COMPETITION BETWEEN T-BOX TRANSCRIPTION FACTORS CONTRIBUTES TO DEVELOPMENTAL DYNAMICS OF THE MOUSE EMBRYO

by

Amy Kristen Wehn

Bachelor of Science, University of Pittsburgh, 2004

Submitted to the Graduate Faculty of

College of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2010

UNIVERSITY OF PITTSBURGH

COLLEGE OF ARTS AND SCIENCES

This thesis was presented

by

Amy Kristen Wehn

It was defended on

June 7, 2010

and approved by

Gerard Campbell, Ph.D. Associate Professor, Biological Sciences

Paula Grabowski, Ph.D. Professor, Biological Sciences

Jeffery Hildebrand, Ph.D. Associate Professor, Biological Sciences

Michael Tsang, Ph.D. Assistant Professor, Biochemistry and Molecular Genetics

Thesis Advisor: Deborah Chapman, Ph.D. Associate Professor, Biological Sciences

COMPETITION BETWEEN T-BOX TRANSCRIPTION FACTORS CONTRIBUTES TO DEVELOPMENTAL DYNAMICS OF THE MOUSE EMBRYO

Amy Kristen Wehn, Ph.D.

University of Pittsburgh, 2010

Tbx6 is a member of the T-box family of transcription factors that is expressed within the primitive streak and presomitic mesoderm of the developing mouse embryo and is critical for the patterning and specification of the paraxial mesoderm. This work has investigated how Tbx6 interacts with other T-box transcription factors both in its endogenous expression domain and when it is ectopically expressed. To this end, we have used a combined genetic, transcriptional, and biochemical approach to investigate potential competition between Tbx6 and T, a T-box transcription factor co-expressed with Tbx6 in the primitive streak and tailbud. Additionally, we have developed a 3-component transgenic system to ectopically express Tbx6 outside of its endogenous domain, driving expression in the formed somites and limb buds. Three-component embryos displayed vertebral, rib, and limb anomalies resembling those in *Tbx15* and *Tbx18* null embryos. We showed that ectopic Tbx6 in these embryos could compete with Tbx15 and Tbx18 at the level of DNA binding. This suggested that the dynamic interplay of co-expressed T-box transcription factors might contribute to observed congenital birth defects resulting from heterozygous loss of a particular family member. While this work focused on one specific family of transcription factors, it implies that transcription factor family members that are related via a conserved DNA binding domain may also compete for downstream targets *in vivo* and that this then contributes to the overall developmental dynamics of the organism.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

First and foremost, this thesis wouldn't be possible without the support and guidance of my advisor- Debbie Chapman. I didn't think I'd find anyone to put up with my crazy energy when I started here- but I found not only someone that accepted that energy, but also gave me a run for my money! Your enthusiasm for science and for life in general was certainly encouraging for me throughout the years- you have been a role model in more ways than one. I was only half-joking when I said that I was going to require my post-doc lab to have someone that bakes/cooks as well as you do. I don't think I can ever again eat a cake from a grocery store bakery without turning my nose up- you've spoiled me for life!

I also need to give a special thank you to my committee members- Gerard Campbell, Paula Grabowski, Jeff Hildebrand, and Michael Tsang. You have been instrumental in helping me to elevate my scientific thinking to a more global level, as well as giving great suggestions in committee meetings. Thank you very much for all of your support and guidance in taking the next step as a post-doc at UNC.

My experience here at Pitt wouldn't have been complete without my partner in crime- Debby Farkas. It was wonderful to have someone else the same year as me in the same lab- we went through the whole deal together—core course, comps, noon seminars, and writing theses! It was great to be able to share that experience with someone else, and I'm surely going to miss you when we move on. Not only will I miss having someone to bounce science ideas off of (that

understands my insanity), but also your friendship. Who else will know not to bother to say, "bless you" the first time I sneeze… because there's always more than one sneeze so you might as well just wait? Or trade puns with me all day long? Or put up with me asking "WHAT?" for the fortieth time when I can't hear a thing? Best of luck in Cecelia's lab and in life, I know you'll do great. You better keep in touch!

I also need to thank other grad students that have been not only wonderful people to discuss science with, but also great friends along the way- especially Jill Dembowski, Elia Crisucci, and Matt Farber. You guys have been great to discuss science and life with. I know you will all move on to do great things- and you guys better keep in touch too!

Then there's the class of people that had little to with the science, but everything to do with keeping me sane, healthy, and most of all, happy for the past six years. I have made some great friends in Pittsburgh that I am surely going to miss. To my "book club girls"- Kelley Malone, Kathryn Smith, Sue Lee, Natalie Nicholas, Katie Margolis, Vanessa Sanchez… you guys have been the best friends I could have ever asked for. It's hard to find people to spend time with that are intelligent and motivated- but also tons of fun- like each and every one of you. Erica Bankowski- I'm not sure what I'm going to do without my other half in NC- your energy and enthusiasm for life will definitely be missed!

Who could get through six years of graduate school without the support of their family? First off, I have to thank Aunt Teddy and "Un-cool" Mike. You have 'adopted' Greg and I and treated us as children of your own- and I can never thank you enough for that. I will greatly miss regular dinners, card games, and picnics. We will definitely come back to visit often! Jen, Luke, Tyler, and Katie- you guys are truthfully like brothers and sisters to me. I will miss having

regular game nights, nights out on the town, yoga classes, and ridiculous Tai Chi and Jazz dance classes.

