
MmTbx21	ETQFIAVTAYQNAEITQLKIDNNPFAKGFRENFES	195
XtTbx21	ETQFIAVTAYQNADITQLKIDHNPFAKGFRTDHCCL	191
DrTbx21	ETQFIAVTAYQNADITQLKIDHNPFAKGFRTDNYDT	195
MmEomes (Tbr2)	ETQFIAVTAYQNTDITQLKIDHNPFAKGFRTDNYDS	195
XtEomes	ETQFIAVTAYQNTDITQLKIDHNPFAKGFRTDNYDS	195
DrEomes	ENQFIAVTAYQNTDITQLKIDHNPFAKGFRTDNYDS	195
DrTbr1	ETQFISVTAYQNTDITQLKIDHNPFAKGFRTDNYDT	195
MmTbr1	ETQFIAVTAYQNTDITQLKIDHNPFAKGFRTDNYDT	195
XtTbr1	ETQFIAVTAYQNTDITQLKIDHNPFAKGFRTDNYDT	195
PpTbr	ETQFIVGTAYQNTDITQLKIDHNPFAKGFRTDNYDS	184
PmTbr	ETQFIVGTAYQNTDITQLKIDHNPFAKGFRTDNYDS	184
PlTbr	ETRFFVGTAYQNTDVTQLKIDYNPFAKGFRTDNYDN	184
LvTbr	ETQFFVGTAYQNTDVTQLKIDYNPFAKGFRTDNYDN	184
SpTbr	ETQFFVGTAYQNTDVTQLKIDYNPFAKGFRTDNYDN	184
HpTbr	ETQFFAVTAYQNTDVTQLKIDYNPFAKGFRTDNYDN	184
SmTbr	ETQFFAVTAYQNTDVTQLKIDYNPFAKGFRTDNYDN	184
PjTbr	ETQFFAVTAYQNTDVTQLKIDYNPFAKGFRTDNFDN	184
BfTbr	ETQFIAVTAYQNTDITQLKIDYNPFAKGFRTDNYDG	189
PfTbr	ETQFFAVTAYQNTDITQLKIDHNPFAKGFRTDNYDS	189
SkTbr	ETQFFAVTAYQNTDITQLKIDHNPFAKGFRTDNYDC	189

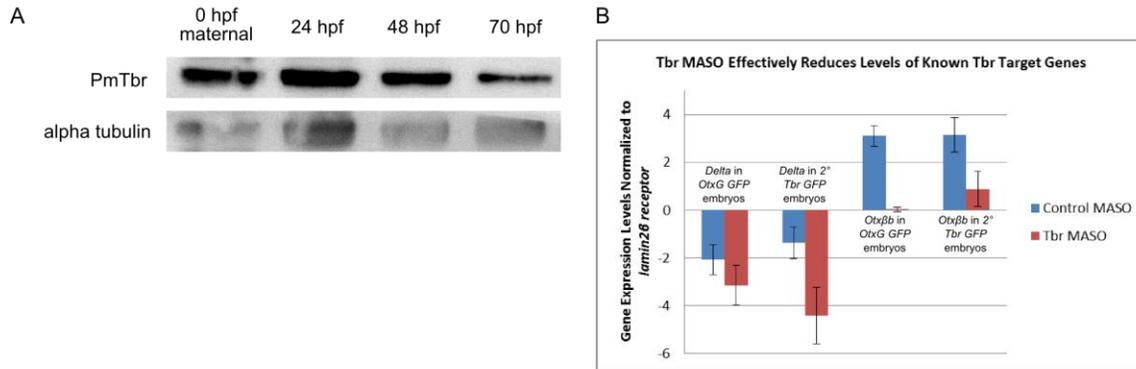
:. *..****:.. ::* ***. *****.* :

Supplemental Figure 2.1: T-box Alignment used in Phylogenetic Tree Construction.

Alignment of DBDs from T-boxes of *Tbr*, *Bra*, and other T-box paralogs from a variety of deuterostome species. This alignment was used to construct the tree in Figure 2.1. All accession numbers are listed next to the gene name. Species abbreviations are as follows: *Bf*- *Branchiostoma floridae*, *Dr*-*Danio rerio*, *Hp*- *Hemicentrotus pulcherrimus*, *Lv*- *Lytechinus variegatus*, *Mm*-*Mus musculus*, *Pf*-*Ptychodera flava*, *Pj*- *Peronella japonica*, *Pl*- *Paracentrotus lividus*, *Pm*- *Patiria miniata*, *Pp*- *Patiria pectinifera*, *Sk*-*Saccoglossus kowalevskii*, *Sm*- *Scaphechinus mirabilis*, *Sp*-*Strongylocentrotus purpuratus*, *Xl*-*Xenopus laevis*, *Xt*- *Xenopus tropicalis*



Supplemental Figure 2.2: Tbr Structural Prediction. A. Structure of sea star Tbr (red) and sea urchin Tbr's (blue) T-box domains modeled based on the structure of *X/Bra* (PDB ID 1XBR) (Muller and Herman 1997) using Phyre (Kelley and Sternberg 2009). B. View of sea star asparagine 389 (red) vs. sea urchin histidine (blue). The sea urchin amino acid is predicted to be poorly positioned to make a hydrogen bond with the DNA backbone. C. The adjacent amino acids (388/478) are also affected by this difference even though both orthologs have an asparagine in this position.



Supplemental Figure 2.3: A. Developmental Western Blot of *PmTbr*. Tbr levels are high maternally (0 hours post-fertilization (h)) and remain high through the gastrula stage (48h). Levels are reduced at 70 h. Alpha-tubulin levels are shown as a loading control. B. Difference in Cycle Number Threshold (Cts) of known Tbr target genes, *Pmdelta* (Accession: ACC62396.1) and *PmotxBb* (Accession: AY263968), normalized to *Pmlamin2B receptor* (a nuclear envelope protein) (Accession: KJ868807). Levels are compared between control and Tbr morpholino oligonucleotide injected siblings. These same samples were used to generate the *GFP* comparison in Figure 2.4E. Data points represent the average of three experiments. Error bars indicate the Standard Error of the Mean.

Amplicon	Method	Forward Primer	Reverse Primer
GFP	qPCR	GCTGGCCGACCATTATCAACA	TGATCCCAGCAGCGGTCA
mCherry	qPCR	CCTCCTCCGAGCGGATGTAC	CGGGCTTCTTGGCCTTGTAG
<i>PmLamin2B receptor</i>	qPCR	GAGCATGCCTAAGCCAGACC	CTCCACCATTGGGCTCCAGTA
<i>PmDelta</i>	qPCR	GTCAGGGTTCCTACTGGCATGT	CAGACATTGGTGGCCATCTT
<i>PmOtxβb</i>	qPCR	GAAAGGATGGATTGCGTCAT	ACCACTCATACTGCGGATTG
PmOtxG CRM set 1	ChIP RT-PCR	GCATAACCCCTGCTTCTGTTGCA	CCTGGCCTCCCATGTCCAATTG
PmOtxG CRM set 2	ChIP RT-PCR	CATGTGCAGCTTATCACTCGTCTG	AGAGGTGGTATTGGGATCTTGTCTG
PmOtxG 1kb upstream	ChIP RT-PCR	CGTGGCAGGCTCAGCAAGTG	GATGATGGCCCAAGCAATGTCATG
PmOtxG 1kb downstream	ChIP RT-PCR	CTCAAATGGCGGGACGGTTACC	CTGCATGGTATGGCAGTGAAT
2° Tbr GFP	Construct generation	GCCAGGGATGTCACTCCTTGTCTGC	GCGAGCAAGGAGGTGACATCCCTGGC
Tbr Deletion	Construct generation	GCCAGGGATACCCACCCCTTGTCTGC	GCGAGCAAGGGTGGGTATCCCTGGC

Supplemental Table 2.2: Primer Sequences

Chapter 3: The GRN for specification of anterior dorsal ganglia in *Patiria miniata* sea star embryos.

3.1 Preface

The following work has been prepared with the intent of submitting it to the journal *Development* in the near future. Although this project seem disconnected from the work described in Chapter 2, I became interested in it when I realized differences in Tbr secondary site binding were mostly likely to take effect in the evolution of the ectoderm among echinoderms, which will be discussed in greater detail in Chapter 4. However, the sea star ectoderm has not been the subject of extensive study; only two papers have really looked extensively at regulatory gene expression and function. Many of the sea urchin studies involving the ectoderm have focused on the apical organ which has not been well-studied at all in the sea star, making evolutionary comparison of this region difficult.

A former graduate student, Kristen Yankura, is a co-first author of this work. Kristen began aspects of this project during the end of her thesis work, and as a result left behind a lot of good data about apical ectoderm patterning, but without a definite story. Based on some of my initial observations as I began to replicate Kristen's experiments, I decided to take the approach of understanding the process of neurogenesis from stem cell to mature neuron. Therefore, the experiments I added focus on neural progenitors and proliferation. Specifically, Kristen performed the experiments depicted in Figures 3.1 E, I and J, 3.5 A and B, 3.6 C, D, G, H, and M-P, and finally Figure 3.7 C and D. Several undergraduates assisted with *in situ* experiments both during Kristen's work on this project and my own: Sowyma Yennam, Lazar Lalone, and Nikita Mishra. This was very helpful in performing such a large number of experiments quickly, and was also a useful learning experience for Kristen and myself.

3.2 Abstract

The anterior-posterior and dorsal-ventral patterning of neurogenic ectoderm has been shown to be well conserved across deuterostomes. It has been hypothesized that the apical organ, a larval sensory structure found in invertebrate deuterostomes, may represent the ancestral state from which the vertebrate forebrain evolved. As such, a greater understanding of how

different subsets of neurons are patterned and specified in invertebrate deuterostomes will inform our understanding of homologous processes in vertebrates and the origins of the central nervous system. We have previously described a gene network that specifies and patterns ciliary band neurons from a pan-neurogenic ectoderm. Here, we describe a unique network that allows cells from the same initial neurogenic ectoderm to instead become the serotonergic neurons of the anterior dorsal ganglia, the apical sensory organ of the sea star larvae. We have found the specification of these serotonergic neurons is dependent on the *lim* homeobox transcription factor, *lhx2/9*, an ortholog of vertebrate *lhx2* and *lhx9*, which are required for the early specification of the forebrain and development of neurons in this structure. *Pmlhx2/9* expression requires pan-neurogenic specification transcription factors, such as *soxc*, dorsal identity provided by *bmp2/4*, and an apical plate territory, defined by *foxq2*. However, serotonergic neurons are specified normally when ciliary band formation is prevented by loss of *foxg*. *Lhx2/9* cells in the *foxq2* territory differentiate into neurons, while those remaining serve as a reserve of proliferating neurogenic cells. Loss of *six3* obliterates this reserve of *lhx2/9* cells by allowing Wnt signaling to expand into this territory, resulting in far fewer serotonergic neurons in larval stages. As is true for the ciliary band neurons, specification and patterning mechanisms are decoupled and therefore can evolve independently, accounting for the diversity of apical organ structure seen among ambulacrarians. More complex patterning lead to the highly specialized regions of the vertebrate forebrain.

3.3 Introduction

An apical sensory organ, located in the most anterior region of the larvae, has been observed in many marine invertebrates, including both deuterostomes like echinoderms (Bisgrove and Burke 1986; Chee and Byrne 1999; Nakajima, Kaneko, et al. 2004; Byrne et al. 2006), and hemichordates (Nakajima, Humphreys, et al. 2004; Nielsen and Hay-Schmidt 2007), and also protostomes such as mollusks (Kempf et al. 1997; Croll 2006), and annelids (Marlow et al. 2014). These structures, sometimes also called anterior ganglia or apical ganglia, can be morphologically diverse, particularly among ambulacrarians, the clade that encompasses hemichordates and echinoderms (reviewed in Byrne et al. 2007). Apical organs are thought to be sensory structures owing to the presence of serotonergic neurons. Serotonin, or 5-hydroxytryptamine, is a neurotransmitter found across metazoans and is important for a variety

of neurophysiological and developmental processes. Serotonergic neurons are found in all metazoans with nervous systems, although it is unknown whether they are specified via homologous mechanisms in diverse taxa. In vertebrates, serotonergic neurons are found in the basal forebrain, spinal cord, and raphe nuclei of the hindbrain, with the exception of placental mammals which only have the raphe nuclei population (Lillesaar 2011). Most studies about serotonergic neuron specification have focused on the raphe neurons, and much less is known about their specification in other regions.

It has been hypothesized that the complex central nervous system (CNS) of vertebrates evolved from an apical organ (Lacalli 1994; Hay-Schmidt 2000) similar to those found in basal deuterostomes, although this has been debated. Others argue that because the CNS is found in disparate taxa, including both arthropods and vertebrates, that the ancestors of all bilaterians also had a CNS (reviewed in Holland et al. 2013). A more recent hypothesis is that vertebrate pineal gland, retina, and anterior hypothalamus all originated from an ancestral apical plate, particularly due to shared top-tier transcriptional regulators like *six3* and *rx* among these structures in vertebrates and the apical plates of invertebrate deuterostomes (Tosches and Arendt 2013). Much work supports the idea that the early patterning of the ectoderm exhibits many similarities across taxa (reviewed in Range 2014), although it's not clear to what extent these early molecular homologies in the apical plate correspond to conserved CNS developmental mechanisms. An understanding of the development of this simple sensory structure at the gene regulatory level will inform studies of the vertebrate forebrain and its origins.

Here we describe the gene regulatory network (GRN) governing the early specification of serotonergic neurons in the apical organ, referred to here as the anterior dorsal ganglia, or simply dorsal ganglia, of the larvae of the sea star, *Patiria miniata*. Both these neurons and the neurons of the larval ciliary bands originate from a pan-neurogenic ectoderm, but overlaying this broad region with different patterning mechanisms allows these populations of neurons to diverge into different neuronal cell types. Because these GRNs are largely uncoupled, different patterns of ciliary bands and apical organ structures can evolve readily, explaining the great diversity seen in these structures. Additionally, this observation may explain why other ambulacrarians, especially directly developing hemichordates, retain a broadly neurogenic ectoderm, which has baffled those seeking to understand the origins of the vertebrate CNS. We find that the apical organ GRN is overlaid on a broadly neurogenic ectoderm, and therefore evolutionary changes in

patterning can explain the diversity seen among apical organ morphologies and the retention of diffuse neural nets in some taxa. Furthermore, there is a large amount of overlap between the apical organ neuron GRN and known vertebrate forebrain patterning and neuronal specification mechanisms. Interestingly, the GRN components used in the sea star apical organ tend to be genes used for similar functions in multiple forebrain sub-regions, especially the retina and hypothalamus. This suggests that sea star's GRN could represent an ancestral mechanism for developing anterior sensory structures that duplicated and diverged as the forebrain evolved, allowing for the specialized regions present in modern vertebrates.

3.4 Results

3.4.1 *PmLhx2/9* is required for the specification of serotonergic neurons

We previously characterized the expression of *elav*, a marker of all differentiated neurons in the sea star embryo. Two groups of neurons were observed; one associated with the pre and post-oral ciliary bands and one located in bilateral anterior dorsal clusters (Yankura et al. 2013). Based on the position of the latter group, we hypothesized that these are the serotonergic neurons of the *P. miniata* larvae since their location is reminiscent of anterior dorsal ganglia serotonergic neurons of related sea stars, *Patiriella regularis* and *Patiria pectinifera* (Chee and Byrne 1999; Nakajima, Kaneko, et al. 2004). We sought to understand how this group of neurons is specified during early development.

In previous work, we observed that a *lim domain homeobox* (*Lhx2*) is also expressed in the anterior dorsal ectoderm of gastrula stage embryos (Yankura et al. 2010). Phylogenetic analysis confirms that we have isolated the echinoderm ortholog of vertebrate *lhx2* (Supplemental Figure 3.1). This gene was duplicated in the vertebrate lineage, after the divergence of echinoderms and chordates, and therefore echinoderm *lhx2* is also orthologous to vertebrate *lhx9*. Thus, we now refer to this gene as *Pmlhx2/9*. We speculated that this transcription factor could also be key to formation of the sea star anterior dorsal ganglia due to the known roles of both *lhx2* and *lhx9* in vertebrate forebrain development (Porter et al. 1997; Bertuzzi et al. 1999; Rétaux et al. 1999). *Lhx2* plays roles in both specifying regional identity and balancing precursor vs. neuron production (Porter et al. 1997; Gordon et al. 2013). As *lhx2* and *lhx9* share some overlapping expression and are redundant in some contexts (Bertuzzi et al.

1999; Rétaux et al. 1999), it is likely that the single *lhx2/9* echinoderm ortholog will have functionality that is similar to both of these. We next characterized the expression of *lhx2/9* more thoroughly. As previously described, *Pmlhx2/9* is expressed as spots in the anterior dorsal ectoderm at 48 hours post-fertilization (hpf) (Figure 3.1A and B) (Yankura et al. 2010). We do not detect expression of *lhx2/9* at earlier stages, such as hatched blastula and early gastrula. By 96hpf, *lhx2/9* is expressed in bilateral clusters in the anterior ectoderm (Figure 3.1C and D), in a very similar position as the previously described dorsal ganglia *elav* expression (Figure 3.1E).

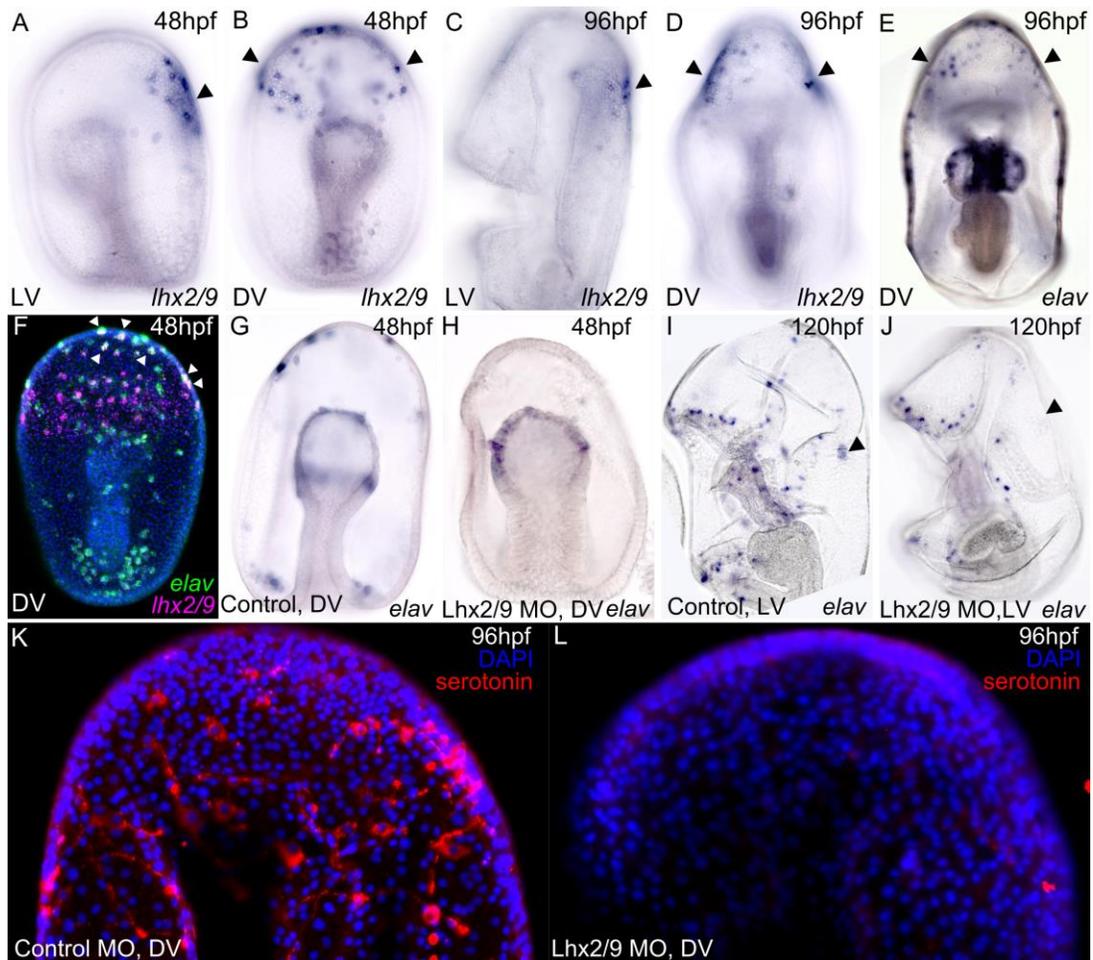


Figure 3.1: Lhx2/9 is required for the specification of serotonergic neurons. A-D.

Expression of *Pmlhx2/9* at 48 hours post fertilization (hpf) and 96hpf respectively occurs in bilateral clusters in the anterior dorsal ectoderm. E. At 96 hpf, *elav*, a marker of differentiating neurons, is also expressed in two anterior dorsal ganglia. F. *Lhx2/9* is co-expressed with *elav* in the apical plate at 48 hpf (arrow heads). Additional *Lhx2/9* cells are present more posterior to

this. When *Lhx2/9* is knocked-down by *Lhx2/9* specific morpholino oligonucleotide (MO), *elav* expression is lost from the apical plate at 48hpf (G vs. H) and the dorsal ganglia (I vs. J) 96hpf, but mesoderm expression (G vs. H) and ciliary band neurons (I vs. J) are unaffected. K. Serotonergic neurons of the dorsal ganglia stained with rabbit anti-serotonin also occur in bilateral clusters in the anterior dorsal ectoderm. L. *Lhx2* MO larvae do not develop serotonergic neurons. In all panels, DV indicates a dorsal view, LV, indicates a lateral view.

Using two-color fluorescent *in situ* hybridization (FISH), we see that at 48hpf, *lhx2/9* and *elav* co-localize in the apical-most ectoderm, although there are additional *lhx2/9*-positive cells located more posteriorly that do not express *elav* (Figure 3.1F). This indicates that at least some of the *lhx2/9* cells will become neurons. Furthermore, when *Lhx2/9* is knocked-down using a morpholino antisense oligonucleotide (MO), *elav* is no longer expressed in the anterior pole of 48hpf embryos (Figure 3.1G vs. H) or in the dorsal ganglia of 120hpf embryos (Figure 3.1I vs. J). Other domains of *elav* expression, namely the mesodermal bulb of 48hpf embryos and ciliary band neurons of 120hpf embryos, are unaffected by the loss of *Lhx2/9*.

We next sought to understand whether the neurons that arise from *lhx2/9* cells will become the serotonergic neurons of the dorsal ganglia. In control 96hpf larvae, serotonin is present in two groups of neurons, localized to the anterior dorsal ectoderm, just as it is in other sea star species (Figure 3.1K) (Chee and Byrne 1999; Nakajima, Kaneko, et al. 2004). *Lhx2/9* knock-down embryos do not express serotonin (Figure 3.1L). Collectively, these experiments demonstrate that *lhx2/9* is required to make the dorsal ganglia, composed of serotonergic neurons. This suggests *lhx2/9* has a conserved role across deuterostomes in specifying anterior neurons.

3.4.2 *Lhx2/9* cells are neural precursors that proliferate by symmetric divisions and generate neurons through asymmetric divisions

Vertebrate *lhx2* and *lhx9* are needed for proliferation of neural precursors in addition to their roles in the early patterning of the forebrain (Hägglund et al. 2011; Chou and O’Leary 2013; Gordon et al. 2013). We wanted to understand whether the function of this gene is conserved in addition to its relative expression domain, and ultimately to what extent the process of neurogenesis is conserved in the anterior dorsal ganglia. To this end, we first determined

whether *lhx2/9* cells are post-mitotic cells or proliferating cells by combining EdU labeling with FISH. Our EdU pulse was very short in order to capture only cells that are actively proliferating. We find that some *lhx2/9* cells are EdU⁺ (Figure 3.2A-A’’’).

Further evidence of this is that we observe that *lhx2/9* cells typically occur in pairs (Figure 3.2B-B’). We also see that in the apical-most ectoderm, *lhx2/9* cells undergo asymmetric divisions, in which one cell expresses *elav* and the other does not (Figure 3.2B-B’’). *Elav* marks post-mitotic neurons, which are therefore on track to differentiate rather than produce additional neural precursors. We hypothesize that the other cell in the pair, which does not express *elav*, will continue to proliferate. We also see pairs of *lhx2/9* cells in which neither cell expresses *elav*, which probably represent cells dedicated to proliferation at that time.

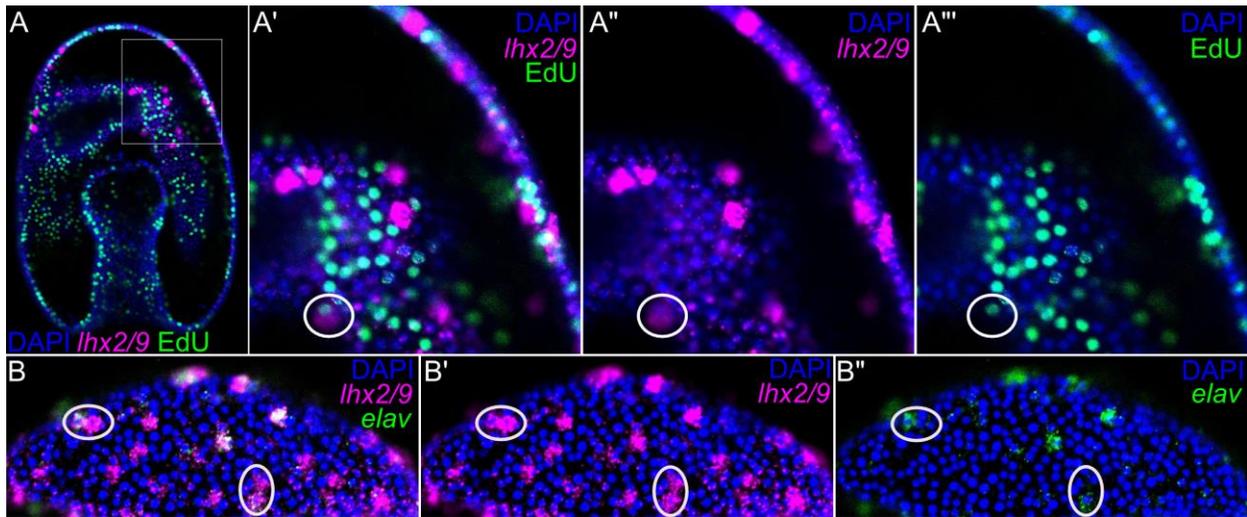


Figure 3.2: Lhx2/9 cells are proliferating neural progenitors. A. Regions of *lhx2/9* expression and EdU labeling in a 48hpf embryo. A’-A’’’. Inset of the indicated region in A. An example of co-expression of *lhx2/9* and EdU label is indicated by a white circle. B-B’’. *Lhx2/9* cells generate *elav*-positive neurons by asymmetric divisions (ex. circled cell pairs). In these pairs, both cells express *lhx2/9*, but only one also expresses *elav*. *Lhx2/9* expression persists in post-mitotic neurons.

3.4.3 *Lhx2/9* cells originate from *soxc*-expressing neural stem cells

Our previous work had also demonstrated that the expression of *elav* in the apical pole domain is dependent on the transcription factor *soxc*, which is expressed broadly throughout the

ectoderm at 48hpf (Yankura et al. 2013). We speculated that it might operate at the top of the GRN because it is present maternally and therefore earlier than other genes involved in neurogenesis. Here we show that *soxc* is required early in neurogenesis, as it is needed for the expression of *lhx2/9*. FISH reveals that *soxc* is expressed in pairs of cells throughout the ectoderm at 48hpf (Figure 3.3A-A’’). *Lhx2/9* is expressed in one of the two *soxc* cells, but only in the dorsal anterior ectoderm (Figure 3. 3A and A’). *Soxc* expression appears to be stronger in the cells that do not also express *lhx2/9* (Figure 3.3A’’), suggesting that these cells are undergoing a transition in their identity. We also see that *soxc* is required for *lhx2/9* expression. *Soxc* MO embryos do not express *lhx2/9* (Figure 3.3 B vs. C).

We now know from this study and our previous work that *soxc* is required to make both known neural cell-types in the sea star larvae (Yankura et al. 2013). Because *soxc* is also expressed in pairs reminiscent of recently divided cells, we wondered whether these could be neural stem cells. To determine this, we also performed EdU/FISH. We find that *soxc* cells are actively proliferating, because they are double-labeled with EdU (Figure 3.3D-D’’). Therefore, we concluded that the *soxc* cells are neural stem cells, in which one daughter cell maintains a stem cell state and the other progresses towards a particular neural fate. Daughter cells that express *lhx2/9* may now be committed to serotonergic vs. ciliary band neuron fate.

In the vertebrate forebrain, *sox11*, a *soxc* ortholog, facilitates the transition from neural stem cell to differentiating post-mitotic neuron by turning on genes such as *lhx2* (Bergsland et al. 2011) and is needed for proper proliferation of neural progenitor cells (Wang et al. 2013). Thus, both ciliary band and dorsal ganglia neurons originate from a pan-neurogenic ectoderm full of *soxc* neural progenitors, but subsequent expression of *lhx2/9* marks commitment towards dorsal ganglia fate. *Soxc*’s role in maintaining cell proliferation is most likely ancient, as its expression has recently been observed in regions of high cell division in ctenophores (Schnitzler et al. 2014). Evolution of ciliary band and dorsal ganglia patterning mechanisms overlaid on this common pan-neurogenic ectoderm could then account for the diverse neuronal configurations seen across metazoans.

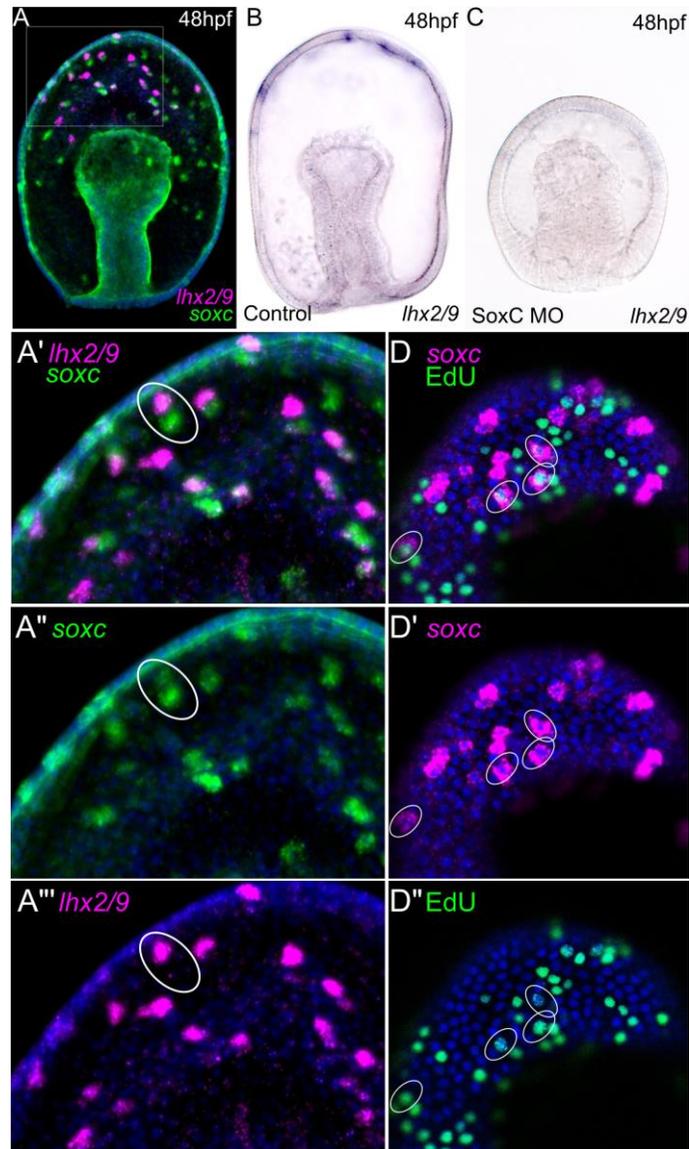


Figure 3.3: *Lhx2/9* cells originate from *soxc*-expressing neural stem cells. A-A'''. Pairs of *soxc* cells are found throughout the pan-neurogenic ectoderm at 48hpf. In some pairs, one cell will also express *lhx2/9*. An example of this is indicated by a white circle in A'-A'''. *Soxc* is required for the expression of *lhx2/9*. *Soxc* MO embryos do not express *lhx2/9* (B vs. C). D-D''. *Soxc* cells actively proliferate. Circles indicate EdU⁺ cells that also express *soxc*.

3.4.4 The Dorsal Ganglia Does Not Require the Ciliary Band Specification GRN

Understanding how different lineages of neurons can differentiate from a common pool of neural progenitors is of broad interest. Previous work that characterized the GRN for ciliary band development revealed that ciliary band neurons and dorsal ganglia neurons are specified by different processes, although they most likely both arise from a pan-neurogenic ectoderm (Yankura et al. 2013). In that work, we demonstrated that *elav* expression in the dorsal ganglia persists when ciliary band specification and patterning are disrupted by the knock-down of Foxg and Ephrin Receptor. Additionally, we show in Figure 3. 1I and J that when *Lhx2/9* is knocked-down, only dorsal ganglia *elav* expression is lost; ciliary band expression is unperturbed. Collectively, these experiments suggest that the ciliary band and dorsal ganglia GRNs are independent aside from their common *soxc* stem cell origin.

Here, we further confirmed that dorsal ganglia specification occurs normally when ciliary band-specific transcription factors are perturbed. First, we show that Foxg is not required for the early expression of *lhx2/9* (Figure 3.4A vs. B). To verify that our Foxg knock-down is effective, and that ciliary band formation is abrogated, we performed two-color FISH with both *lhx2/9* and *onecut*. We previously demonstrated that *onecut*, which is normally restricted to the ciliary bands, is dramatically upregulated in Foxg morphant larvae (Yankura et al. 2013). Here, we show that 96hpf Foxg morphants not only express *lhx2/9* normally, but they do so even though *onecut* is inappropriately expressed in the *lhx2/9* territory (Figure 3.4C vs. D). Therefore, the two populations of neurons found in the early sea star larvae are specified via different GRNs.

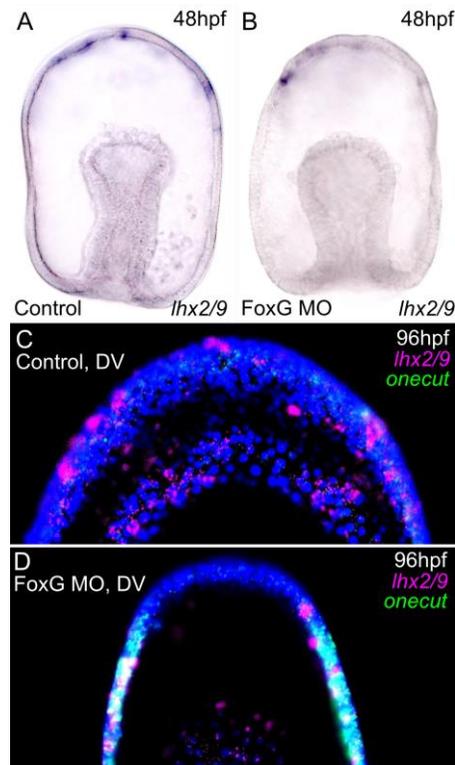


Figure 3.4: Lhx2/9 expression is not dependent on components specific to the ciliary band neuron gene regulatory network. A. Normal expression of *lhx2/9* at 48hpf. B. When FoxG, which is crucial for specifying ciliary band neurons, is knocked-down, *lhx2/9* is still expressed normally. C. At 96hpf, *lhx2/9*, but not *onecut* is normally expressed in the anterior dorsal ectoderm. D. When Foxg is knocked-down, *lhx2/9* expression is normal, even though *onecut* is inappropriately upregulated in this same region.

3.4.5 Dorsal/Ventral patterning is needed to specify the Dorsal Ganglia

We wanted to understand the regulatory environment that causes some *soxc* progenitors to turn on *lhx2/9* and eventually differentiate into serotonergic neurons. This environment must include a dorsal-ventral patterning component in order to explain the dorsally-restricted expression pattern of *lhx2/9*. In many developmental systems, including other model echinoderms such as the sea urchin, Nodal and BMP signaling are critical to this process (Saudemont et al. 2010). Additionally, BMP2 and BMP4 have previously been implicated in regulating *lhx2* expression in the mouse forebrain (Monuki et al. 2001). We previously characterized the expression of *nodal*, *bmp2/4*, and BMP2/4's effector p-Smad1/5/8 at blastula

stages (24hpf) (Yankura et al. 2013). In this work, we found that *nodal* is required for the ventral expression of ciliary band transcription factor *foxg*, while *bmp2/4* signaling prevents expansion of *foxg* into the dorsal ectoderm. P-Smad1/5/8 is present throughout the dorsal ectoderm at 24hpf, indicating that BMP2/4 signaling is active there. We hypothesized that this same D/V patterning mechanism may localize *lhx2/9* to the dorsal ectoderm.