Sheryl, Don, (Mom and Dad Wehn!), Jason, and Mashell- although you guys are all the way across the country, your support and love has been much appreciated through this journey and as I take the next step and move to North Carolina. Don't worry, "Ellie" will still come and visit- and you can come visit us in North Carolina- bring your golf clubs!

Mom, Dad, and Jon- you have always believed in me- sometimes when I didn't believe in myself- and been my biggest cheerleaders. I truthfully wouldn't be where I am right now without you behind me all the way. I can't say thank you enough.

Elaysia- you will ALWAYS be my little sister. Although I was supposed to be there for you- you have been there for me in a way that I could never thank you enough for- just by being you. You are such a beautiful, intelligent, out-going, and absolutely wonderful young woman. You encourage me every day to be the best that I can be- because I wanted to set a great example for you. I know you can do anything you want to do in life. I can't wait until you come and visit me in North Carolina- and I promise to come back and visit you.

Last but certainly not least- my wonderful husband Gregory- who has been there for me every step of the way. Thank you for sticking by my side through thick and thin. You have afforded me every opportunity to complete this journey and have attempted to do everything to make it easier for me- from vacuuming while I was studying or writing my comps, celebrating with me, to just being there for me when I had a bad day. I am truly at a loss for words as to how to say thank you for everything you have done for me. I can't wait to start the next chapter in our life!

1.0 INTRODUCTION

Generation of a multi-cellular organism from the fertilized egg is a remarkable process involving cell proliferation, cell differentiation, morphogenesis, and pattern formation. These carefully orchestrated processes ultimately result in formation of a functional organism and require many genes that are essential for survival of the embryo. Tissue-specific transcription factors regulate distinct gene expression profiles in the developing embryo resulting in the generation of different cell types that exhibit distinct behaviors and functions. The T-box family of transcription factors are critical components of cell type specification, differentiation, and proliferation during embryonic development and when mutated result in specific syndromes in humans (Naiche et al., 2005b). Our lab is primarily interested in two T-box transcription factors, T and Tbx6, and their roles in mesoderm development in the mouse embryo. Understanding how these critical transcription factors regulate mesoderm development can elucidate potential roles in congenital human birth defects.

1.1 DEVELOPMENT OF THE MOUSE EMBRYO

After fertilization, the single cell mouse embryo undergoes several rounds of cell division to form a ball of cells known as the blastocyst. The blastocyst consists of two different cell types; the outside layer, the trophectoderm, which will give rise to the extra-embryonic tissues that

support development of the embryo and contribute to the formation of the placenta, and the pluripotent inner cell mass (ICM), which will form the embryo proper. Implantation into the mother's uterus occurs at approximately embryonic day (e)4.5, after which the embryo will then elongate and form a hollow, cup-shape. As the embryo elongates, the ICM will contribute to the epiblast, from which the definitive ectoderm, mesoderm and endoderm are derived, as well as the extraembryonic mesoderm of the yolk sac, allantois, and amnion (Tam et al., 2006). The ectoderm will give rise to neural tissue and epidermis; mesoderm will form the kidneys, gonads, heart, skeletal muscle, and the skeleton; and endoderm will form the liver, pancreas, lungs, and gut (Tam and Loebel, 2007).

1.1.1 Gastrulation and formation of the mesoderm

During gastrulation the three germ layers of the embryo are established and assume their proper position in the embryo. It is from these germ layers that all of the fetal tissues will develop. Gastrulation commences at e6.5 with the formation of the primitive streak (PS) at the posterior end of the embryo. At the PS, epiblast cells down-regulate E-cadherin, undergo an epithelial-to-mesenchymal transition (EMT) then ingress through the PS (Zohn et al., 2006). The PS extends anteriorly as cells continue to ingress and migrate away. Fgf8, Fgf receptor-1 (Fgfr1), and Eomesodermin (Eomes, a T-box transcription factor) are crucial for this process, as embryos lacking either Fgf8, Fgfr1 or Eomes signaling fail to undergo the necessary EMT, and cells accumulate in the vicinity of the PS but fail to ingress (Ciruna and Rossant, 2001; Deng et al., 1994; Russ et al., 2000; Sun et al., 1999).

The timing and position that cells exit the PS determine the type of mesoderm generated: axial (node, notochord), paraxial (somites), intermediate (kidneys, gonads), or lateral plate

(splanchnic, somatic mesoderm) (Tam and Beddington, 1987). Our lab is primarily interested in the specification and patterning of the paraxial mesoderm (PAM). The presumptive PAM migrates out of the anterior portion of the PS and assumes a position lateral to the axial mesoderm thus establishing the bilateral presomitic mesoderm (PSM) (Hatada and Stern, 1994). After closure of the neuropore at e10.5, mesoderm, including PAM, is generated by the tailbud, a mass of cells at the caudal extremity of the embryo, which continues to extend the body axis until approximately e13.5 of development (Tam and Tan, 1992). In the chick, removal of the PS and node terminates axis elongation (Packard, 1978). Similarly, in the early somite stage mouse embryo, removal of the PS terminates somite formation and truncates the body axis (Smith, 1964). These studies demonstrate the necessity of the PS for generating all posterior structures.