To test this, we injected BMP2/4 MO and assayed for *lhx2/9* expression at 48hpf. Without BMP2/4 signaling, *lhx2/9* is not expressed (Figure 3.5A vs. B). Likewise, these embryos do not produce serotonergic neurons in later development (Figure 3.5C vs. D). BMP2/4 MO embryos are radialized and never develop normal bipinnaria larval morphology. To account for the absence of serotonin arising from a developmental delay, we compared 96hpf BMP2/4 MO embryos to 72hpf control MO injected embryos. Even at this earlier stage, the control embryos have serotonergic neurons in the anterior dorsal embryo, suggesting that the lack of such neurons in the Bmp2/4 MO is due to a BMP signaling requirement. Importantly, *soxc* is expressed normally in 48hpf BMP2/4 MO (Figure 3.5E vs. F), even though these same embryos cannot produce *lhx2/9* positive cells. Therefore, BMP2/4 signaling is not involved in creating the pan-ectodermal neuron precursors, but instead sculpts a territory in which some of these precursors will be directed towards becoming serotonergic neurons.

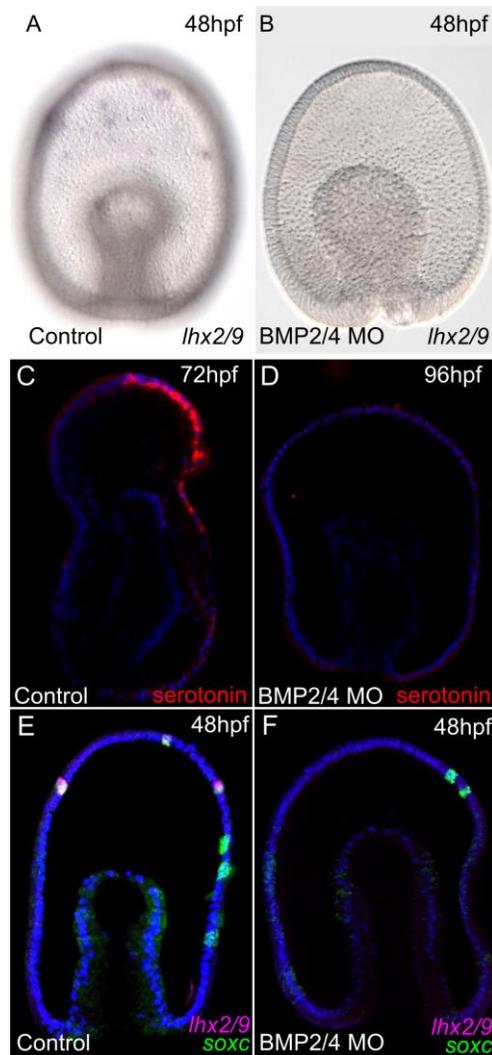


Figure 3.5: BMP2/4 Signaling is needed for *lhx2/9* expression and provides dorsal regional identity. A. Expression of *lhx2* on the dorsal surface of a 48hpf control embryo. B. *Lhx2/9* expression is lost when BMP2/4 is knocked-down. C-D. Serotonergic neurons do not develop in BMP2/4 MO embryos (D), although this is not due to a developmental delay as 72hpf control embryos already have serotonergic neurons (C). E-F. BMP2/4 knock-down affects *lhx2/9*'s expression, but not *soxc*'s. In BMP2/4 MO embryos, *soxc* is expressed in the anterior ectoderm, but none of these cells do also express *lhx2/9*, as seen in control embryos. This indicates that BMP2/4 provides patterning to neural stem cells as opposed to promoting or prohibiting their formation.

3.4.6 Dorsal ganglia require an apical pole domain, defined by *foxq2* and a lack of Wnt expression

We next sought to understand how *lhx2/9* expression is restricted to the anterior ectoderm and what promotes the differentiation of these neural precursors. In the sea urchin, this domain is initially delineated by the forkhead transcription factor, *foxq2*, as well as a sine oculis homeobox transcription factor, *six3* (Yaguchi et al. 2008; Wei et al. 2009). We previously characterized the expression of these genes in *P. miniata* and found that while *foxq2* is expressed only in the anterior-most ectoderm, *six3* is expressed in a much broader ectodermal domain, which extends below the mesodermal bulb at gastrula stages (Yankura et al. 2010). Here, we characterize the functions of *foxq2* and *six3* in the specification of the apical pole domain, and ultimately, the dorsal ganglia.

We find that *foxq2* is required for the expression of *lhx2/9* at 48hpf, as no *lhx2/9* is seen in *foxq2* knock-down embryos (Figure 3.6A vs. B). Likewise, *elav* is not expressed in the apical pole of *foxq2* morphants (Figure 3.6C vs. D). *Foxq2* is also needed for the expression of other genes shown to be important to the specification of the animal pole domain, *dkk3*, a secreted wnt agonist (Figure 3.6E vs. F), and *zic*, a transcription factor (Figure 3.6G vs. H)(Range 2014). All *elav* in the apical pole co-localizes with *foxq2*, indicating an additional requirement for *foxq2* in the differentiation of serotonergic neurons from *lhx2/9* positive cells (Figure 3.6I).

Interestingly, the expression domains of *foxq2* and *lhx2/9* only partially overlap at 48hpf (Figure 3.6J). In the sea urchin, *foxq2* is expressed broadly in early blastula stages and eventually becomes restricted to the apical-most ectoderm (Yaguchi et al. 2008). It may be that Foxq2 protein is still be present in the broader *lhx2/9* domain, assuming a similar spatial progression of *foxq2* expression occurs in *P. miniata*. Alternatively, Foxq2 may promote *lhx2/9* expression indirectly via a more diffusible signaling molecule, such as its known target, *dkk3*. We speculated that the *lhx2/9* cells that do not co-localize with *foxq2* may represent a progenitor population because they do not immediately differentiate into *elav* positive cells. To this end, we wanted to test whether all *lhx2/9* cells will differentiate into neurons if co-localized with *foxq2*. In Rx morphant embryos, *foxq2* expands posteriorly (Figure 3.6J vs. K). This places all *lhx2/9* cells into the *foxq2* territory (Figure 3.6K) and as a result all of these cells also express *elav* rather than maintaining a progenitor state (Figure 3.6L).

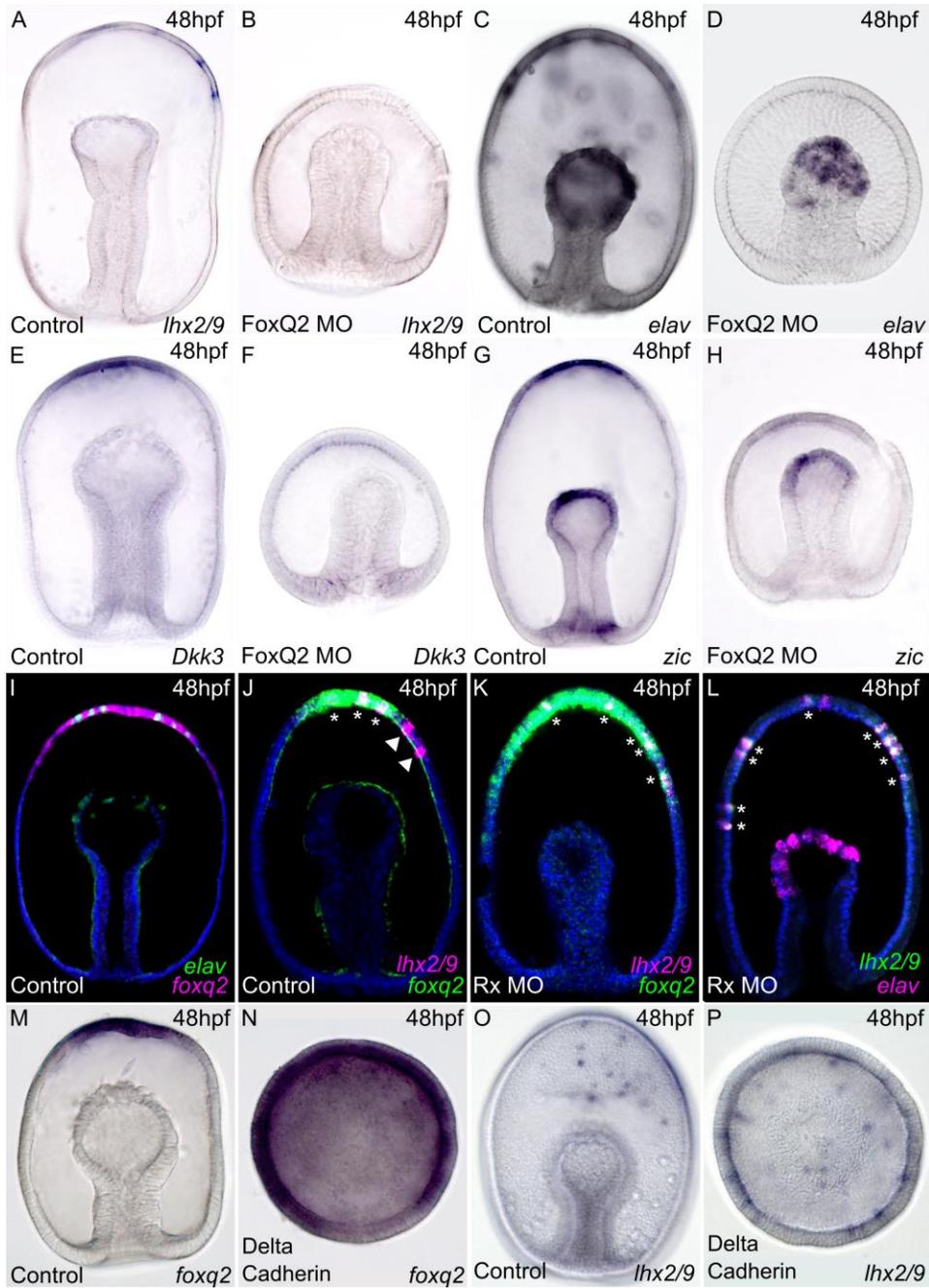


Figure 3.6: *Lhx2* requires an anterior pole domain defined by *foxq2* for specification and also to differentiate into neurons. A-B. Normal expression of *lhx2* (A) at 48hpf is lost when Foxq2 is knocked-down (B). C-D. Likewise, *elav* expression (C) is lost from the anterior pole when Foxq2 is knocked-down (D). *Elav* expression in the mesoderm is unaffected. Foxq2 is also required for the expression of other conserved apical pole domain genes, *dkk3* (E vs. F), and *zic* (G vs. H). I. *Elav* (green) and *foxq2* (magenta) completely co-express at 48hpf, suggesting a requirement for *foxq2* in neuron differentiation. J. *Lhx2* (magenta) expression only partially co-

localizes with *foxq2* (green) at 48hpf. Asterisks indicate co-expression, while arrows denote cells that only express *lhx2/9*. K. Rx MO causes expansion of *foxq2* and complete overlap with the *lhx2/9* territory. L. In Rx MO embryos, all *lhx2/9* cells are also *elav* positive, indicating that *foxq2* promotes inappropriate differentiation of otherwise proliferating *lhx2/9* cells. M-N. *Foxq2* is normally expressed only in the apical ectoderm (M). In the absence of wnt signaling, created by the introduction of *delta cadherin*, *foxq2* is expressed ubiquitously (N). As a result, *lhx2/9* expression also expands throughout the embryo (O-P).

It is well established that there is an important interplay between wnt signaling and neuronal development in vertebrates (Glinka et al. 1997; Houart et al. 2002; Andoniadou et al. 2011) and echinoderms (Range et al. 2013; Yankura et al. 2013), in which a gradient of Wnt establishes the AP axis and the anterior most ectoderm requires local Wnt antagonism. We hypothesize that *foxq2* establishes the apical pole domain by protecting this region from wnt signaling. However, wnt signaling also restricts *foxq2* expression to the apical pole. When canonical wnt signaling is abolished through introduction of *delta cadherin* mRNA (Wikramanayake et al. 1998; Logan et al. 1999), *foxq2* expression becomes ubiquitous (Figure 3.6M vs. N). This is in agreement with previous work which demonstrated that echinoderm embryos become ubiquitously neurogenic when c-Wnt signaling is lost (Range et al. 2013; Yankura et al. 2013). Additionally, *delta cadherin* embryos express *lhx2/9* (Figure 3.6O vs. P) and *elav* (Yankura et al. 2013) in spots throughout the entire embryo. The presence of *elav* throughout the entirety of these embryos suggests that all *lhx2/9* cells differentiate into neurons in this environment, most likely due to *foxq2*'s ubiquitous expression.

3.4.7 *Six3* Regulates the Size and Position of the Dorsal Ganglia through Wnt Signaling, but not its Initial Specification

In spite of its importance in creating the anterior neuroectoderm of the sea urchin, we did not find a requirement for *six3* in the initial specification of the sea star apical pole domain (Figure 3.7). Studies in the sea urchin have demonstrated that *six3* is needed for the expression of genes that define the apical pole domain, such as *foxq2*, *zic*, *dkk3*, and *sFrp1/5*, and therefore it is considered to be the “master regulator” of this territory (Yaguchi et al. 2008; Wei et al. 2009; Range 2014). We previously demonstrated that *six3* is crucial for anterior/posterior patterning, but not the specification of the ciliary band neurons (Yankura et al. 2013). Although

we were able to reproduce the stunted anterior ectoderm characteristic of Six3 morphants, we found that expression of *foxq2*, *lhx2/9*, and *elav* are not affected by loss of Six3 aside from having a smaller available territory (Figure 3.7A-C, vs. E-G). In the sea star, *foxq2* instead performs the role of setting up the apical pole domain and is responsible for regulating this set of target genes. This is unsurprising as *six3* is expressed in a much broader territory in the sea star ectoderm, while *foxq2* is restricted to the apical pole.

We wanted to understand what role *six3* might play in dorsal ganglia formation, especially since it has a strongly conserved role in not only in specifying the vertebrate forebrain, but also anterior neuroectoderm development across metazoans (Lagutin et al. 2003; Wei et al. 2009; Steinmetz et al. 2010). Later in development, 96hpf Six3 MO embryos have very little *lhx2/9* expression, and very few serotonergic neurons (Figure 3.7D vs. H, I vs. J), in spite of relatively normal earlier specification. Knock-down of Six3 did result in expansion of *wnt8* into the anterior ectoderm (Yankura et al. 2013), such that it abuts *foxq2* (Figure 3.7K vs. L). Figure 3.7 M vs. N demonstrates that while normally the more posterior *lhx2/9* cells do not express *elav* and instead continue to proliferate, Six3 morphants do not have this *lhx2/9* progenitor population. All *lhx2/9* cells in these morphants become *elav* positive right away. This results in a loss of the proliferating reserve of *lhx2/9* cells, and explains why a dorsal ganglia phenotype does not emerge in Six3 morphants until later development.

However, it is important to note that *lhx2/9* neither requires *six3* to be expressed, nor does it need *six3* in order to differentiate into a serotonergic neuron. This is similar to our previous finding that ciliary band neurons differentiate in Six3 morphants, albeit in a shifted AP position with respect to control larvae (Yankura et al. 2013). This is in stark contrast to the closely-related sea urchin model, which is unable to produce neurons of any kind in the absence of Six3 (Wei et al. 2009). We instead find that Six3's main role in dorsal ganglia development is to repress Wnt signaling and therefore create a broader domain in which *lhx2/9* can be expressed and drive neural progenitor proliferation, as is seen in the vertebrate forebrain (Lagutin et al. 2003; Lavado et al. 2008). In vertebrates, wnt signaling has been shown to impact the number of cortical neurons by promoting differentiation over proliferation (Hirabayashi et al. 2004; Munji et al. 2011). Our results indicate that this function is conserved in the sea star apical organ.

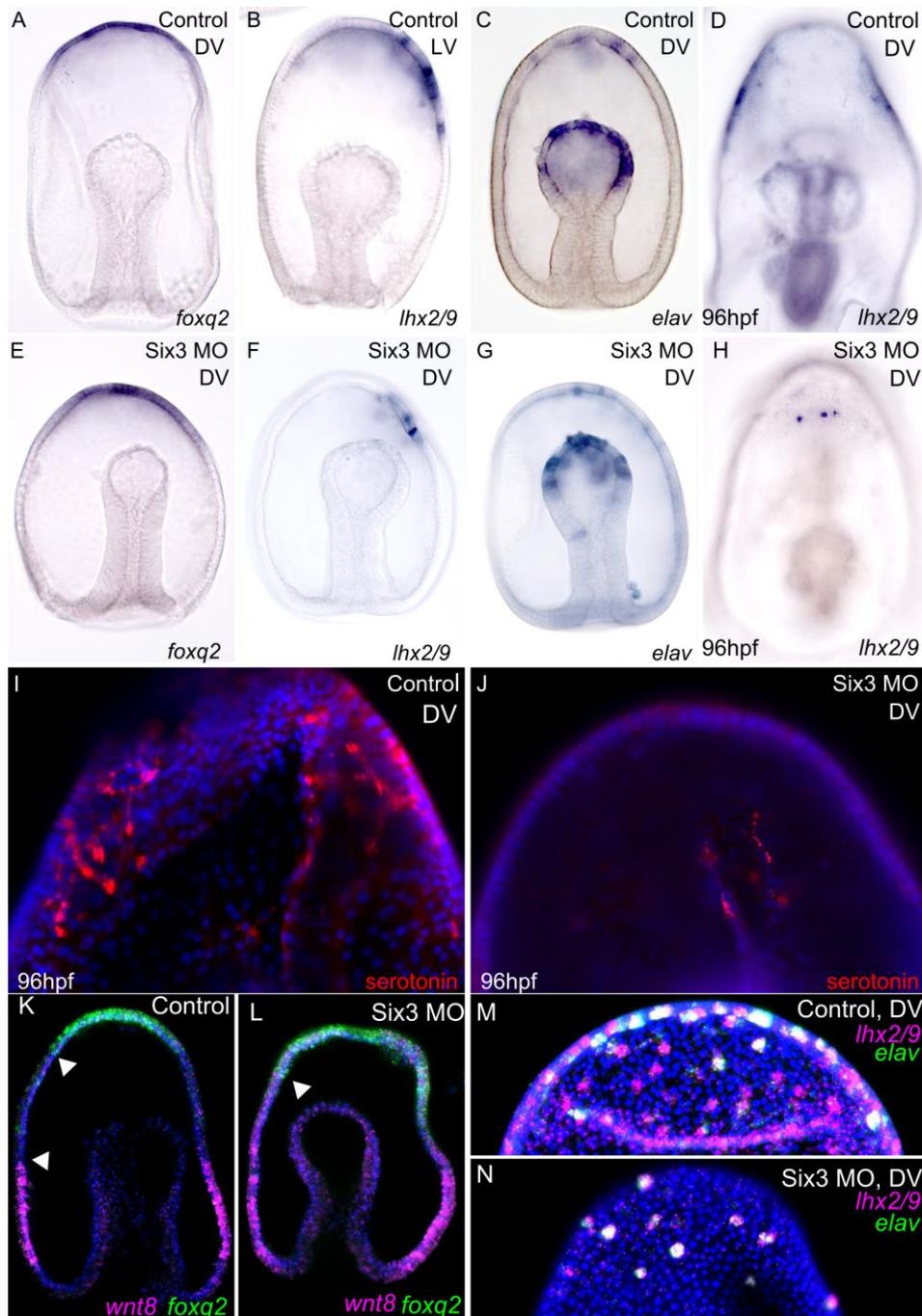


Figure 3.7: Six3 does not affect the initial specification of the apical pole domain, but instead maintains a proliferative *lhx2/9* territory free from both *foxq2* expression Wnt signaling. A-C, E-G. Apical pole domain specification is unaffected by the loss of *six3*, as revealed by the normal expression patterns of *foxq2* (A vs. E), *lhx2* (B vs. F) and *elav* (C vs. G) at 48hpf. Six3 morphants have a truncated anterior ectoderm, so these genes are expressed in a proportionally smaller area. Six3 morphants have fewer *lhx2/9* cells at 96hpf (D vs. H) and also a

smaller number of serotonergic neurons (I vs. J). K-L. *Six3* morphants have expanded *wnt8* expression, such that there is no space between the *wnt8* and *foxq2* expression domains. M. *Lhx2/9* is normally co-expressed with *elav* in the apical-most ectoderm, and expressed alone more posteriorly. N. In *six3* morphants, all *lhx2/9* cells are in the *foxq2* domain, causing them to differentiate, and thus are *elav* positive.

3.5 Discussion

Much previous work has shown similarity between the patterning of invertebrate ectodermal territories and the vertebrate forebrain (reviewed by Range 2014). Here we show that not only the expression domains, but also the functionality of these genes are conserved in the development of an echinoderm apical organ structure. The entirety of this process is depicted in our model shown in Figure 3.8.

Additionally, we find that two sea star nervous systems, the ciliary bands and dorsal ganglia, are specified from the same original pan-neurogenic ectoderm, but are patterned independently. Moreover, they seem to represent different types of neurons, as the dorsal ganglia neurons express serotonin, while the neurons of the ciliary bands do not. It has been observed in several echinoderm models that the serotonergic neurons associate physically with the ciliary bands (Byrne et al. 2007; Bishop et al. 2013). Our work demonstrates that the serotonergic neurons do not require ciliary band specific transcription factors in their early specification, although it is possible that they make use of similar patterning cues later in development, such as Ephrin signaling, to migrate to the same location. Understanding the processes that allow different neuronal cell types to be specified from a common stem cell pool is important, but not well-understood. This work enhances our knowledge of such mechanisms.

In vertebrates, *lhx2* is needed for the specification and growth of the telencephalon, or anterior forebrain, as well as the developing retina. It has been shown to play roles in both maintaining cortical progenitors and differentiation of the neurons that arise from such progenitors (Bulchand et al. 2001; Ando et al. 2005; Hägglund et al. 2011). In the mouse neocortex, *lhx2* is needed to maintain the proliferative state of neocortical progenitors; conditional knock-out mice form a much smaller neocortex with prematurely differentiating neurons (Chou and O'Leary 2013). Other work in mice suggests that *lhx2* is a selector gene, which directs stem cells to become cortical progenitors and suppresses alternate fates (Mangale

et al. 2008). Furthermore, in zebrafish embryos, *Lhx2/Lhx9* double knock-outs have stalled neurogenesis in the thalamus; progenitors accumulate but do not reach their terminal neural state (Peukert et al. 2011). We have previously shown that the AP axis of the sea star is patterned much like the brain or anterior-most region of a vertebrate embryo (Yankura et al. 2010). Therefore, in the sea star, *lhx2/9* is expressed in a region thought to be molecularly homologous to the forebrain region. Here, we also demonstrate that *lhx2/9* also performs similar roles in the sea star compared to vertebrates. We see that *lhx2/9* promotes proliferation of neural progenitors and their eventual differentiation into serotonergic neurons. *Elav*, a marker of post-mitotic neurons, is a conserved target gene of *lhx2/9* in sea stars (shown here) and vertebrates (Peukert et al. 2011). Expression of a *lhx2* (*apterous*) in the central nervous system has been observed in the annelid *Neanthes arenaceodentata* suggesting this function of *lhx2/9* may be ancient and deeply conserved (Marlow et al. 2014).

In all bilaterians except for chordates, the BMP side of the embryo is dorsal and the Nodal/Chordin side of the embryo is ventral, which has led to the theory that the dorsal-ventral axis was inverted in the chordate lineage (Nübler-Jung and Arendt 1996). This is supported by the fact that BMP is expressed on the dorsal side of on in the closest relatives of chordates, hemichordates (Lowe et al. 2006) and echinoderms (Saudemont et al. 2010). In hemichordates, BMP is not required for neurogenesis, but is needed to pattern neuronal cell types within a diffuse neuroectoderm (Lowe et al. 2006). We see a similar use of BMP signaling in neuronal patterning, as a particular neuronal type, the serotonergic neurons, require BMP, but pan-neurogenic genes such as *soxc* are unaffected by *Bmp2/4* knock-down. Interestingly, BMP2 and BMP4 have been shown to regulate the expression of *lhx2* in the mouse dorsal telencephalon in a bimodal fashion, in which high BMP blocks *lhx2* expression, but low BMP signaling promotes it (Monuki et al. 2001). If *lhx2* expression in the sea star is also sensitive to BMP2/4 levels, it could explain why two lateral dorsal ganglia form as opposed to a single ganglion in the dorsal-most anterior ectoderm. Distance from the concentrated source of the BMP activity could be crucial for positioning these neurons. This seems to be a common feature of neuronal development across metazoans; animals with a central nervous system require low levels BMP to produce neurons, but high levels are inhibitory. In *Drosophila*, BMP4 represses neurogenesis on the dorsal side of the embryo in a concentration dependent manner defined by BMP4 and Chordin diffusion (Biehs et al. 1996). In chordates, neurons only form in ventral regions, which

have very low levels of BMP signaling owing to the presence of BMP agonists such as Chordin and Noggin (De Robertis and Kuroda 2004).

Foxq2 has a highly conserved function in invertebrates. It has been implicated in the specification of the apical organ in numerous metazoans including even cnidarians (Sinigaglia et al. 2013), but is conspicuously absent in the vertebrate lineage (Kaestner et al. 2000; Mazet et al. 2003). *Foxq2* is expressed in the apical organ, and in many cases coincides with serotonergic neuron location, in mollusks (Santagata et al. 2012), annelids (Marlow et al. 2014), hemichordates (Fritzenwanker et al. 2014), cephalochordates (Yu et al. 2003), echinoderms (sea urchins) (Yaguchi et al. 2008) and here we show also in another echinoderm, the sea star. In many of these systems, the functional roles of *foxq2* in apical organ specification have not been tested, but in all those who have, *foxq2* plays an integral role in the process. We find that *foxq2* has multiple functions in this dorsal ganglia development; it is required for the expression of *lhx2/9* and for differentiation of *lhx2/9* cells into serotonergic neurons. At least part of this may be tied to maintaining a Wnt-free domain by providing positive inputs into a variety of Wnt agonists, such as *dkk3*. In vertebrates, although *lhx2* and *lhx9* are important for neurogenesis, they have not been implicated in the production of serotonergic neurons per se. Loss of *foxq2* in these organisms might have decoupled *lhx2* from this process and allowed this transcription factor to be co-opted by other neural GRNs.

In vertebrates, loss of *Six3* does not alter the early specification of the anterior neural region; however, it does ultimately lead to truncation of the forebrain, presumably due to expansion of Wnt signaling into this region (Lagutin et al. 2003; Lavado et al. 2008). This is reminiscent of the phenotype observed in our sea star *Six3* morphants, however it is quite different from what has been reported for euechinoid sea urchins, which are unable to make neurons of any sort without *Six3* (Wei et al. 2009). In pencil urchins, unlike euechinoids, *six3* is not restricted to the apical pole, but is expressed in lateral clusters within the ciliary bands. Serotonergic neurons eventually develop here (Bishop et al. 2013). This suggests that the discrete apical organ seen in euechinoids is probably a derived structure which in part evolved through progressively restricted expression of *six3*, and the lateral dorsal ganglia morphology is more representative of the ancestral deuterostome state.

Our work suggests that *six3*'s primary function in development of the dorsal ganglia is to repress Wnt signaling such that its effects are restricted to the posterior half of the embryo. Wnt

signaling has been shown to have posteriorizing effects in the early embryos of a variety of deuterostomes (Nordström et al. 2002; Pani et al. 2012; Range et al. 2013) and is typically found in a gradient that directs A/P axis specification. *Six3*'s wide domain in the sea star allows a broader area than the apical pole domain to be permissive to the progression of neuronal precursors. When *six3* function is blocked, this region of proliferating neural progenitors is lost, resulting in fewer serotonergic neurons and a restriction of these neurons to a small region in the anterior of the embryo. This phenotype is much like the normal pattern of serotonergic neurons in the sea urchin (Bisgrove and Burke 1986; Wei et al. 2009), suggesting that changes in *six3* expression and function may have contributed to differing apical organ morphologies among echinoderms. Importantly, *six3* is expressed in an apically-restricted domain in sea urchins, suggesting that it does not repress Wnt signaling more posterior territories in order to delineate a neural proliferative zone (Wei et al. 2009).

Our model (Figure 3.8) suggests that serotonergic neurons are patterned within a broadly neurogenic ectoderm by the intersect of BMP2/4 signaling, and Wnt repression, directed by *foxq2* and *six3*. Therefore, changing the boundaries of these signaling pathways, either by changing the expression patterns of the overlying transcription factors, such as *rx*, or by changing regulatory inputs completely will result in evolution of apical organ morphology. Previous work in many organisms has pointed to widely conserved early patterning of the ectoderm and potential homology between apical organs and the vertebrate forebrain. Here, we demonstrate that there is also conservation in the mechanisms used to generate neural precursors and the trajectory they take to become mature neurons. Some important parts of this forebrain network (e.g. *eya*) are not found in the apical plate, but are present in the mesodermal bulb, which is incidentally the other expression domain of *six3* (Yankura et al. 2010). These parts of the network could have therefore been moved into or out of the apical plate region via *six3* as a top-tier regulator. Later addition of these genes to the apical organ would have allowed for greater cell-type and sub-region diversification. Many of these early patterning and subsequent neurogenic components are used in multiple contexts in vertebrate forebrain development, especially the telencephalon, hypothalamus, and developing eyes, suggesting that the ancestral GRN might have been similar to what we find in echinoderms, and subsequently duplicated and diverged to allow for many specialized forebrain regions in vertebrates.

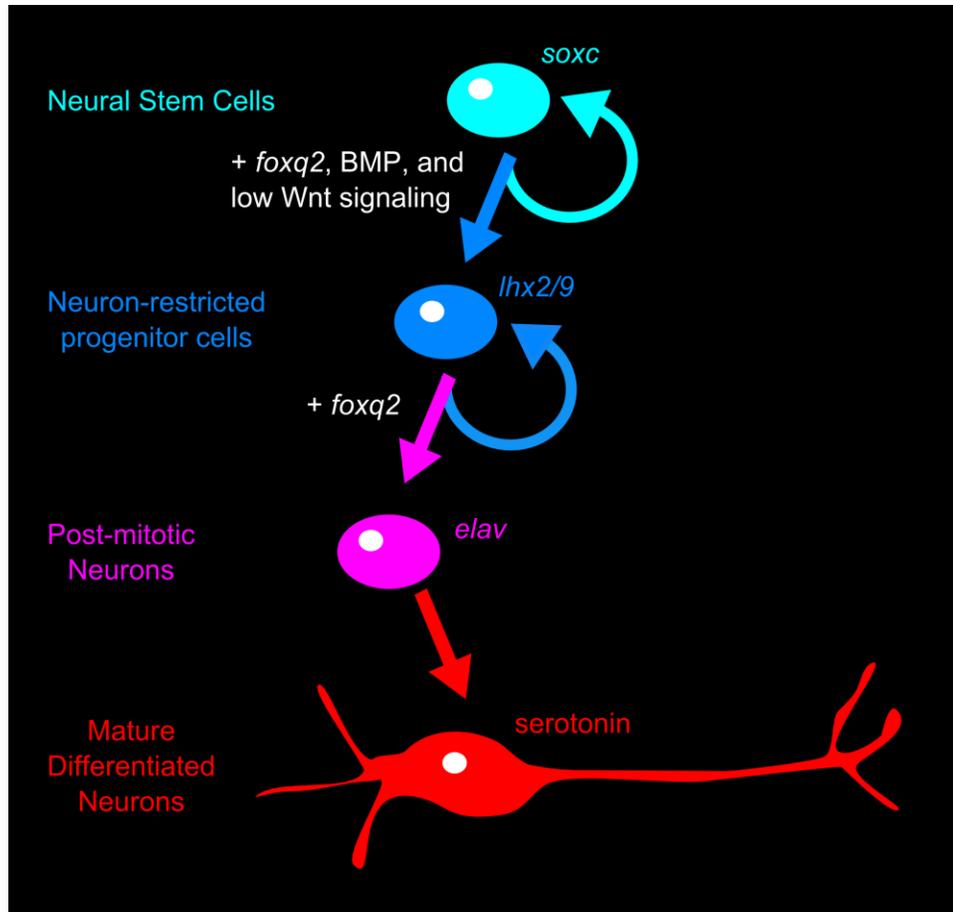


Figure 3.8: Model of sea star serotonergic neurogenesis.

3.6 Methods

Embryo Culture and Microinjections: *Patiria miniata* embryos were cultured and injected as described previously (Cheatle Jarvela and Hinman 2014). All embryos were reared in seawater at 15°C until fixation. Morpholino antisense oligonucleotides were designed by GeneTools.

Morpholino sequences are as follows:

PmLhx2/9: 5' CCGGTTGCAAAGTGAAATACATTCA 3'

PmRx: 5' GCACAGCTCCAACCCAGATAGCATC 3'

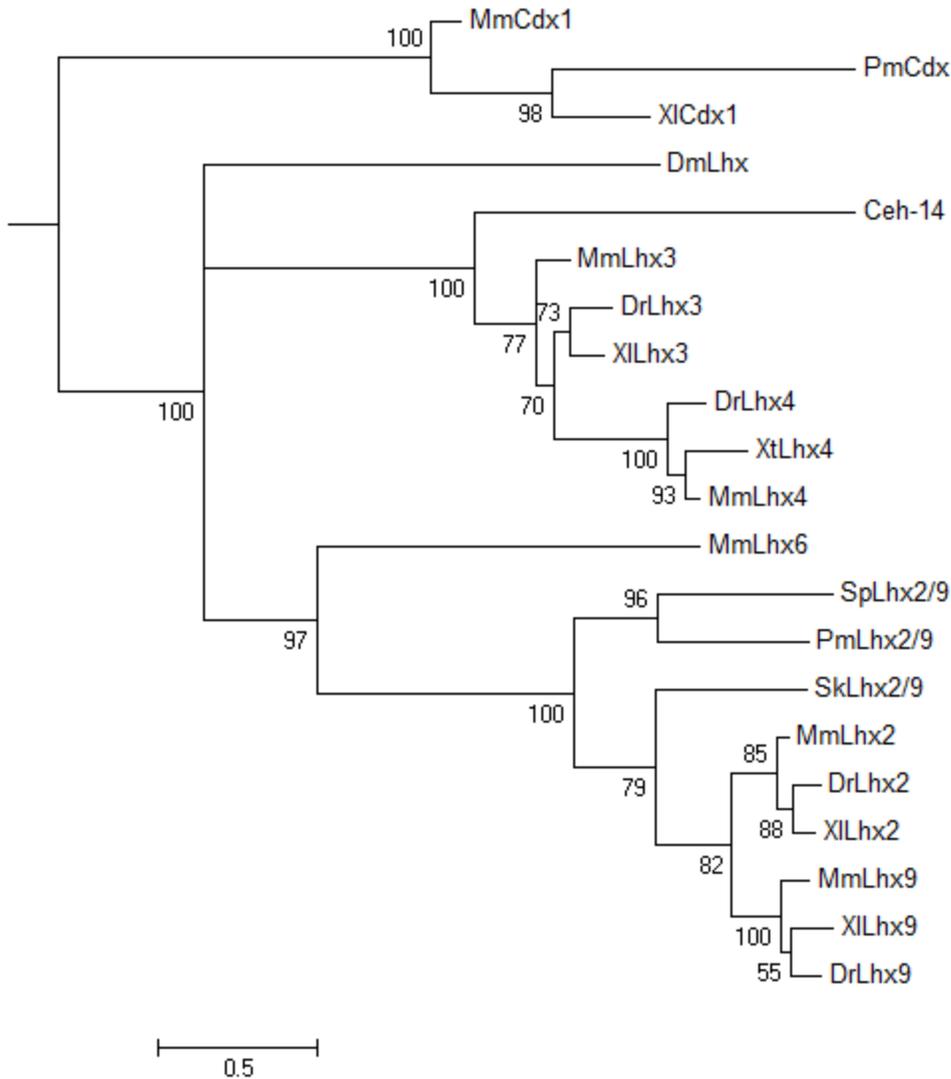
Other sequences can be found in our previous publication (*PmBmp2/4*, *PmFoxg*, *PmFoxq2*, *PmSix3*, *PmSoxc*)(Yankura et al. 2013).

Whole Mount *In Situ* Hybridization (WMISH) and Fluorescent *In Situ* Hybridization

(FISH): WMISH was performed as previously described (Hinman, Nguyen, and Davidson 2003; Yankura et al. 2010) using digoxigenin-or dinitrophenol labeled antisense RNA probes. FISH embryos were mounted in Slowfade media (Life Technologies) and imaged by confocal microscopy with a Carl Zeiss LSM-510 Meta DuoScan Inverted Confocal Microscope.