Cells derived from the PS at e7.5 and transplanted under the kidney capsule generate teratomas that contain a variety of mature tissues originating from each of the three germ layers, demonstrating the pluripotent capacity of cells in the PS. Identical transplant experiments performed with caudal fragments of e8.5 and e9.5 embryos revealed a prominence of mesodermal derivatives such as muscle, adipose, and skeletal elements, as well as glandular and intestinal epithelium. These teratomas derived from later staged explants showed discernibly less mature neural tissue in e8.5 grafts and only trace amounts in e9.5 grafts, and displayed differentiated tissue types earlier than grafts from e7.5 embryos. These observations indicate a progressive restriction of the histogenic capacity of the original progenitor populations, with the e7.5 fragments generating adult tissues belonging to both the cranial and caudal body levels, but older fragments giving rise to tissues predominantly found in the caudal level. This progressive restriction of histogenetic capacity may be a genuine reflection of the functional capacity of the PS in normal anterior-posterior development of the embryo (Tam, 1984).

1.1.2 Somitogenesis

A hallmark of vertebrate embryogenesis is the formation of the somites, regularly spaced blocks of tissue flanking either side of the central neural tube. Somites are formed as the PSM segments in an anterior-to-posterior fashion via a mesenchymal-to-epithelial transition (MET) (Duband et al., 1987). The somite whose boundary is forming is denoted somite 0 (S0); the next somite anterior to this is S+1, and so on. After segmentation, somites become patterned by signals from the notochord, neural tube, surface ectoderm, and lateral mesoderm. The dorsal part of the somite will differentiate into the dermamyotome under the influence of signals from the surface ectoderm and roof-plate of the neural tube, while the ventral portion receives signals from the notochord and floor-plate of the neural tube to become the sclerotome (Figure 1). The dermamyotome, which will give rise to the dermatome and myotome, and sclerotome will ultimately go on to form the dermis, skeletal muscle, ribs, and vertebrae of the adult animal, respectively (Christ and Ordahl, 1995).

Maturation of the PSM along the anterior-posterior axis coincides with changes in molecular properties of the cells and gene expression profiles. Within the PS/PSM, expression of the secreted signaling molecules *wnt3a* and *Fgf8* is at the highest level posteriorly, with lower levels in the anterior (Aulehla et al., 2003; Dubrulle et al., 2001). Fgf8 plays a key role in maintaining the undifferentiated mesenchymal state of the PSM, and wnt3a is required for proper anterior-posterior boundary positioning in the segmenting PSM (Aulehla et al., 2003; Dubrulle et al., 2001; Dunty et al., 2008). Additionally, retinoic acid (RA, a signaling molecule derived from vitamin A) is found in a reverse gradient to that of *Fgf8* and *wnt3a* and functions to coordinate somite formation across the left-right axis of the embryo (Moreno and Kintner, 2004). Embryos that lack *Retinaldehyde dehydrogenase type 2 (Raldh2)*, the gene encoding the retinoic

biosynthetic enzyme necessary for processing vitamin A to RA, display small somites with left-right asymmetry defects due to a loss of coordination of clock oscillations across the left-right axis (Vermot and Pourquie, 2005).

Expression of several other genes oscillates to comprise a molecular segmentation clock. Molecular evidence of the segmentation clock was first revealed with the discovery of the cyclic expression of *c-hairy1* in the chick, which begins in as broad expression in the posterior (stage I) and progressively proceeds as a wavefront (stage II) which is subsequently restricted as an intensified band in the anterior PSM (stage III), and coincides with the amount of time it takes one somite to form embryo. *c-hairy1* is a transcription factor of the hairy/enhancer-of-split (HES) family that is downstream of the Notch signaling pathway (Palmeirim et al., 1997).

In the mouse, hairy/HES family of transcription factors, *Hes1* and *Hey2*, are expressed in a similar fashion to *c-hairy* (Jouve et al., 2000; Leimeister et al., 2000). Each wave of cycling genes are expressed every 90 minutes, and are progressively restricted to the anterior region of the PSM, where the next segment will form (Pourquie, 2001). Notch signaling pathway components play critical roles in somitogenesis and establishing the periodicity of the segmentation clock. Defects in somitogenesis have been reported in mice with mutations in *Notch1, Dll1, Dll3, Lfgn,* and *RBP-Jκ* (Conlon et al., 1995; Hrabe de Angelis et al., 1997; Kusumi et al., 1998; Oka et al., 1995; Zhang and Gridley, 1998), with mutations in *Notch1, RBP-Jκ,* and *Dll1* ultimately causing embryonic lethality. Notch is a large transmembrane receptor. Upon binding of the membrane-bound Notch ligand, Delta (Dll), the receptor undergoes a γ-secretase proteolytic cleavage event catalyzed by the presenilin family of proteins to release the Notch intracellular domain (NICD) (Selkoe and Kopan, 2003). The NICD then translocates to the nucleus where it associates with the transcription factor RBP-Jκ to activate expression of

downstream targets (Artavanis-Tsakonas et al., 1999). Affinity of Notch for its ligands can be modified by glycosylation in the Golgi by Fringe proteins (Artavanis-Tsakonas et al., 1999). Interestingly, the expression of *Lunatic Fringe* (*Lfgn*) in the mouse PSM is reminiscent of *c-hairy1* expression (Aulehla and Johnson, 1999; Evrard et al., 1998). While Fringe family members can act to either potentiate or inhibit Notch signaling, Lfgn has been shown to inhibit Notch mediated signaling in the anterior PSM to allow segmentation (Dale et al., 2003).