Immunofluorescence: Embryos were fixed in 4% paraformaldehyde/MOPS (100mM MOPS, 2mM MgSO₄, 1mM EGTA, 80mM NaCl, 0.1% Triton x-100) at room temperature for 90 minutes. Embryos were stored in 70% EtOH at -20°C until use. Embryos were stepped into MAB/0.1% Triton x-100 (100mM malic acid, 150mM NaCl), washed four times in MAB/0.1% Triton x-100, and blocked in 2% BSA/MAB/0.1% Triton x-100. Embryos were incubated with 1:250 rabbit anti-serotonin (Sigma), diluted in 2% BSA block, overnight at 4°C. Embryos were then washed four times with MAB/0.1% Triton x-100, and incubated with 1:1000 anti-rabbit alexa-fluor 568, diluted in 2% BSA block. Embryos were rinsed twice, incubated with 1:10,000 DAPI for 20 minutes, and rinsed four more times in MAB/0.1% Triton x-100. Embryos were imaged as described for FISH.

EdU Labeling: EdU labeling was performed using the Click-It Plus EdU 488 Imaging Kit (Life Technologies), with some modifications to the manufacturer's instructions. Briefly, EdU was added to embryo culture at a final concentration of 10 µM and incubated for 15 minutes at 15°C. Embryos were fixed as processed for FISH as previously described. After completion of the FISH protocol, embryos were washed three times in PBS, permeabilized in 0.5% Triton-x100/PBS, and then washed three times in 3% BSA/PBS. Embryos were incubated for 30 minutes in Click-It cocktail (prepared according to manufacturer's instructions). After three additional PBS washes, nuclei were stained with DAPI imaged by confocal microscopy as described above.



Supplemental Figure 3.1: PmLhx2/9 is an ortholog of vertebrate Lhx2 and Lhx9. Tree topology was determined using a MrBayes model (TOPALI v2.5), and is based on a character alignment that includes full length sequences from Genbank. Lengths of branches are drawn to the scale indicated and bootstrap values indicate the number of times a node was supported in 100 replicates. Homeobox transcription factor Cdx and *Drosophila* lim homeobox were used to create outgroups. Ce- *Caenorhabditis elegans*, Dm- *Drosophila melanogaster*, Dr-*Danio rerio*, Mm-*Mus musculus*, Pm- *Patiria miniata*, Sk-*Saccoglossus kowalevskii*, Sp-*Strongylocentrotus purpuratus*, Xl-*Xenopus laevis*, Xt- *Xenopus tropicalis*

Chapter 4: Conclusions and Future Directions

4.1 Conclusions

In this work, we have studied the evolution of gene regulatory networks (GRNs) from two different perspectives. First, we demonstrate that the transcription factor Tbr has evolved differences in secondary site binding, both within the echinoderm lineage and also among deuterostomes (Chapter 2) (Cheatle Jarvela et al. 2014). However, the primary binding site is conserved across both of these evolutionary distances. We next demonstrate that many aspects of the GRN controlling neurogenesis is conserved between the sea star and the vertebrate forebrain, including not only the early patterning mechanisms and also the progression of neural stem cells to differentiated neurons (Chapter 3). In both chapters, we briefly speculate about differences between the neurogenic ectoderm of the sea star and the sea urchin. The work in Chapter 2 suggests that differences in secondary site usage have caused sea urchin Tbr to lose ectodermal target genes. Meanwhile, the work in Chapter 3 describes a few striking differences in early patterning, function of key apical organ genes, and later neuron number and placement between sea stars and sea urchins. Many of these differences entail different regulatory connections between genes, which are most likely explained by CRM evolution. However, our preliminary data also potentially ties the biochemical change in Tbr DNA-specificity to this difference. In the remainder of this chapter, I will put forth the hypothesis that the changes outlined in Chapter 2 may have in part caused the morphological differences described in Chapter 3. Additionally, I describe preliminary work and future experiments aimed towards tying these bodies of work together.

4.2 Background and Rationale: Tbr Function within Echinoderms and across Deuterostomes

At first glance, it seems that the most interesting aspect of Tbr's evolution in echinoderms is its potential role in creating a novel cell type in the sea urchin lineage, the primary mesenchyme cells, or PMCs. In the sea star, *tbr* is expressed in the ectoderm and endomesoderm during blastula stages (Hinman et al. 2007), while it is expressed only in the PMCs of similar stages of sea urchin (Croce et al. 2001; Oliveri et al. 2002). However, based on

the available data, we do not think this new lineage came into existence due to Tbr's changed binding preference, although it could have played a smaller role in making this lineage what it is today after it emerged. For one thing, *tbr* seems to be a fairly recent addition to the PMC GRN. In this network, *tbr* has relatively few connections to morphoregulatory genes, while *alx1* and *ets1* do most of the heavy-lifting (Rafiq et al. 2012). More importantly, sea cucumbers, a closer relative to the sea urchins than sea stars, have a PMC lineage that expresses *alx1* but not *tbr* (McCauley et al. 2012). These cells secrete skeleton, but the result is not as extensive or well-patterned as what is seen in sea urchins. This work also demonstrates that in sea cucumbers, *tbr* is expressed predominately in other mesodermal cell types, and potentially also in the ectoderm. Collectively, these studies indicate that *tbr* joined the PMC GRN network after skeleton-producing PMCs evolved.

The sea star endomesoderm is predicted to be similar to the ancestral territory that gave rise to the PMC lineage; PMCs evolved by carving out a sub-territory within the broader endomesoderm (McCauley et al. 2012). A substantial proportion of the genes regulated by *tbr* in the sea urchin are conserved in the sea star endomesoderm, suggesting that this gene hasn't acquired much novel functionality that would help the PMCs become different than other mesodermal cell types. What little we know about *tbr*'s target genes in the endomesoderm vs. the PMCs thus far supports this idea; target genes such as *erg*, *ets1/2*, and *delta* are conserved between species (Oliveri et al. 2002; Hinman and Davidson 2007; McCauley et al. 2010). A case where *tbr* regulates an endomesoderm target in the sea star but not the sea urchin, *otx*, has been attributed to a CRM change (Hinman et al. 2007). Other important mesodermal targets of sea star Tbr, such as *hex*, are not maintained in the sea urchin. Loss of such connections might be the result of reduced secondary site binding in the sea urchin. However this seems unlikely. Exclusive PMC expression of *tbr* in sea urchins is an extremely recent development; irregular sea urchins, such as sand dollars, express *tbr* in both the PMCs and throughout the rest of the endomesoderm (Minemura et al. 2009). This suggests that differences in secondary binding did not drastically affect *tbr*'s broad endomesodermal functions, as they might be conserved even in specific sea urchin clades. If *SpTbr* has does have fewer critical functions in the broader endomesoderm and ectoderm than other echinoderms, it would have more flexibility to develop a more restricted expression pattern and gain functions in the PMC lineage. Aside from that potential advantage, this change does not seem to have impacted the novel PMC lineage.

We do expect that *PmTbr* has important functions in the ectoderm that *SpTbr* is unable to perform, and that this contributed to the large difference in neuron location and number between these species. Previous work as demonstrated that *PmTbr* regulates the expression of *Pmotx* through the *OtxG* module, both in the endomesoderm and in the ectoderm (Hinman et al. 2007). Other work from our lab has predicted that *otx* acts near the top of neurogenic GRNs in the sea star and may have roles in both early patterning of the pan-neurogenic ectoderm as well as later roles in the developing ciliary band neurogenic territory (Hinman, Nguyen, and Davidson 2003; Yankura et al. 2013). As of now, we do not know whether Tbr also functions in the patterning and specification of the apical organ neurogenic territory, but we hypothesize that it does since it is expressed there.

In vertebrates, Tbr2 (Eomesodermin) is crucial to the very early patterning of the mesoderm, just as it is in echinoderms (reviewed by Showell et al. 2004). It is also expressed in the telencephalon region of the forebrain later in *xenopus* development (Ryan et al. 1998) and induces neural stem cells to become intermediate neural progenitors in this region (Sessa et al. 2008). Conditional inactivation of Tbr2 expression during early brain development leads to microcephaly due to a severely reduced number of neural progenitors (Arnold et al. 2008).

PmTbr is likely to have some conserved roles in the apical organ, which has similar patterning and progression of neurogenesis compared to the vertebrate forebrain (Chapter 3)(Yankura et al. 2010). Obviously, the vertebrate forebrain is a great deal more complex than an echinoderm apical organ. We discovered in Chapter 2 that *PmTbr* recognizes a different secondary site than *MmEomes*. Additionally, *PmTbr* is orthologous to three vertebrate paralogs, Tbr1, Tbr2, and Tbx21. Therefore, we expect this will result in some conserved and some non-conserved Tbr functions between the echinoderm apical organ and the vertebrate forebrain. Thus, detailed study of Tbr's functions in the sea star ectoderm may shed light on the evolution of the vertebrate forebrain.

4.3 Experimental Approach and Preliminary Data: Comparative ChIP-seq

The work described in Chapter 2 and Chapter 3 introduces many new questions. Was the use of conserved, primary binding sites important to maintaining Tbr's roles in early mesoderm development across 800 million years of deuterostome evolution? What consequence did a change in secondary site binding have on echinoderm development? Is this change in any way

connected to morphological diversity in echinoderm apical organs and ciliary band patterns? Did Tbr have important roles in the establishment of the novel sea urchin PMC lineage that have been overlooked? To answer these questions, we must have comprehensive knowledge of Tbr's target genes and binding site usage in both organisms.

To obtain such data, we have performed ChIP-seq on mesenchyme blastula stage sea urchins and hatched blastula stage sea stars. We chose these stages because they are developmentally equivalent as far as we can tell, and ideally staged to capture both skeletogenesis in the sea urchin and endomesodermal, as well as early ectodermal patterning functions in the sea star. Study of *PmTbr*'s neurogenesis functions may require that we also obtain ChIP data for a later stage, as much of early neurogenesis occurs during gastrula stages. ChIP methodology is described in Appendix 1. Illumina HiSeq was performed at the Yale Center for Genome Analysis. Preliminary MACS models (Zhang et al. 2008) predict approximately 50,000 sea star peaks and 7,000 sea urchin peaks. We expect the actual number of peaks will be much smaller as we are currently filtering out redundancy from PCR amplification and removing reads that do not map uniquely to their respective genomes. We are also in the process of obtaining RNA-seq data for Tbr morphants to aid in validating this ChIP data as well as to shed further light on Tbr's function. As a proof of principle, Figure 4.1A demonstrates that our ChIP-seq experiment was a success. Our MACS model has predicted a peak downstream of *Pmotx* that almost exactly coincides with the published OtxG sequence's location (compare purple track to turquoise track).

We speculate that Tbr may perform roles in the broad neurogenic ectoderm, ciliary band GRN, and also apical organ GRN. First, we know that *tbr* is expressed dynamically in the ectoderm, appearing first throughout the anterior half of the ectoderm, and later becoming restricted to the ciliary bands (Hinman et al. 2007)(Chapter 2, Figure 2.5). Additionally, we find a ChIP peak within a few kb of the *Pmwnt8* gene (Figure 4.1B), which had been shown to be crucial to defining the A/P limit of neurogenesis for the apical organ (Chapter 3) and also is important to establishing the proper A/P location of the ciliary bands (Yankura et al. 2013). Knock-down of Tbr results in a similar up-regulation of *wnt8* compared to what we have previously described for Six3 knock-down embryos (Figure 4.1C and Chapter 3) (Yankura et al. 2013). Here, a sea star zygote was allowed to cleave into two-cells, and one of the resulting blastomeres was injected with Tbr MO. The injected side of the embryo expresses *wnt8* in all

but the apical-most ectoderm, which is a much broader domain than seen in the uninjected half of the embryo. Interestingly, Tbr neither affects the expression of *six3* (Figure 4.1D vs. E) nor does it produce ChIP peak in the vicinity of the *six3* locus. This suggests that Six3 might actually regulate *wnt8* via Tbr instead of acting on *wnt8* directly. Six3 and Tbr may also repress *wnt8* independently, but if that is the case neither can do so sufficiently alone. The putative *wnt8* CRM contains primary Tbr motifs, so the loss of this connection in sea urchins must be tied to changes in *tbr* expression rather than changes to Tbr DNA-specificity.

Additionally, we find a ChIP peak near *PmRx* (Figure 4.1F). We suspect that Tbr will activate rather than repress *rx* because they are expressed in largely overlapping anterior ectodermal domains (Hinman et al. 2007; Yankura et al. 2010). This peak contains a centrally located secondary Tbr site (Figure 4.1G), suggesting that this interaction would have been altered as Tbr's binding preference diverged. In Chapter 3, we show that *rx* helps to maintain a proliferative zone for *lhx2/9*-positive neural progenitors by repressing *foxq2* expression in this territory. *Foxq2* induces differentiation of these progenitors. Therefore, the ultimate number of serotonergic neurons in the larvae will be likely be reduced when *rx* expression is abrogated due to a lack of proliferative precursors, same as we observe in Six3 morphants. If Tbr indeed affects both the anterior and posterior boundary of the proliferative zone by regulating both *wnt8* and *rx*, then the Tbr morphants should have very few serotonergic neurons.

Furthermore, in sea urchins, which do not have ectodermal Tbr to delineate this proliferative zone, only a handful of serotonergic neurons are ever produced, and they are located in the apical-most ectoderm rather than in the broader range seen in sea stars (Bisgrove and Burke 1986). It could be that when Tbr lost the ability to regulate *rx* through secondary site binding, the result would have been a very truncated apical organ. We observe that this phenotype is not lethal to Six3 morphant sea star embryos (Chapter 3), and therefore there shouldn't have been strong selection pressure against such a change in the ancestor of sea urchins. Tbr's function in preventing *wnt8* expansion would have been moot at this point since the proliferative zone would have already been non-existent; therefore, degenerative changes to the CRMs of other genes, for example, *wnt8* and *tbr* itself, would have easily followed the initial loss of Tbr secondary site binding. The opposite could also be true. If Tbr lost its ectodermal expression domain through a CRM change, there would be less selection pressure to maintain the secondary site, and it would then have been lost in the lineage leading to sea urchins.

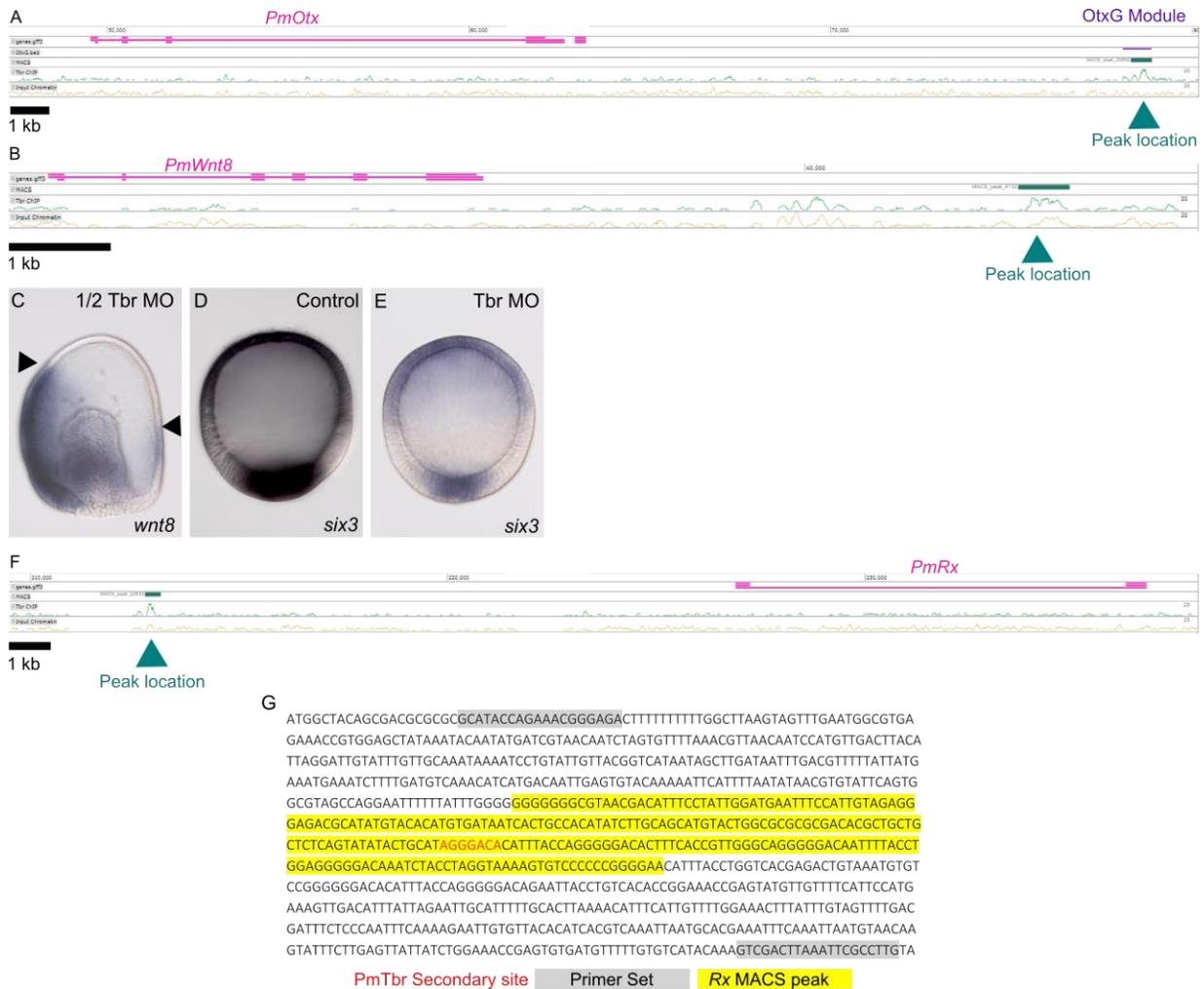


Figure 4.1: A, B, F. Visualization of genomic regions including tracks that indicate the locations and enrichment of gene CDS (pink), MACS peak (turquoise), Tbr ChIP (green), and Input Chromatin (yellow). A. Genomic region surrounding *PmOtx*. Purple indicates the location of the published *OtxG* CRM (Hinman et al., 2007). B. Genomic region surrounding *PmWnt8*. C. Tbr represses the expression of *wnt8* in the ectoderm. The left half of the embryo has been injected with a morpholino against Tbr. The right half of the embryo is uninjected. *wnt8*'s expression is abrogated in the injected half, such that it mimics the phenotype of Six3 MO embryos. D. Normal expression of *six3* at hatched blastula stage (Image: Kristen Yankura). E. Expression of *six3* is not significantly affected by Tbr knock-down. F. Genomic region surrounding *PmRx*. G. Sequence of Rx MACS peak (yellow) and surrounding sequence. A centrally located *PmTbr* secondary site (red) is likely to be the explanatory site for this peak.

Much experimental evidence is required to validate this evolutionary scenario; we must demonstrate that Tbr does indeed regulate *rx* and determine the apical organ phenotype of Tbr morphants. We also must determine if the identified peak regions correspond to functional regulatory regions and compare the CRM regions of sea urchin and sea star *rx* and *wnt8*. Expression pattern and genomic data from other echinoderm species would also be helpful to elucidate the order of events leading to changes in apical organ morphology.

Furthermore, we do not yet know the causative amino acid changes in the Tbr DNA-binding domain that lead to differences in secondary site binding. This is required to determine when secondary site binding was lost in the echinoderm lineage, and whether the timing

corresponds to apical organ morphological changes. In Chapter 2, we predicted that two amino acid changes that occur in suspected DNA-binding positions are responsible. To this end, we have produced Tbr DNA-binding domain constructs where these amino acids have been replaced with the sea star version in the sea urchin protein and vice versa. We have also generated chimeric proteins in which the N-terminal half is from sea star and the C-terminal from sea urchin and vice versa to begin to tease out other potential functional changes. These mutant DNA-binding domains will ultimately be tested for DNA-binding affinity and specificity by SPR as described in Chapter 2.

Finally, there are opportunities to learn about the impact of conserved primary vs. evolving secondary site binding on the evolution of the vertebrate forebrain. Recently published ChIP-seq data from gastrula stage *Xenopus* embryos offers an ideal comparison for our own datasets (Gentsch et al. 2013). Tbrain has duplicated and diverged into three paralogs in the vertebrate lineage: Tbr1, Tbr2 (Eomesodermin), and Tbx21 (Chapter 2, Figure 2.1). The binding specificities of Tbr1 and Tbx21 (T-bet) have not been determined by Protein-binding microarray, so it is unclear whether either of these uses the same secondary motif as *PmTbr*. Both *tbr1* and *tbr2* are expressed in the developing telencephalon (Bulfone et al. 1999). Meanwhile, *tbx21* is expressed exclusively in olfactory bulb, a different forebrain region than the other paralogs, and also in the thymus during development (Faedo et al. 2002). It is therefore possible that Eomesodermin had opportunities to sub-functionalize and neo-functionalize, as described in Chapter 1. An intriguing possibility is that as these new paralogs emerged, they may have diverged in secondary site binding and contributed to the regionalization of the vertebrate forebrain. A study of this would be an interesting line of research, but beyond the scope of this thesis. However, only *tbr2* seems to be expressed during the earlier gastrula stages we are interested in, so comparisons of Tbr functions in early patterning and neurogenesis may be straightforward (Bulfone et al. 1999). Thus despite this potential complexity in paralog functions and DNA-binding, we could still learn much about early neurogenesis in vertebrates vs. echinoderms by comparing ChIP-seq datasets.

Appendix 1: ChIP Method

This protocol has been modified from methods used by Richard Myers's lab (Mortazavi et al., 2006) and Craig Nelson's lab (personal communication).

For best results, make all buffers fresh daily, with the exception of PBS and Micrococcal nuclease buffer. All steps must be performed on ice or in a cold-room unless otherwise specified.

A. Embryo Collection and Cross-Linking

1. Collect embryos. Aim for 1×10^8 cells. A blastula stage embryo has roughly 1000 cells, so collect approximately 100,000. This can be scaled up or down as needed. I estimate the number of embryos by taking 1 mL of well-mixed culture and counting the number of embryos. Repeat three times and calculate the average embryos/mL. Then measure an appropriate volume of culture and collect embryos with a 100 μ m filter cup. Rinse embryos into a 50 mL conical tube (on ice).
2. Allow embryos to settle on ice for several minutes. Then remove as much sea water as possible. Add ice cold PBS up to 36 mL.
3. Fix in 1% formaldehyde. Formaldehyde comes at a 37% solution: add 1 mL to 36mL PBS+ embryos. Mix gently, but thoroughly. Fix for 10 minutes at room temperature.
4. Add 2.5M glycine to a final concentration of 0.125 M (2mL) to stop cross-linking. Mix gently.
5. Pellet embryos at 2,000 RPM for 5 minutes at 4°C.
6. Place cells on ice and remove fix. Fill the tube with ice cold PBS and gently resuspend cells. Pellet embryos at 2,000 RPM for 5 minutes at 4°C.
7. Remove PBS and snap-freeze cell pellet. Store at -80°C. The Myers's lab recommends preparing as much sample as possible at this point because the cells are stable stored this way indefinitely.

B. Preparation of Sheared Chromatin

1. Wash pellet twice with 1 mL ice cold PBS per $1-2 \times 10^7$ cells. Resuspend pellet gently. Centrifuge at 3,000 RPM at 4°C for 5 minutes after each wash.
2. Resuspend pellet in 1 mL cell lysis buffer and transfer to 1.5 mL Eppendorf tubes. Incubate on ice for 10 minutes.
3. Pass lysate through a 25 gauge needle five times. Gently expel material from the syringe to avoid producing foam.
4. Centrifuge at 5,000 RPMs for 4 minutes at 4°C to pellet nuclei.
5. Remove supernatant and resuspend pellet in 1 mL Micrococcal nuclease buffer. Add 1 μ L of nuclease to each tube (diluted in Micrococcal nuclease buffer).
6. Incubate at 37°C for 20 minutes. Mix tubes by inverting them every 3-5 minutes.
7. Stop digestion by adding 100 μ L of 0.5M EDTA.
8. Centrifuge at 5,000 RPMs for 4 minutes at 4°C to pellet nuclei.
9. Resuspend in nuclei lysis buffer. Incubate on ice for 10 minutes.
10. Centrifuge at 13,000 RPMs for 10 minutes at 4°C to pellet remaining cellular debris.
11. Transfer all supernatants to a new 15 mL conical tube and mix gently by inverting the tube.

12. Aliquot chromatin (200 μ L per aliquot). Make one 50 μ L aliquot and set aside for immediate quantification. Flash-freeze chromatin and store at -80°C until use.

C. Determination of Chromatin Fragment Size and Yield

1. Add 450 μ L of IP buffer to 50 μ L chromatin aliquot.
2. Reverse cross-link and RNase-treat by adding 20 μ L of 5M NaCl and 1 μ L of 10 mg/mL of RNase A. Incubate at 65°C for 4-5 hours. If needed, the sample can be stored at -20°C after this step.
3. Add 10 μ L 0.5 M EDTA, 20 μ L 1M Tris (pH 6.5) and 2 μ L 20 mg/mL Proteinase K. Incubate at 45°C for 1 hour.
4. Phenol/chloroform extract twice, followed by chloroform only.
5. Precipitate DNA with 53 μ L 3M NaAc and 1.3 mL ice cold 100% EtOH. Incubate at -20°C for a minimum of 1 hour. This step may be performed overnight.
6. Centrifuge at 12,000 RPM for 20 minutes at 4°C .
7. Resuspend in 50 μ L of LoTE and spec. Calculate the total μg of chromatin in the 50 μ L sample. Multiply by four to get the μg of chromatin per 200 μ L IP aliquot.
8. Run 10 μ L on a 1% gel to check fragment size. Ideal samples will have a mean length of 200-300 bp and range from 100 to 500 bp in length.

D. Immunoprecipitation

1. Couple antibody to beads
 - a. Prepare 5 mg/mL BSA in PBS. Filter Sterilize.
 - b. Add 50 μ L resuspended magnetic Protein-G beads to 1 mL PBS/BSA for each experiment. Plan to have a mock IP to pair with each ChIP (2 tubes/experiment)
 - c. Vortex briefly. Pull down beads with magnetic rack and remove supernatant. Resuspend in 1 mL PBS/BSA.
 - d. Repeat 3x
 - e. Add 10 μg antibody, or an equal volume of pre-immune sera for mock-IP
 - f. Rock o/n at 4°C
2. Pull-down chromatin
 - a. Thaw two 50 μg chromatin samples per experiment on ice. Bring each up to 1 mL with RIPA buffer.
 - b. At the same time, wash beads 3x in PBS/BSA. Do not vortex to resuspend! Mix very gently so as not to decouple the antibody. After the third wash, resuspend beads in 100 μ L RIPA buffer.
 - c. Add each set of beads to 1 mL chromatin.
 - d. Rock o/n at 4°C
3. Isolate Chromatin.
 - a. Pull-down beads. Discard supernatant.
 - b. Wash 3x in LiCl buffer for 3 minutes each. Resuspend VERY gently. Clumps of beads are indicative of a successful experiment. Do not force these apart, but do dislodge the clumps from the tube to promote removal of unbound chromatin. Performing washes on a rocker seems to allow for ideal mixing conditions.
 - c. Wash 1x in TE buffer.

- d. Resuspend beads in 300 μ L TE (now mix hard enough to break-up bead clumps). Add 9 μ L 10% SDS (in dH₂O) and 15 μ L 20 mg/mL Proteinase K. Incubate at 65°C o/n.
- e. Thaw an aliquot of chromatin to use as the input control. Bring up to 400 μ L with TE and add 9 μ L 10% SDS (in dH₂O) and 15 μ L 20 mg/mL Proteinase K. Incubate this overnight at 65°C along with the IP samples.
- f. Vortex beads and transfer supernatant to new tubes. Leave total chromatin in heat block until step h.
- g. Resuspend beads in 100 μ L TE. Add 11 μ L 5M NaCl. Vortex. Collect beads. Add the supernatants to the first collection.
- h. Add 4 μ L of 20 mg/mL glycogen to all tubes.
- i. Phenol/Chloroform extract twice. Chloroform extract once.
- j. Add 40 μ L NaAc, and 1 mL cold 100% EtOH. Precipitate for a minimum of 1 hour, but as long as overnight.
- k. Spin down DNA at 12,000 RPM for 20 min at 4°C. Dry pellet and resuspend in 50 μ L LoTE. Spec samples.
- l. Make a 10x dilution with 5 μ L into 50 μ L TE for future qPCR. Store this and the concentrated sequencing stock at -80°C.

Buffers: Add protease inhibitors to each buffer just prior to use. Make fresh at the start of experiments.

Cell Lysis:	Component	Volume/10mL	Nuclei Lysis:	Component	Volume/10 mL
	0.5M PIPES	100 μ L		1M Tris pH 8	500 μ L
	1M KCl	850 μ L		0.5 M EDTA	200 μ L
	10 % NP40	500 μ L		10% SDS	1 mL
	Nuclease-free dH ₂ O	8.55 mL		Nuclease-free dH ₂ O	8.3 mL

IP Buffer:	Component	Volume/10mL	LiCl Buffer:	Component	Volume/10 mL
	10% SDS	10 μ L		1M Tris pH 7.5	1 mL
	0.5 M EDTA	24 μ L		7.5M LiCl	666 μ L
	1M Tris-HCl, pH8	167 μ L		10% NP40	1 mL
	Triton x-100	110 μ L		20% DOC*	500 μ L
	Nuclease-free dH ₂ O	9.12 mL		Nuclease-free dH ₂ O	6.83 mL

RIPA Buffer:	Component	Volume/10mL	LoTE Buffer:	Component	Volume/10 mL
	10x PBS	1 mL		1 M Tris-HCl, pH 8	100 μ L
	10% NP40	1 mL		0.5 M EDTA	4 μ L
	20% DOC	100 μ L	TE Buffer:	Component	Volume/10mL

10% SDS	100 μ L	1 M Tris-HCl, pH 8	100 μ L
Nuclease-free dH ₂ O	7.8 mL	0.5 M EDTA	20 μ L

Micrococcal Nuclease Buffer: 10x Stock comes with nuclease. Dilute in Nuclease-free dH₂O. However, the buffer runs out before the enzyme does. 1xBuffer is 50 mM Tris HCl, 5mM CaCl₂, pH 7.9

* 20% DOC can be made ahead and stored in the fridge for several months, HOWEVER, it must be completely dissolved back into solution prior to use in buffers. Heat tube in a 37 °C water bath for an hour ahead of making buffers.

Abbreviations:

PBS= Phosphate Buffered Saline

NaAc= Sodium Acetate

EtOH= Ethanol

BSA= Bovine Serum Albumin

DOC= Sodium Deoxycholate

References

- Adamska M, Matus DQ, Adamski M, Green K, Rokhsar DS, Martindale MQ, Degnan BM. 2007. The evolutionary origin of hedgehog proteins. *Curr. Biol. CB* 17:R836–837.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403–410.
- Ando H, Kobayashi M, Tsubokawa T, Uyemura K, Furuta T, Okamoto H. 2005. Lhx2 mediates the activity of Six3 in zebrafish forebrain growth. *Dev. Biol.* 287:456–468.
- Andoniadou CL, Signore M, Young RM, Gaston-Massuet C, Wilson SW, Fuchs E, Martinez-Barbera JP. 2011. HESX1- and TCF3-mediated repression of Wnt/ β -catenin targets is required for normal development of the anterior forebrain. *Dev. Camb. Engl.* 138:4931–4942.
- Arnold SJ, Huang G-J, Cheung AFP, Era T, Nishikawa S-I, Bikoff EK, Molnár Z, Robertson EJ, Groszer M. 2008. The T-box transcription factor Eomes/Tbr2 regulates neurogenesis in the cortical subventricular zone. *Genes Dev.* 22:2479–2484.
- Arnone MI, Davidson EH. 1997. The hardwiring of development: organization and function of genomic regulatory systems. *Dev. Camb. Engl.* 124:1851–1864.
- Arnoult L, Su KFY, Manoel D, Minervino C, Magriña J, Gompel N, Prud'homme B. 2013. Emergence and Diversification of Fly Pigmentation Through Evolution of a Gene Regulatory Module. *Science* 339:1423–1426.
- Badis G, Berger MF, Philippakis AA, et al. 2009. Diversity and complexity in DNA recognition by transcription factors. *Science* 324:1720–1723.
- Baker CR, Tuch BB, Johnson AD. 2011. Extensive DNA-binding specificity divergence of a conserved transcription regulator. *Proc. Natl. Acad. Sci. U. S. A.* 108:7493–7498.
- Berger MF, Badis G, Gehrke AR, et al. 2008. Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell* 133:1266–1276.
- Berger MF, Bulyk ML. 2006. Protein binding microarrays (PBMs) for rapid, high-throughput characterization of the sequence specificities of DNA binding proteins. *Methods Mol. Biol. Clifton NJ* 338:245–260.
- Berger MF, Bulyk ML. 2009. Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. *Nat. Protoc.* 4:393–411.
- Berger MF, Philippakis AA, Qureshi AM, He FS, Estep PW, Bulyk ML. 2006. Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nat. Biotechnol.* 24:1429–1435.
- Bergsland M, Ramsköld D, Zaouter C, Klum S, Sandberg R, Muhr J. 2011. Sequentially acting Sox transcription factors in neural lineage development. *Genes Dev.* 25:2453–2464.
- Bertuzzi S, Porter FD, Pitts A, Kumar M, Agulnick A, Wassif C, Westphal H. 1999. Characterization of Lhx9, a novel LIM/homeobox gene expressed by the pioneer neurons in the mouse cerebral cortex. *Mech. Dev.* 81:193–198.

- Biehs B, François V, Bier E. 1996. The *Drosophila* short gastrulation gene prevents Dpp from autoactivating and suppressing neurogenesis in the neuroectoderm. *Genes Dev.* 10:2922–2934.
- Bisgrove BW, Burke RD. 1986. Development of Serotonergic Neurons in Embryos of the Sea Urchin, *Strongylocentrotus purpuratus*. *Dev. Growth Differ.* 28:569–574.
- Bishop CD, MacNeil KEA, Patel D, Taylor VJ, Burke RD. 2013. Neural development in *Eucidaris tribuloides* and the evolutionary history of the echinoid larval nervous system. *Dev. Biol.* 377:236–244.
- Blekhman R, Marioni JC, Zumbo P, Stephens M, Gilad Y. 2010. Sex-specific and lineage-specific alternative splicing in primates. *Genome Res.* 20:180–189.
- Bourbon HM, Martin-Blanco E, Rosen D, Kornberg TB. 1995. Phosphorylation of the *Drosophila* engrailed protein at a site outside its homeodomain enhances DNA binding. *J. Biol. Chem.* 270:11130–11139.
- Brayer KJ, Lynch VJ, Wagner GP. 2011. Evolution of a derived protein-protein interaction between HoxA11 and Foxo1a in mammals caused by changes in intramolecular regulation. *Proc. Natl. Acad. Sci. U. S. A.* 108:E414–420.
- Britten RJ, Davidson EH. 1971. Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *Q. Rev. Biol.* 46:111–138.
- Bulchand S, Grove EA, Porter FD, Tole S. 2001. LIM-homeodomain gene *Lhx2* regulates the formation of the cortical hem. *Mech. Dev.* 100:165–175.
- Bulfone A, Martinez S, Marigo V, Campanella M, Basile A, Quaderi N, Gattuso C, Rubenstein JLR, Ballabio A. 1999. Expression pattern of the *Tbr2* (Eomesodermin) gene during mouse and chick brain development. *Mech. Dev.* 84:133–138.
- Busser BW, Shokri L, Jaeger SA, Gisselbrecht SS, Singhanian A, Berger MF, Zhou B, Bulyk ML, Michelson AM. 2012. Molecular mechanism underlying the regulatory specificity of a *Drosophila* homeodomain protein that specifies myoblast identity. *Dev. Camb. Engl.* 139:1164–1174.
- Bustamante CD, Fledel-Alon A, Williamson S, et al. 2005. Natural selection on protein-coding genes in the human genome. *Nature* 437:1153–1157.
- Byrne M, Nakajima Y, Chee FC, Burke RD. 2007. Apical organs in echinoderm larvae: insights into larval evolution in the Ambulacraria. *Evol. Dev.* 9:432–445.
- Byrne M, Sewell MA, Selvakumaraswamy P, Prowse TAA. 2006. The Larval Apical Organ in the Holothuroid *Chiridota gigas* (Apodida): Inferences on Evolution of the Ambulacrarian Larval Nervous System. *Biol. Bull.* 211:95–100.
- Cameron RA, Samanta M, Yuan A, He D, Davidson E. 2009. SpBase: the sea urchin genome database and web site. *Nucleic Acids Res.* 37:D750–D754.
- Carroll SB. 2005. *Endless Forms Most Beautiful: The New Science of Evo Devo and the Making of the Animal Kingdom*. W. W. Norton & Company
- Carroll SB. 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* 134:25–36.
- Cheatle Jarvela AM, Brubaker L, Vedenko A, Gupta A, Armitage BA, Bulyk ML, Hinman VF. 2014. Modular Evolution of DNA Binding Preference of a Tbrain Transcription Factor Provides a Mechanism for Modifying Gene Regulatory Networks. *Mol. Biol. Evol.*