After each wave of Notch signaling, expression of *Dll3* becomes stabilized in the anterior of the newly segmented somite, while *Dll1* becomes stabilized in the posterior, setting up molecular differences between the anterior and posterior halves of somites. These molecular differences are critical for both initial segmentation and later for resegmentation. During resegmentation the sclerotome region segments again with the posterior part of one somite fusing to the anterior half of the next somite to form the vertebral unit. Thus, a vertebral unit is composed of cells from two adjacent somites, resulting in the attachment of one muscle group to two consecutive vertebrae (Huang et al., 2000a) (Figure 1). Additionally, proper anterior-posterior specification is required for migration of the spinal ganglia, which migrate exclusively through the anterior somite halves. A defect in anterior-posterior patterning of the somites can therefore also have deleterious effects on the peripheral nervous system in addition to the vertebrae (Bronner-Fraser, 2000).

The importance of proper somite patterning is highlighted by several human congenital birth defects. Mutations in *DLL3* and a missense mutation in *LFNG* contribute to the etiology of spondylocostal dysostosis, a congenital birth defect that results in hemivertebrae and fusion of ribs resulting in a shortened body axis (Bulman et al., 2000; Sparrow et al., 2006). In addition,

improper pattering of the somites may lead to more common vertebral disorders, such as congenital scoliosis, which is genetically complex (Sewell and Kusumi, 2007).

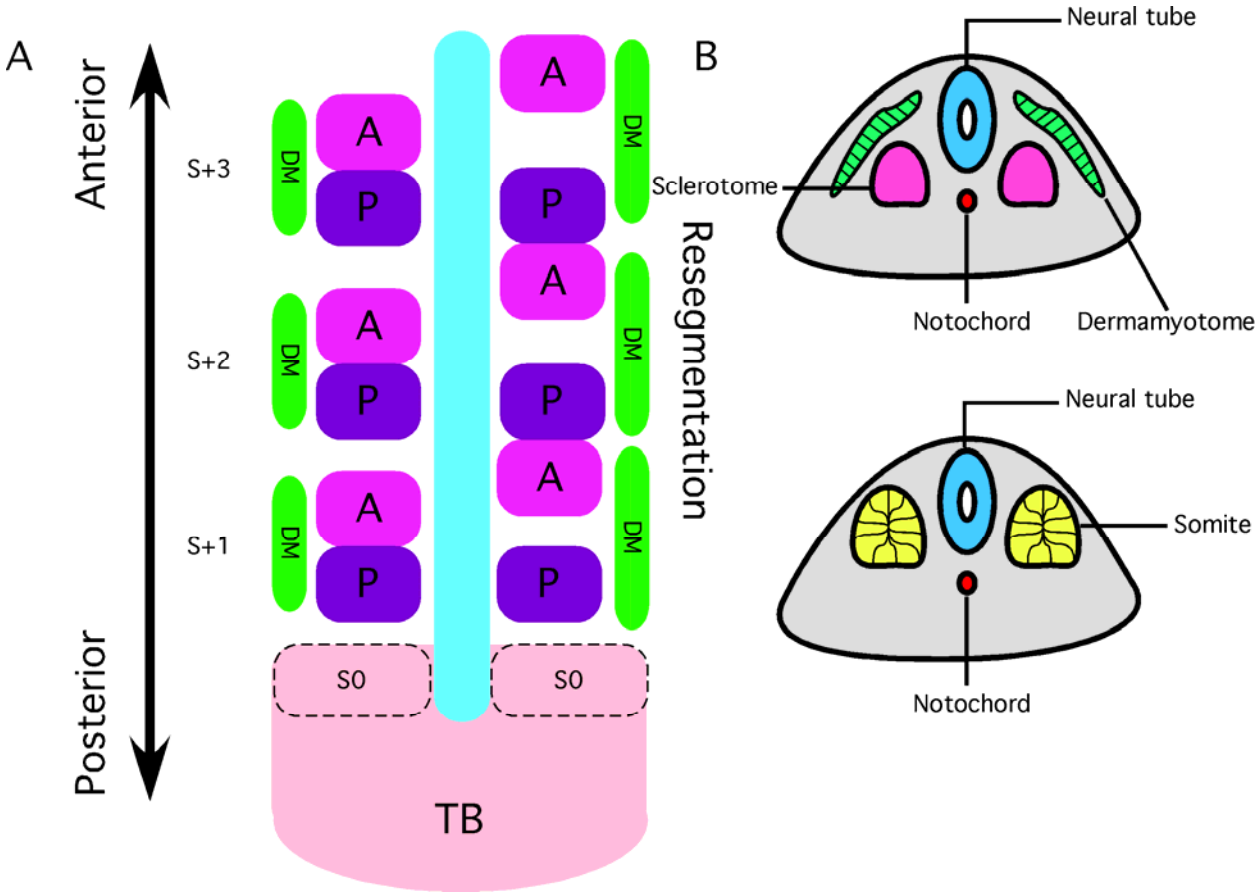

Figure 1: Schematic of Mouse Somitogenesis

(A) Somite segmentation from the mesenchymal tailbud (TB) forms an epithelial somite that flanks either side of the central neural tube (blue). Establishment of anterior (A) – posterior (P) polarity is essential for proper re-segmentation of the somites, which joins the posterior portion of each sclerotome with the anterior half of the next most posterior sclerotome, allowing for proper patterning of the vertebrae and attachment of muscles. (B) Cross-section depicting differences in anterior and posterior somite maturation. Somitic cells closest to the surface ectoderm become dermamyotome (DM), which gives rise to the dermatome and myotome. Ventral somitic cells receive signals from the notochord and ventral neural tube to become the sclerotome.