- Chee F, Byrne M. 1999. Development of the Larval Serotonergic Nervous System in the Sea Star *Patiriella regularis* as Revealed by Confocal Imaging. *Biol. Bull.* 197:123–131.
- Chen L, Zhao P, Wells L, Amemiya CT, Condie BG, Manley NR. 2010. Mouse and zebrafish *Hoxa3* orthologues have nonequivalent in vivo protein function. *Proc. Natl. Acad. Sci. U. S. A.* 107:10555–10560.
- Chou S-J, O’Leary DDM. 2013. Role for *Lhx2* in corticogenesis through regulation of progenitor differentiation. *Mol. Cell. Neurosci.* 56:1–9.
- Chung H-R, Löhr U, Jäckle H. 2007. Lineage-specific expansion of the zinc finger associated domain ZAD. *Mol. Biol. Evol.* 24:1934–1943.
- Conlon FL, Fairclough L, Price BM, Casey ES, Smith JC. 2001. Determinants of T box protein specificity. *Dev. Camb. Engl.* 128:3749–3758.
- Corsinotti A, Kapopoulou A, Gubelmann C, Imbeault M, Santoni de Sio FR, Rowe HM, Mouscaz Y, Deplancke B, Trono D. 2013. Global and stage specific patterns of Krüppel-associated-box zinc finger protein gene expression in murine early embryonic cells. *PLoS One* 8:e56721.
- Croce J, Lhomond G, Lozano JC, Gache C. 2001. *ske-T*, a T-box gene expressed in the skeletogenic mesenchyme lineage of the sea urchin embryo. *Mech. Dev.* 107:159–162.
- Croll RP. 2006. Development of embryonic and larval cells containing serotonin, catecholamines, and FMRFamide-related peptides in the gastropod mollusc *Phestilla sibogae*. *Biol. Bull.* 211:232–247.
- Daburon V, Mella S, Plouhinec J-L, Mazan S, Crozatier M, Vincent A. 2008. The metazoan history of the COE transcription factors. Selection of a variant HLH motif by mandatory inclusion of a duplicated exon in vertebrates. *BMC Evol. Biol.* 8:131.
- Davidson CJ, Tirouvanziam R, Herzenberg LA, Lipsick JS. 2005. Functional Evolution of the Vertebrate Myb Gene Family B-Myb, but Neither A-Myb nor c-Myb, Complements *Drosophila* Myb in Hemocytes. *Genetics* 169:215–229.
- Davidson EH. 2010. *The Regulatory Genome: Gene Regulatory Networks In Development And Evolution.* Academic Press
- Degnan BM, Vervoort M, Larroux C, Richards GS. 2009. Early evolution of metazoan transcription factors. *Curr. Opin. Genet. Dev.* 19:591–599.
- Duboule D, Dollé P. 1989. The structural and functional organization of the murine HOX gene family resembles that of *Drosophila* homeotic genes. *EMBO J.* 8:1497–1505.
- Enard W, Gehre S, Hammerschmidt K, et al. 2009. A humanized version of *Foxp2* affects cortico-basal ganglia circuits in mice. *Cell* 137:961–971.
- Enard W, Przeworski M, Fisher SE, Lai CSL, Wiebe V, Kitano T, Monaco AP, Pääbo S. 2002. Molecular evolution of FOXP2, a gene involved in speech and language. *Nature* 418:869–872.
- Faedo A, Ficara F, Ghiani M, Aiuti A, Rubenstein JLR, Bulfone A. 2002. Developmental expression of the T-box transcription factor T-bet/Tbx21 during mouse embryogenesis. *Mech. Dev.* 116:157–160.
- Fritzenwanker JH, Gerhart J, Freeman RM, Lowe CJ. 2014. The Fox/Forkhead transcription factor family of the hemichordate *Saccoglossus kowalevskii*. *EvoDevo* 5:17.

- Fuchikami T, Mitsunaga-Nakatsubo K, Amemiya S, et al. 2002. T-brain homologue (HpTb) is involved in the archenteron induction signals of micromere descendant cells in the sea urchin embryo. *Dev. Camb. Engl.* 129:5205–5216.
- Galant R, Carroll SB. 2002. Evolution of a transcriptional repression domain in an insect Hox protein. *Nature* 415:910–913.
- Gao Y, Lan Y, Ovitt CE, Jiang R. 2009. Functional equivalence of the zinc finger transcription factors Osr1 and Osr2 in mouse development. *Dev. Biol.* 328:200–209.
- Garrett-Engle CM, Siegal ML, Manoli DS, Williams BC, Li H, Baker BS. 2002. *intersex*, a gene required for female sexual development in *Drosophila*, is expressed in both sexes and functions together with *doublesex* to regulate terminal differentiation. *Dev. Camb. Engl.* 129:4661–4675.
- Gentsch GE, Owens NDL, Martin SR, Piccinelli P, Faial T, Trotter MWB, Gilchrist MJ, Smith JC. 2013. In Vivo T-Box Transcription Factor Profiling Reveals Joint Regulation of Embryonic Neuromesodermal Bipotency. *Cell Rep.* 4:1185–1196.
- Glinka A, Wu W, Onichtchouk D, Blumenstock C, Niehrs C. 1997. Head induction by simultaneous repression of Bmp and Wnt signalling in *Xenopus*. *Nature* 389:517–519.
- Gomez C, Ozbudak EM, Wunderlich J, Baumann D, Lewis J, Pourquié O. 2008. Control of segment number in vertebrate embryos. *Nature* 454:335–339.
- Gordân R, Murphy KF, McCord RP, Zhu C, Vedenko A, Bulyk ML. 2011. Curated collection of yeast transcription factor DNA binding specificity data reveals novel structural and gene regulatory insights. *Genome Biol.* 12:R125.
- Gordon PJ, Yun S, Clark AM, Monuki ES, Murtaugh LC, Levine EM. 2013. Lhx2 balances progenitor maintenance with neurogenic output and promotes competence state progression in the developing retina. *J. Neurosci. Off. J. Soc. Neurosci.* 33:12197–12207.
- Guerreiro I, Nunes A, Woltering JM, Casaca A, Nóvoa A, Vinagre T, Hunter ME, Duboule D, Mallo M. 2013. Role of a polymorphism in a Hox/Pax-responsive enhancer in the evolution of the vertebrate spine. *Proc. Natl. Acad. Sci.* 110:10682–10686.
- Häggglund A-C, Dahl L, Carlsson L. 2011. Lhx2 is required for patterning and expansion of a distinct progenitor cell population committed to eye development. *PLoS One* 6:e23387.
- Halder G, Callaerts P, Gehring WJ. 1995. Induction of ectopic eyes by targeted expression of the *eyeless* gene in *Drosophila*. *Science* 267:1788–1792.
- Hanes SD, Brent R. 1989. DNA specificity of the bicoid activator protein is determined by homeodomain recognition helix residue 9. *Cell* 57:1275–1283.
- Hanes SD, Riddihough G, Ish-Horowitz D, Brent R. 1994. Specific DNA recognition and intersite spacing are critical for action of the bicoid morphogen. *Mol. Cell. Biol.* 14:3364–3375.
- Hay-Schmidt A. 2000. The evolution of the serotonergic nervous system. *Proc. Biol. Sci.* 267:1071–1079.
- Hedges SB, Dudley J, Kumar S. 2006. TimeTree: a public knowledge-base of divergence times among organisms. *Bioinform. Oxf. Engl.* 22:2971–2972.
- Heffer A, Shultz JW, Pick L. 2010. Surprising flexibility in a conserved Hox transcription factor over 550 million years of evolution. *Proc. Natl. Acad. Sci. U. S. A.* 107:18040–18045.

- Hinman VF, Davidson EH. 2007. Evolutionary plasticity of developmental gene regulatory network architecture. *Proc. Natl. Acad. Sci. U. S. A.* 104:19404–19409.
- Hinman VF, Nguyen A, Davidson EH. 2007. Caught in the evolutionary act: precise cis-regulatory basis of difference in the organization of gene networks of sea stars and sea urchins. *Dev. Biol.* 312:584–595.
- Hinman VF, Nguyen AT, Cameron RA, Davidson EH. 2003. Developmental gene regulatory network architecture across 500 million years of echinoderm evolution. *Proc. Natl. Acad. Sci. U. S. A.* 100:13356–13361.
- Hinman VF, Nguyen AT, Davidson EH. 2003. Expression and function of a starfish Otx ortholog, AmOtx: a conserved role for Otx proteins in endoderm development that predates divergence of the eleutherozoa. *Mech. Dev.* 120:1165–1176.
- Hirabayashi Y, Itoh Y, Tabata H, Nakajima K, Akiyama T, Masuyama N, Gotoh Y. 2004. The Wnt/ β -catenin pathway directs neuronal differentiation of cortical neural precursor cells. *Development* 131:2791–2801.
- Hittinger CT, Carroll SB. 2008. Evolution of an insect-specific GROUCHO-interaction motif in the ENGRAILED selector protein. *Evol. Dev.* 10:537–545.
- Hoekstra HE, Coyne JA. 2007. The locus of evolution: evo devo and the genetics of adaptation. *Evol. Int. J. Org. Evol.* 61:995–1016.
- Holland LZ, Carvalho JE, Escrava H, Laudet V, Schubert M, Shimeld SM, Yu J-K. 2013. Evolution of bilaterian central nervous systems: a single origin? *EvoDevo* 4:27.
- Holland PWH. 2013. Evolution of homeobox genes. *Wiley Interdiscip. Rev. Dev. Biol.* 2:31–45.
- Horton AC, Gibson-Brown JJ. 2002. Evolution of developmental functions by the Eomesodermin, T-brain-1, Tbx21 subfamily of T-box genes: insights from amphioxus. *J. Exp. Zool.* 294:112–121.
- Hoser M, Potzner MR, Koch JMC, Bösl MR, Wegner M, Sock E. 2008. Sox12 deletion in the mouse reveals nonreciprocal redundancy with the related Sox4 and Sox11 transcription factors. *Mol. Cell. Biol.* 28:4675–4687.
- Houart C, Caneparo L, Heisenberg C, Barth K, Take-Uchi M, Wilson S. 2002. Establishment of the telencephalon during gastrulation by local antagonism of Wnt signaling. *Neuron* 35:255–265.
- Howard-Ashby M, Materna SC, Brown CT, Chen L, Cameron RA, Davidson EH. 2006. Gene families encoding transcription factors expressed in early development of *Strongylocentrotus purpuratus*. *Dev. Biol.* 300:90–107.
- Huntley S, Baggott DM, Hamilton AT, Tran-Gyamfi M, Yang S, Kim J, Gordon L, Branscomb E, Stubbs L. 2006. A comprehensive catalog of human KRAB-associated zinc finger genes: Insights into the evolutionary history of a large family of transcriptional repressors. *Genome Res.* 16:669–677.
- Jaffe L, Ryoo HD, Mann RS. 1997. A role for phosphorylation by casein kinase II in modulating Antennapedia activity in *Drosophila*. *Genes Dev.* 11:1327–1340.
- Jovelin R. 2009. Rapid sequence evolution of transcription factors controlling neuron differentiation in *Caenorhabditis*. *Mol. Biol. Evol.* 26:2373–2386.
- Kaestner KH, Knochel W, Martinez DE. 2000. Unified nomenclature for the winged helix/forkhead transcription factors. *Genes Dev.* 14:142–146.

- Kawashima T, Kawashima S, Tanaka C, et al. 2009. Domain shuffling and the evolution of vertebrates. *Genome Res.* 19:1393–1403.
- Kelley LA, Sternberg MJE. 2009. Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protoc.* 4:363–371.
- Kempf SC, Page LR, Pires A. 1997. Development of serotonin-like immunoreactivity in the embryos and larvae of nudibranch mollusks with emphasis on the structure and possible function of the apical sensory organ. *J. Comp. Neurol.* 386:507–528.
- Kersting AR, Bornberg-Bauer E, Moore AD, Grath S. 2012. Dynamics and Adaptive Benefits of Protein Domain Emergence and Arrangements during Plant Genome Evolution. *Genome Biol. Evol.* 4:316–329.
- Keyte A, Smith KK. 2012. Heterochrony in somitogenesis rate in a model marsupial, *Monodelphis domestica*. *Evol. Dev.* 14:93–103.
- Keyte AL, Smith KK. Heterochrony and developmental timing mechanisms: Changing ontogenies in evolution. *Semin. Cell Dev. Biol.* [Internet]. Available from: <http://www.sciencedirect.com/science/article/pii/S1084952114001888>
- Kim DS, Hahn Y. 2011. Identification of novel phosphorylation modification sites in human proteins that originated after the human-chimpanzee divergence. *Bioinforma. Oxf. Engl.* 27:2494–2501.
- Kim DS, Hahn Y. 2012. Gains of ubiquitylation sites in highly conserved proteins in the human lineage. *BMC Bioinformatics* 13:306.
- King MC, Wilson AC. 1975. Evolution at two levels in humans and chimpanzees. *Science* 188:107–116.
- Konopka G, Bomar JM, Winden K, et al. 2009. Human-specific transcriptional regulation of CNS development genes by FOXP2. *Nature* 462:213–217.
- Krebs CJ, Larkins LK, Khan SM, Robins DM. 2005. Expansion and diversification of KRAB zinc-finger genes within a cluster including Regulator of sex-limitation 1 and 2. *Genomics* 85:752–761.
- Lacalli TC. 1994. Apical Organs, Epithelial Domains, and the Origin of the Chordate Central Nervous System. *Am. Zool.* 34:533–541.
- Lagutin OV, Zhu CC, Kobayashi D, et al. 2003. Six3 repression of Wnt signaling in the anterior neuroectoderm is essential for vertebrate forebrain development. *Genes Dev.* 17:368–379.
- Lavado A, Lagutin OV, Oliver G. 2008. Six3 inactivation causes progressive caudalization and aberrant patterning of the mammalian diencephalon. *Dev. Camb. Engl.* 135:441–450.
- Lee H-L, Irish VF. 2011. Gene duplication and loss in a MADS box gene transcription factor circuit. *Mol. Biol. Evol.* 28:3367–3380.
- Levitt M. 2009. Nature of the protein universe. *Proc. Natl. Acad. Sci.* 106:11079–11084.
- Lillesaar C. 2011. The serotonergic system in fish. *J. Chem. Neuroanat.* 41:294–308.
- Liu H, Chang L-H, Sun Y, Lu X, Stubbs L. 2014. Deep vertebrate roots for mammalian zinc finger transcription factor subfamilies. *Genome Biol. Evol.* 6:510–525.
- Logan CY, Miller JR, Ferkowicz MJ, McClay DR. 1999. Nuclear beta-catenin is required to specify vegetal cell fates in the sea urchin embryo. *Development* 126:345–357.

- Löhr U, Pick L. 2005. Cofactor-interaction motifs and the cooption of a homeotic Hox protein into the segmentation pathway of *Drosophila melanogaster*. *Curr. Biol.* CB 15:643–649.
- Van Loosdregt J, Coffey PJ. 2014. Post-translational modification networks regulating FOXP3 function. *Trends Immunol.*
- Lowe CJ, Terasaki M, Wu M, et al. 2006. Dorsoventral Patterning in Hemichordates: Insights into Early Chordate Evolution. *PLoS Biol* 4:e291.
- Lynch VJ, Brayer K, Gellersen B, Wagner GP. 2009. HoxA-11 and FOXO1A cooperate to regulate decidual prolactin expression: towards inferring the core transcriptional regulators of decidual genes. *PLoS One* 4:e6845.
- Lynch VJ, May G, Wagner GP. 2011. Regulatory evolution through divergence of a phosphoswitch in the transcription factor CEBPB. *Nature* 480:383–386.
- Lynch VJ, Tanzer A, Wang Y, Leung FC, Gellersen B, Emera D, Wagner GP. 2008. Adaptive changes in the transcription factor HoxA-11 are essential for the evolution of pregnancy in mammals. *Proc. Natl. Acad. Sci. U. S. A.* 105:14928–14933.
- Lynch VJ, Wagner GP. 2008. Resurrecting the role of transcription factor change in developmental evolution. *Evol. Int. J. Org. Evol.* 62:2131–2154.
- Macindoe I, Glockner L, Vukasin P, Stennard FA, Costa MW, Harvey RP, Mackay JP, Sunde M. 2009. Conformational stability and DNA binding specificity of the cardiac T-box transcription factor Tbx20. *J. Mol. Biol.* 389:606–618.
- Mangale VS, Hirokawa KE, Satyaki PRV, et al. 2008. Lhx2 Selector Activity Specifies Cortical Identity and Suppresses Hippocampal Organizer Fate. *Science* 319:304–309.
- Maricic T, Günther V, Georgiev O, et al. 2013. A recent evolutionary change affects a regulatory element in the human FOXP2 gene. *Mol. Biol. Evol.* 30:844–852.
- Marlow H, Tosches MA, Tomer R, Steinmetz PR, Lauri A, Larsson T, Arendt D. 2014. Larval body patterning and apical organs are conserved in animal evolution. *BMC Biol.* 12:7.
- Mazet F, Yu J-K, Liberles DA, Holland LZ, Shimeld SM. 2003. Phylogenetic relationships of the Fox (Forkhead) gene family in the Bilateria. *Gene* 316:79–89.
- McCauley BS, Weideman EP, Hinman VF. 2010. A conserved gene regulatory network subcircuit drives different developmental fates in the vegetal pole of highly divergent echinoderm embryos. *Dev. Biol.* 340:200–208.
- McCauley BS, Wright EP, Exner C, Kitazawa C, Hinman VF. 2012. Development of an embryonic skeletogenic mesenchyme lineage in a sea cucumber reveals the trajectory of change for the evolution of novel structures in echinoderms. *EvoDevo* 3:17.
- McGinnis N, Kuziora MA, McGinnis W. 1990. Human Hox-4.2 and *Drosophila* deformed encode similar regulatory specificities in *Drosophila* embryos and larvae. *Cell* 63:969–976.
- McGinnis W, Garber RL, Wirz J, Kuroiwa A, Gehring WJ. 1984. A homologous protein-coding sequence in *Drosophila* homeotic genes and its conservation in other metazoans. *Cell* 37:403–408.
- Meador S, Ponting CP, Lunter G. 2010. Massive turnover of functional sequence in human and other mammalian genomes. *Genome Res.* 20:1335–1343.