1.2 T-BOX GENES AND REGULATION OF DEVELOPMENT

Tissue-specific transcription factors regulate distinct gene expression profiles in the developing embryo resulting in the generation of different cell types with unique behaviors. Genetic and molecular evidence has linked the T-box family of transcription factors to important processes in cell type specification, differentiation, and proliferation in a variety of different model organisms including humans. This family of transcription factors share a conserved DNA binding domain (DBD) comprised of approximately 180 amino acids (Tada and Smith, 2001).

The T-box family of transcription factors play critical roles during embryonic development such that when mutated result in specific syndromes in humans (Naiche et al., 2005b). The common DBD shared between T-box transcription factors raises the question of how T-box transcription factors that are co-expressed establish specificity for downstream targets. Research has only begun to uncover the variety of ways in which T-box transcription factors interact when co-expressed to regulate transcription of downstream targets. Many studies have revealed synergistic, additive, and competitively antagonistic mechanisms in which T-box factors can interact with endogenous enhancers.

1.2.1 Precedence for competition between T-box transcription factors

The founding member of the T box transcription family is Brachyury, or T. The role of T in development is discussed in more detail in section 1.3. Initial experiments with T in a PCR-based binding site selection experiment showed that T binds to the twenty nucleotide palindromic sequence T(G/C)ACACCTAGGTGTGAAATT (Kispert and Hermann, 1993). All T-box proteins tested to date can bind the T palindromic target sequence *in vitro*, however *in*

vivo they have unique targets suggesting that other regions of T-box proteins or non-conserved changes within the T-domain determine target selection (Carreira et al., 1998; Hsueh et al., 2000; Macindoe et al., 2009; Tada et al., 1998; White and Chapman, 2005). The conserved nature of the T-box DBD allows for all T-box transcription factors to bind to the core half-site, 5'-AGGTGT-3', which is contained within the palindrome, with bases outside of this being more variable. The crystal structure for the DBDs of *Xenopus* homolog of T, Xbra, and human TBX3 bound to the canonical palindromic binding site revealed that conserved amino acids made contacts with the DNA, supporting the strong conservation of binding preferences (Coll et al., 2002; Muller and Herrmann, 1997) (Figure 2). In tissues where more than one T-box transcription factor are expressed, functional specificity is at least in part determined by preferences for orientation and spacing of the half-sites, differences in bases flanking the conserved core, and physical interactions with differing co-factors (Conlon et al., 2001; Messenger et al., 2005).

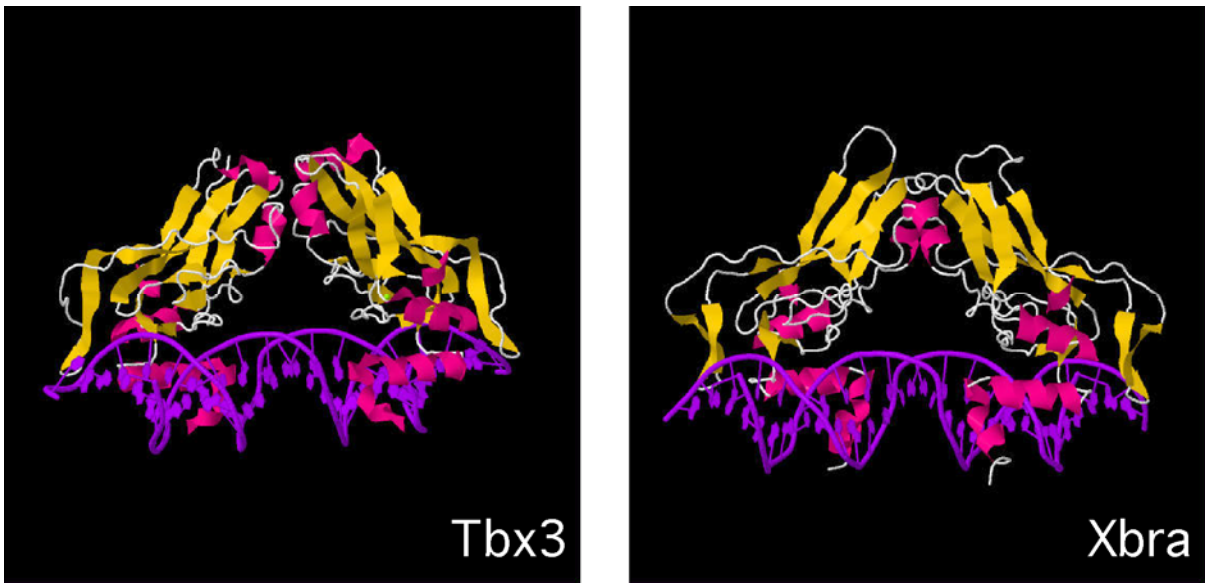

Figure 2: Crystal structures of Xbra and TBX3 in complex with the palindromic binding site

Crystal structure of the DNA binding domain of *Xenopus* Brachyury (Xbra) and human TBX3 in complex with the 24 bp palindromic T-box binding site reveals strong conservation of overall structure and amino acids that directly contact the bases. Tbx3 binds as two monomers to the palindromic sequence, while Xbra homodimerizes. Crystal structures were obtained from the protein data bank at www.pdb.org, and originally published in (Coll et al., 2002; Muller and Herrmann, 1997).

There are 22 identified human T-box proteins, and thus there are many instances *in vivo* where multiple T-box proteins are co-expressed within a particular tissue. We are interested in how multiple T-box proteins coordinate the regulation of downstream targets. Due to the conserved nature of the DBD, it is possible for co-expressed T-box factors to compete with each other for targets, cooperate in activation or repression of targets, or have completely different targets due to differences in binding preferences.

In vivo, many T-box proteins compete for binding sites within the enhancers of common target genes. For example, in mouse heart formation, Tbx2, a transcriptional repressor, is expressed in a subdomain of cardiac progenitors that also express the transcriptional activator

Tbx5. Tbx2 and Tbx5 compete for binding to T-box binding elements in the *ANF* enhancer (Habets et al., 2002). Additionally, Tbx18 (a transcriptional repressor) can compete with Tbx5 for binding the *ANF* enhancer in the sinus horn mesenchyme of the heart, which is devoid of *ANF* expression (Farin et al., 2007). Competition has also been observed between T-box transcription factors in zebrafish mesoderm formation. The function of no-tail (ntl) is required in the dorsal mesoderm to promote differentiation of *MyoD* expressing adaxial cells, which is antagonized by tbx6 within the ventrolateral mesoderm. Ectopic expression of *tbx6* within the *ntl* expression domain is sufficient to elicit a *ntl* hypomorphic phenotype (Goering et al., 2003).

Although it is apparent that the preferences of different T-box factors for orientation and spacing of half-sites and sequences flanking the core differ, what exactly contributes to the differences in specificity is still a point of contention. Conlon et al. (2001) suggested that specificity for differing targets could lie within non-conserved amino acid within the T-domain in residues that contact the DNA (Conlon et al., 2001). These studies in *Xenopus* with the co-expressed mesoderm inducing T-box factors *Xbra, Eomes,* and *VegT* confirmed differing preferences for orientation and spacing of half-sites, and sequences flanking the core half-site. VegT and Eomes induced *Pintovallavis, Xsox17α, goosecoid, Xwnt8, Mix1, Xwnt11, and Bix4,* while Xbra induced only *Xwnt11, Bix4,* and weakly *Pintavallavis*. Studies with chimeric versions of these proteins where the DBDs were swapped revealed that much of the specificity of T-box transcription factors lies within the DBD, such that chimeric versions of the proteins are able to activate downstream targets according to the T-box sequence they contain. Superposition of the VegT and Eomes DBD onto the pre-exisiting Xbra crystal structure revealed two amino acid differences at DNA contact points. The amino acid substitution at position 214 is conserved, replacing an alanine in Xbra with a glycine in Eomes and VegT, and was therefore

not further investigated. All T homologs contain a basic lysine (K) at residue 149, while VegT and Eomes contain the neutral polar asparagine (N), leading to the hypothesis that the amino acid at position 149 may confer specificity. N149K mutations within VegT and Eomes resulted in loss of activation of the VegT and Eomes targets *goosecoid,* and *Pintavallavis,* but not the Xbra specific *Xwnt11* and *Bix4*, suggesting that the residue at 149 may play a key role in specificity (Conlon et al., 2001).

Whether amino acid 149 is involved in conferring specificity is debatable as residue 149 contacts only the phosphate backbone, and does not make a base-specific contact (Muller and Herrmann, 1997). In addition, the reciprocal mutation (K149N) within Xbra does not result in activation of VegT and Eomes specific targets such as *goosecoid.* Moreover, others have found that the N149K mutations within VegT and Eomes only confered an overall decrease in activation, not a switch in target gene preference (Messenger et al., 2005). Altogether, the data suggests that specificity may not be conferred by this particular amino acid (Messenger et al., 2005; Muller and Herrmann, 1997).

If the specificity of Xbra, VegT, and Eomes does not lie within residue 149, where does the specificity come from? Messenger et al. (2005) suggested that the specificity of Xbra lies within the N-terminal domain, and demonstrated that chimeric proteins where the N-terminal domains are swapped between VegT and Xbra behave differently. The N-terminal domain of Xbra directly interacts with Smad1, the BMP pathway component, in a BMP-dependent manner. Xbra, in conjunction with Smad1, activates the expression of *Xenopus Eomesodermin* (*Xom)*, a direct repressor of *goosecoid* expression (Messenger et al., 2005; Trindade et al., 1999). Additionally, morpholino mediated knockdown of *Xom* allows Xbra to induce expression of *goosecoid.* Therefore, it has been proposed that it is an indirect pathway that is responsible for

the apparent inability for Xbra to activate *goosecoid*, such that Xbra mediated expression of *Xom* represses any Xbra mediated expression of *goosecoid* (Messenger et al., 2005).

More recent studies have investigated the binding kinetics of T-box factors (Macindoe et al., 2009). As might be expected, differences in binding site affinity can contribute to transcriptional activity at specific enhancers. Although T-box factors all bind to the same consensus binding sequence, surface plasmon resonance (SPR) experiments with the DBDs of Tbx20, Tbx2, and Tbx5 and immobilized half-site core sequence revealed differences in the relative affinities and kinetics of binding between factors. Tbx5-DBD binds more tightly to the half-site, with a binding affinity of $0.2x10^{-6}$ M. Tbx20-DBD and Tbx2-DBD bound with an affinity of $0.9x10^{-6}$ M and $1.5x10^{-6}$ M, respectively. In addition, Tbx5-DBD associated with the oligonucleotide twenty times quicker than either Tbx20-DBD or Tbx2-DBD, and only disassociated five times quicker. This indicates that specificity for particular orientation and spacing of half-sites in addition to overall binding dynamics such as on- and off-rates contribute to the overall activation ability of T-box transcription factors (Macindoe et al., 2009).