- De Mendoza A, Sebé-Pedrós A, Šestak MS, Matejčić M, Torruella G, Domazet-Lošo T, Ruiz-Trillo I. 2013. Transcription factor evolution in eukaryotes and the assembly of the regulatory toolkit in multicellular lineages. *Proc. Natl. Acad. Sci. U. S. A.* 110:E4858–4866.
- Milne I, Wright F, Rowe G, Marshall DF, Husmeier D, McGuire G. 2004. TOPALi: software for automatic identification of recombinant sequences within DNA multiple alignments. *Bioinforma. Oxf. Engl.* 20:1806–1807.
- Minemura K, Yamaguchi M, Minokawa T. 2009. Evolutionary modification of T-brain (*tbr*) expression patterns in sand dollar. *Gene Expr. Patterns GEP* 9:468–474.
- MONOD J, JACOB F. 1961. Teleonomic mechanisms in cellular metabolism, growth, and differentiation. *Cold Spring Harb. Symp. Quant. Biol.* 26:389–401.
- Monuki ES, Porter FD, Walsh CA. 2001. Patterning of the Dorsal Telencephalon and Cerebral Cortex by a Roof Plate-Lhx2 Pathway. *Neuron* 32:591–604.
- Mortazavi A, Thompson ECL, Garcia ST, Myers RM, Wold B. 2006. Comparative genomics modeling of the NRSF/REST repressor network: From single conserved sites to genome-wide repertoire. *Genome Res.* 16:1208–1221.
- Müller CW, Herrmann BG. 1997. Crystallographic structure of the T domain-DNA complex of the Brachyury transcription factor. *Nature* 389:884–888.
- Munji RN, Choe Y, Li G, Siegenthaler JA, Pleasure SJ. 2011. Wnt Signaling Regulates Neuronal Differentiation of Cortical Intermediate Progenitors. *J. Neurosci.* 31:1676–1687.
- Nakagawa S, Gisselbrecht SS, Rogers JM, Hartl DL, Bulyk ML. 2013a. DNA-binding specificity changes in the evolution of forkhead transcription factors. *Proc. Natl. Acad. Sci.* 110:12349–12354.
- Nakagawa S, Gisselbrecht SS, Rogers JM, Hartl DL, Bulyk ML. 2013b. DNA-binding specificity changes in the evolution of forkhead transcription factors. *Proc. Natl. Acad. Sci.* 110:12349–12354.
- Nakajima Y, Humphreys T, Kaneko H, Tagawa K. 2004. Development and neural organization of the tornaria larva of the Hawaiian hemichordate, *Ptychodera flava*. *Zoolog. Sci.* 21:69–78.
- Nakajima Y, Kaneko H, Murray G, Burke RD. 2004. Divergent patterns of neural development in larval echinoids and asteroids. *Evol. Dev.* 6:95–104.
- Nelson CS, Fuller CK, Fordyce PM, Greninger AL, Li H, DeRisi JL. 2013. Microfluidic affinity and ChIP-seq analyses converge on a conserved FOXP2-binding motif in chimp and human, which enables the detection of evolutionarily novel targets. *Nucleic Acids Res.* 41:5991–6004.
- Nguyen B, Stanek J, Wilson WD. 2006. Binding-linked protonation of a DNA minor-groove agent. *Biophys. J.* 90:1319–1328.
- Nielsen C, Hay-Schmidt A. 2007. Development of the enteropneust *Ptychodera flava*: ciliary bands and nervous system. *J. Morphol.* 268:551–570.
- Nordström U, Jessell TM, Edlund T. 2002. Progressive induction of caudal neural character by graded Wnt signaling. *Nat. Neurosci.* 5:525–532.
- Nowick K, Fields C, Gernat T, Caetano-Anolles D, Kholina N, Stubbs L. 2011. Gain, loss and divergence in primate zinc-finger genes: a rich resource for evolution of gene regulatory differences between species. *PLoS One* 6:e21553.

- Nowick K, Gernat T, Almaas E, Stubbs L. 2009. Differences in human and chimpanzee gene expression patterns define an evolving network of transcription factors in brain. *Proc. Natl. Acad. Sci.* 106:22358–22363.
- Nowick K, Hamilton AT, Zhang H, Stubbs L. 2010. Rapid sequence and expression divergence suggest selection for novel function in primate-specific KRAB-ZNF genes. *Mol. Biol. Evol.* 27:2606–2617.
- Nowick K, Stubbs L. 2010. Lineage-specific transcription factors and the evolution of gene regulatory networks. *Brief. Funct. Genomics* 9:65–78.
- Nübler-Jung K, Arendt D. 1996. Enteropneusts and chordate evolution. *Curr. Biol.* 6:352–353.
- Oliveri P, Carrick DM, Davidson EH. 2002. A regulatory gene network that directs micromere specification in the sea urchin embryo. *Dev. Biol.* 246:209–228.
- Pani AM, Mullarkey EE, Aronowicz J, Assimacopoulos S, Grove EA, Lowe CJ. 2012. Ancient deuterostome origins of vertebrate brain signalling centres. *Nature* 483:289–294.
- Papaiouannou VE, Silver LM. 1998. The T-box gene family. *BioEssays News Rev. Mol. Cell. Dev. Biol.* 20:9–19.
- Parker DS, White MA, Ramos AI, Cohen BA, Barolo S. 2011. The cis-regulatory logic of Hedgehog gradient responses: key roles for gli binding affinity, competition, and cooperativity. *Sci. Signal.* 4:ra38.
- Pérez JC, Fordyce PM, Lohse MB, Hanson-Smith V, DeRisi JL, Johnson AD. 2014. How duplicated transcription regulators can diversify to govern the expression of nonoverlapping sets of genes. *Genes Dev.* 28:1272–1277.
- Peterson KA, Nishi Y, Ma W, et al. 2012. Neural-specific Sox2 input and differential Gli-binding affinity provide context and positional information in Shh-directed neural patterning. *Genes Dev.* 26:2802–2816.
- Peukert D, Weber S, Lumsden A, Scholpp S. 2011. Lhx2 and Lhx9 determine neuronal differentiation and compartment in the caudal forebrain by regulating Wnt signaling. *PLoS Biol.* 9:e1001218.
- Pisani D, Feuda R, Peterson KJ, Smith AB. 2012. Resolving phylogenetic signal from noise when divergence is rapid: a new look at the old problem of echinoderm class relationships. *Mol. Phylogenet. Evol.* 62:27–34.
- Plaitakis A, Spanaki C, Mastorodemos V, Zaganas I. 2003. Study of structure-function relationships in human glutamate dehydrogenases reveals novel molecular mechanisms for the regulation of the nerve tissue-specific (GLUD2) isoenzyme. *Neurochem. Int.* 43:401–410.
- Pocock R, Mione M, Hussain S, Maxwell S, Pontecorvi M, Aslam S, Gerrelli D, Sowden JC, Woollard A. 2008. Neuronal function of Tbx20 conserved from nematodes to vertebrates. *Dev. Biol.* 317:671–685.
- Porter FD, Drago J, Xu Y, et al. 1997. Lhx2, a LIM homeobox gene, is required for eye, forebrain, and definitive erythrocyte development. *Dev. Camb. Engl.* 124:2935–2944.
- Prasad MS, Sauka-Spengler T, LaBonne C. 2012. Induction of the neural crest state: Control of stem cell attributes by gene regulatory, post-transcriptional and epigenetic interactions. *Dev. Biol.* 366:10–21.
- Prud'homme B, Gompel N, Carroll SB. 2007. Emerging principles of regulatory evolution. *Proc. Natl. Acad. Sci. U. S. A.* 104 Suppl 1:8605–8612.
- Quenneville S, Turelli P, Bojkowska K, Raclot C, Offner S, Kapopoulou A, Trono D. 2012. The KRAB-ZFP/KAP1 system contributes to the early embryonic establishment of site-specific DNA methylation patterns maintained during development. *Cell Rep.* 2:766–773.

- Rafiq K, Cheers MS, Etensohn CA. 2012. The genomic regulatory control of skeletal morphogenesis in the sea urchin. *Dev. Camb. Engl.* 139:579–590.
- Ramos AI, Barolo S. 2013. Low-affinity transcription factor binding sites shape morphogen responses and enhancer evolution. *Philos. Trans. R. Soc. B Biol. Sci.* 368:20130018.
- Range R. 2014. Specification and positioning of the anterior neuroectoderm in deuterostome embryos. *Genes.* N. Y. N 2000 52:222–234.
- Range RC, Angerer RC, Angerer LM. 2013. Integration of Canonical and Noncanonical Wnt Signaling Pathways Patterns the Neuroectoderm Along the Anterior-Posterior Axis of Sea Urchin Embryos. *PLoS Biol.* [Internet] 11. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3545869/>
- Rebeiz M, Williams TM. 2011. Experimental approaches to evaluate the contributions of candidate cis-regulatory mutations to phenotypic evolution. *Methods Mol. Biol.* Clifton NJ 772:351–375.
- Rétaux S, Rogard M, Bach I, Failli V, Besson M-J. 1999. Lhx9: A Novel LIM-Homeodomain Gene Expressed in the Developing Forebrain. *J. Neurosci.* 19:783–793.
- De Robertis EM, Kuroda H. 2004. Dorsal-ventral patterning and neural induction in *Xenopus* embryos. *Annu. Rev. Cell Dev. Biol.* 20:285–308.
- Ronshaugen M, McGinnis N, McGinnis W. 2002. Hox protein mutation and macroevolution of the insect body plan. *Nature* 415:914–917.
- Rowan S, Siggers T, Lachke SA, Yue Y, Bulyk ML, Maas RL. 2010. Precise temporal control of the eye regulatory gene Pax6 via enhancer-binding site affinity. *Genes Dev.* 24:980–985.
- Rubinstein M, de Souza FSJ. 2013. Evolution of transcriptional enhancers and animal diversity. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 368:20130017.
- Ryan K, Butler K, Bellefroid E, Gurdon JB. 1998. *Xenopus* eomesodermin is expressed in neural differentiation. *Mech. Dev.* 75:155–158.
- Sakabe NJ, Aneas I, Shen T, Shokri L, Park S-Y, Bulyk ML, Evans SM, Nobrega MA. 2012. Dual transcriptional activator and repressor roles of TBX20 regulate adult cardiac structure and function. *Hum. Mol. Genet.* 21:2194–2204.
- Santagata S, Resh C, Hejnal A, Martindale MQ, Passamanek YJ. 2012. Development of the larval anterior neurogenic domains of *Terebratalia transversa* (Brachiopoda) provides insights into the diversification of larval apical organs and the spiralian nervous system. *EvoDevo* 3:3.
- Saudemont A, Haillet E, Mekpoh F, et al. 2010. Ancestral regulatory circuits governing ectoderm patterning downstream of Nodal and BMP2/4 revealed by gene regulatory network analysis in an echinoderm. *PLoS Genet.* 6:e1001259.
- Schmidt D, Wilson MD, Ballester B, et al. 2010. Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science* 328:1036–1040.
- Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9:671–675.
- Schnitzler CE, Simmons DK, Pang K, Martindale MQ, Baxevanis AD. 2014. Expression of multiple Sox genes through embryonic development in the ctenophore *Mnemiopsis leidyi* is spatially restricted to zones of cell proliferation. *EvoDevo* 5:15.

- Sebé-Pedrós A, Mendoza A de, Lang BF, Degnan BM, Ruiz-Trillo I. 2011. Unexpected Repertoire of Metazoan Transcription Factors in the Unicellular Holozoan *Capsaspora owczarzaki*. *Mol. Biol. Evol.* 28:1241–1254.
- Sessa A, Mao C-A, Hadjantonakis A-K, Klein WH, Broccoli V. 2008. *Tbr2* directs conversion of radial glia into basal precursors and guides neuronal amplification by indirect neurogenesis in the developing neocortex. *Neuron* 60:56–69.
- Shoguchi E, Satoh N, Maruyama YK. 2000. A starfish homolog of mouse *T-brain-1* is expressed in the archenteron of *Asterina pectinifera* embryos: possible involvement of two *T-box* genes in starfish gastrulation. *Dev. Growth Differ.* 42:61–68.
- Showell C, Binder O, Conlon FL. 2004. *T-box* genes in early embryogenesis. *Dev. Dyn. Off. Publ. Am. Assoc. Anat.* 229:201–218.
- Sievers F, Wilm A, Dineen D, et al. 2014. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7:539–539.
- Siggers T, Gordân R. 2014. Protein-DNA binding: complexities and multi-protein codes. *Nucleic Acids Res.* 42:2099–2111.
- Siggers T, Reddy J, Barron B, Bulyk ML. Diversification of Transcription Factor Paralogs via Noncanonical Modularity in C2H2 Zinc Finger DNA Binding. *Mol. Cell* [Internet]. Available from: <http://www.sciencedirect.com/science/article/pii/S1097276514005279>
- Sinigaglia C, Busengdal H, Leclère L, Technau U, Rentzsch F. 2013. The bilaterian head patterning gene *six3/6* controls aboral domain development in a cnidarian. *PLoS Biol.* 11:e1001488.
- Smith KK. 2003. Time's arrow: heterochrony and the evolution of development. *Int. J. Dev. Biol.* 47:613–621.
- Smith ST, Jaynes JB. 1996. A conserved region of engrailed, shared among all *en-*, *gsc-*, *Nk1-*, *Nk2-* and *msh-class* homeoproteins, mediates active transcriptional repression in vivo. *Dev. Camb. Engl.* 122:3141–3150.
- Srivastava M, Simakov O, Chapman J, et al. 2010. The Amphimedon *queenslandica* genome and the evolution of animal complexity. *Nature* 466:720–726.
- Steinmetz PRH, Urbach R, Posnien N, et al. 2010. *Six3* demarcates the anterior-most developing brain region in bilaterian animals. *EvoDevo* 1:14.
- Stern DL. 2000. Evolutionary developmental biology and the problem of variation. *Evol. Int. J. Org. Evol.* 54:1079–1091.
- Tagawa K, Satoh N, Humphreys T. 2001. Molecular studies of hemichordate development: a key to understanding the evolution of bilateral animals and chordates. *Evol. Dev.* 3:443–454.
- Taghli-Lamalle O, Hsia C, Ronshaugen M, McGinnis W. 2008. Context-dependent regulation of Hox protein functions by CK2 phosphorylation sites. *Dev. Genes Evol.* 218:321–332.
- Taneri B, Snyder B, Novoradovsky A, Gaasterland T. 2004. Alternative splicing of mouse transcription factors affects their DNA-binding domain architecture and is tissue specific. *Genome Biol.* 5:R75.
- Teichmann SA, Babu MM. 2004. Gene regulatory network growth by duplication. *Nat. Genet.* 36:492–496.
- Tolkunova EN, Fujioka M, Kobayashi M, Deka D, Jaynes JB. 1998. Two distinct types of repression domain in engrailed: one interacts with the groucho corepressor and is preferentially active on integrated target genes. *Mol. Cell. Biol.* 18:2804–2814.

- Tosches MA, Arendt D. 2013. The bilaterian forebrain: an evolutionary chimaera. *Curr. Opin. Neurobiol.* 23:1080–1089.
- Tsui D, Vessey JP, Tomita H, Kaplan DR, Miller FD. 2013. FoxP2 regulates neurogenesis during embryonic cortical development. *J. Neurosci. Off. J. Soc. Neurosci.* 33:244–258.
- Tuch BB, Galgoczy DJ, Hernday AD, Li H, Johnson AD. 2008. The evolution of combinatorial gene regulation in fungi. *PLoS Biol.* 6:e38.
- Wang M, Caetano-Anollés G. 2009. The evolutionary mechanics of domain organization in proteomes and the rise of modularity in the protein world. *Struct. Lond. Engl.* 1993 17:66–78.
- Wang VY, Hassan BA, Bellen HJ, Zoghbi HY. 2002. *Drosophila* atonal fully rescues the phenotype of Math1 null mice: new functions evolve in new cellular contexts. *Curr. Biol. CB* 12:1611–1616.
- Wang W, Grimmer JF, Van De Water TR, Lufkin T. 2004. Hmx2 and Hmx3 homeobox genes direct development of the murine inner ear and hypothalamus and can be functionally replaced by *Drosophila* Hmx. *Dev. Cell* 7:439–453.
- Wang Y, Lin L, Lai H, Parada LF, Lei L. 2013. Transcription factor Sox11 is essential for both embryonic and adult neurogenesis. *Dev. Dyn. Off. Publ. Am. Assoc. Anat.* 242:638–653.
- Wei Z, Yaguchi J, Yaguchi S, Angerer RC, Angerer LM. 2009. The sea urchin animal pole domain is a Six3-dependent neurogenic patterning center. *Dev. Camb. Engl.* 136:1179–1189.
- Wikramanayake AH, Huang L, Klein WH. 1998. β -Catenin is essential for patterning the maternally specified animal-vegetal axis in the sea urchin embryo. *Proc. Natl. Acad. Sci.* 95:9343–9348.
- Williams TM, Selegue JE, Werner T, Gompel N, Kopp A, Carroll SB. 2008. The Regulation and Evolution of a Genetic Switch Controlling Sexually Dimorphic Traits in *Drosophila*. *Cell* 134:610–623.
- Wittkopp PJ, Kalay G. 2012. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat. Rev. Genet.* 13:59–69.
- Wray GA. 2007. The evolutionary significance of cis-regulatory mutations. *Nat. Rev. Genet.* 8:206–216.
- Yaguchi S, Yaguchi J, Angerer RC, Angerer LM. 2008. A Wnt-FoxQ2-nodal pathway links primary and secondary axis specification in sea urchin embryos. *Dev. Cell* 14:97–107.
- Yankura KA, Koechlein CS, Cryan AF, Cheadle A, Hinman VF. 2013. Gene regulatory network for neurogenesis in a sea star embryo connects broad neural specification and localized patterning. *Proc. Natl. Acad. Sci.* 110:8591–8596.
- Yankura KA, Martik ML, Jennings CK, Hinman VF. 2010. Uncoupling of complex regulatory patterning during evolution of larval development in echinoderms. *BMC Biol.* 8:143.
- Yu J-K, Holland ND, Holland LZ. 2003. AmphifoxQ2, a novel winged helix/forkhead gene, exclusively marks the anterior end of the amphioxus embryo. *Dev. Genes Evol.* 213:102–105.
- Zhang J, Dean AM, Brunet F, Long M. 2004. Evolving protein functional diversity in new genes of *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* 101:16246–16250.
- Zhang Y, Liu T, Meyer CA, et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9:R137.

Zhu C, Byers KJRP, McCord RP, et al. 2009. High-resolution DNA-binding specificity analysis of yeast transcription factors. *Genome Res.* 19:556–566.

Zhu X, Ahmad SM, Aboukhalil A, et al. 2012. Differential regulation of mesodermal gene expression by *Drosophila* cell type-specific Forkhead transcription factors. *Dev. Camb. Engl.* 139:1457–1466.