It is clear that different T-box transcription factors differentially activate targets and have differing preferences for orientation and spacing of half-sites. Whether or not amino acid substitutions within the T-domain or other portions of the protein lend specificity for a particular target is currently unresolved.

1.2.2 Interaction with co-factors partially determines transcriptional activity

As the DBD is the only region of homology shared among T-box transcription factors, regions outside of the DBD can interact with a wide variety of other specific factors. For the most part, the activation and/or repression domains are found within the C-terminus of the protein (Conlon

et al., 2001). Thus far most T-box factors have been found to act as a transcriptional activator or repressor (Kawamura et al., 2008), however Tbx20 contains both a strong activation and repression domain, and therefore can act as both a transcriptional activator or repressor depending on the context (Brown et al., 2005). In the zebrafish, the Groucho/TLE associated co-repressor, Ripply1, associates with No-tail (Ntl) and Tbx24 to convert them from transcriptional activators to repressors (Kawamura et al., 2008).

The interaction of T-box factors with other transcription factors can also result in synergistic activation of target genes. T-box interacting factors include members of the homeodomain, GATA zinc finger, and LIM domain families (Naiche et al., 2005b). Tbx2 and Tbx5 both interact with Nkx2.5 in the developing heart (Habets et al., 2002), suggesting that competition for common co-factors could also potentially regulate their activity. T-box transcription factors can also heterodimerize. In the developing heart, it has been shown that Tbx20 and Tbx5 can directly interact via the N-terminal and T-box region to synergistically activate downstream targets (Brown et al., 2005).

1.2.3 T-box transcription factors and human health

Most of the known autosomal human syndromes related to mutation of T-box transcription factors occur in heterozygous individuals (haploinsufficiency). Presumably, homozygous mutations would result in embryonic lethality resulting in miscarriage (Naiche et al., 2005b). A comparison of human and mouse phenotypes associated with mutation of T-box factors is presented in Table 1. Similarities between mouse and human phenotypes associated with haploinsufficiency of T-box factors allow for the mouse to serve as an exceptional model system of human congenital disorders associated with T-box factors.

Heterozygous mutations in human *TBX1*, *TBX3*, *TBX4*, *TBX5*, and *TBX22* result in the human syndromes DiGeorge, ulna-mammary, small patella, Holt-Oram, and X-linked cleft palate with ankyloglossia, respectively (Baldini, 2003; Bamshad et al., 1997; Basson et al., 1997; Bongers et al., 2004; Braybrook et al., 2001). Recessive mutations in *TBX15* and *TBX19* result in Cousin's syndrome and adrenocorticotrophic hormone (ACTH) deficiency, respectively (Asteria, 2002; Lausch et al., 2008). We hypothesize that in syndromes associated with haploinsufficiency, the presence of other T-box transcription factors in the same tissues may play a role in the phenotypes observed.

Investigation into ten different naturally occurring missense alleles of *TBX22* reveals that those causing the X-linked cleft palate phenotype occur in residues within the DBD that are in or near predicted DNA contact points (Andreou et al., 2007). This results in proteins that are unable to bind DNA stably as assayed by electrophoretic mobility shift assays (EMSAs). Interestingly, a G118C missense mutation within TBX22 occurs in a conserved DNA contact residue and results in the inability to bind to DNA (Andreou et al., 2007), similar to the situation with Tbx5, where a missense mutation at the same position (G80R) also abolishes DNA binding and leads to Holt-Oram syndrome (Ghosh et al., 2001). Mutations within the DBD of TBX20 also have been found to contribute to the etiology of congenital heart disease (Liu et al., 2008). In contrast, a mutation in TBX22 outside of DNA the contact points in α helix 2 (E187K) does not alter DNA binding (Andreou et al., 2007). In total, this suggests that mutation of critical DNA contact residues of T-box transcription factors reduce binding ability and may contribute to the etiology of congenital birth defects.

In addition to contributing to congenital birth defects, misregulation of several T-box transcription factors has been implicated in the development of cancer. Several lines of evidence

link T-box transcription factors to control of cell proliferation and tumorigenesis. *TBX2* is located within a segment of DNA that is often upregulated in breast and pancreatic cancers, resulting in over-expression of *TBX2* in 10% of breast tumors (Barlund et al., 2000). Murine Tbx2 represses the expression of N-myc1 in the atrioventricular canal and outflow tract of the developing heart, where low proliferation is observed (Cai et al., 2005). N-myc is required for early cardiac proliferation, and over-expression has been linked to childhood tumors (Davis et al., 1993). Additionally, over-expression of both *TBX2* and *TBX3* has been molecularly linked to the p53 pathway and ultimately results in genomic instability (Carlson et al., 2001; Jacobs et al., 2000). Genomic rearrangements that increase the copy number of T, which is essential for notochord patterning, are implicated in the development of familial chordomas (Yang et al., 2009).

Clearly, unraveling how T-box factors interact when co-expressed to control transcription of downstream targets is of utmost importance to understand the developmental basis of several congenital birth defects, as well as some cases of cancer. This makes understanding the T-box family of transcription factors of relevant to human health.

Table 1: Comparison of human and mouse phenotypes associated with mutation of T-box transcription factors

Gene	Human Syndrome	Human Phenotype	Mouse heterozygous phenoype	Mouse homozygous phenotype
T (Brachyury)	Congenital vertebral malformations (haploinsufficiency)	Fusions/malformations of vertebral column	Shortened tail	Severe truncation of body axis; lethal
Tbx1	DiGeorge (haploinsufficiency)	Craniofacial, glandular, vascular and heart abnormalities	Thymus and vascular abnormalities	Craniofacial, glandular, vascular, heart abnormalities; lethal
Tbx3	Ulnar-mammary (haploinsufficiency)	Hypoplastic mammary glands, abnormal external genitalia, limb abnormalities	Hypoplastic mammary glands, abnormal external genitalia	Yolk sac, limb, mammary gland defects; lethal
Tbx4	Small patella (haploinsufficiency)	Reduced patella	Reduced allantois growth	Allantois, hindlimb defects; lethal
Tbx5	Holt-Oram (haploinsufficiency)	Heart and hand abormalities	Heart abnormalities	Severe heart malformations; lethal
Tbx15	Cousin's (recessive)	Craniofacial, limb malformations	Normal	Craniofacial defects, limb defects, pigment pattern alterations; viable
Tbx19	ACTH deficiency (recssive)	Deficient production of adrenocorticotropic hormone from pituitary	Normal	ACTH deficiency, pigment defects; viable
Tbx22	X-linked cleft palate with ankyloglossia (haploinsufficiency)	Cleft palate, ankyloglossia	Not known	Not known

Table adapted from (Naiche et al., 2005b).

1.3 BRACHYURY IS ESSENTIAL FOR MESODERM FORMATION

Brachuyruy (*T*) was first discovered as a spontaneous mutation deleting a 160 kb region on chromosome 17, resulting in a short tail phenotype in heterozygotes and embryonic lethality in homozygotes with severe axis truncations (Dobrovolskaia-Zavadskaia, 1927). Early mutant analyses suggested that T played a critical role in mesoderm formation, although the specific function of T was not elucidated until the advent of modern molecular techniques. Cloning of the *T* gene revealed that the original *T* mutation resulted from a 200 kilo-base (kb) deletion (Herrmann et al., 1990). Furthermore, cloning of *T* allowed for the first biochemical studies to elucidate the function of T as a tissue specific transcription factor which binds to a 24 base-pair (bp) palindromic sequence (Kispert and Hermann, 1993; Kispert et al., 1995). Other members of the family of T-box transcription factors were identified through their homology to the T-domain. Degenerate PCR primers for conserved regions of the T-box amplified a family of transcription factors with similar DNA binding properties (Bollag et al., 1994). Initial studies performed with T paved the way for similar studies performed with other later discovered T-box family members to further characterize the family.

1.3.1 T expression patterns and mutant analysis

In the mouse, *T* transcripts are first detectable at e5.5 in the proximal epiblast in a population of cells by the embryonic-extraembryonic junction (Thomas and Beddington, 1996). At e7.0, *T* is present within the nascent PS, and at later stages is present in the node and notochord, however it is down-regulated in the paraxial and lateral plate mesoderm. Expression of *T* persists in the tailbud until axis extension is complete at e13.5 (Wilkinson et al., 1990) (Figure 3B).

T heterozygous mutants have a short tail, while *T* homozygous null embryos display severe axis truncations posterior to the forelimb bud, suggesting dosage of T is critical (Figure 3C). The homozygous mutant phenotype is first morphologically apparent at e8.5 as mutants fail to extend the notochord, and have a small allantois (Chesley, 1935). *T* homozygous mutants have a discontinuous notochord and form only 7-9 abnormal, anterior somites. The allantois, notochord, and PAM are all derived from the PS (Beddington et al., 1992). Ultimately, embryonic death at e10.0 occurs due to failure of the allantois, the embryonic precursor to the umbilical cord, to connect to the chorion.

The phenotypes observed in *T* mutants are likely the result of defective mesodermal cell movements in the PS, which results in cells accumulating in the streak. *T/T* mutant embryonic stem (ES) cells injected into a wild-type blastocyst displayed migration defects reminiscent of *T* null embryos at gastrulation and do not colonize trunk paraxial or lateral mesoderm, suggesting that *T* acts cell autonomously (Beddington et al., 1992; Wilson et al., 1995). These observations led to the hypothesis that the T protein could be required to alter cell adhesion molecules during EMT so that nascent mesoderm can properly migrate out of the PS. Cell mobility studies with zebrafish *ntl*, the ortholog of T, also support this hypothesis, noting that later staged convergent-extension movements were defective, suggesting conservation of function (Glickman et al., 2003). Additionally, the *T* mutant phenotype is strikingly similar to the phenotype of *integrin α5* null embryos, suggesting that *T* may act upstream. Integrin α5 is involved in cell-matrix recognition, and may therefore play a role in cell migration during gastrulation (Goh et al., 1997).

Further evidence for migration defects in the *T* null mutant embryos comes from studies of explanted PS tissue. Mesodermal cells isolated from PS-staged mouse embryos cultured *in*

vitro on extracellular matrix (ECM) exhibit behaviors similar to that inside of the embryo in terms of cell shape, rate of movement, and cell division (Hashimoto et al., 1987). PS tissue explanted from a *T* null mutant displayed identical cell morphology and rate of cell division, but had a slower rate of movement as compared to wild-type cells (39.4 μm/h and 52.8 μm/h, respectively). In wild-type populations approximately 10% of cells moved rapidly (over 80 μm/h), whereas no cells moved rapidly in *T* null populations, further suggesting that loss of *T* affects cell migration that contributes to the *T* mutant phenotype.

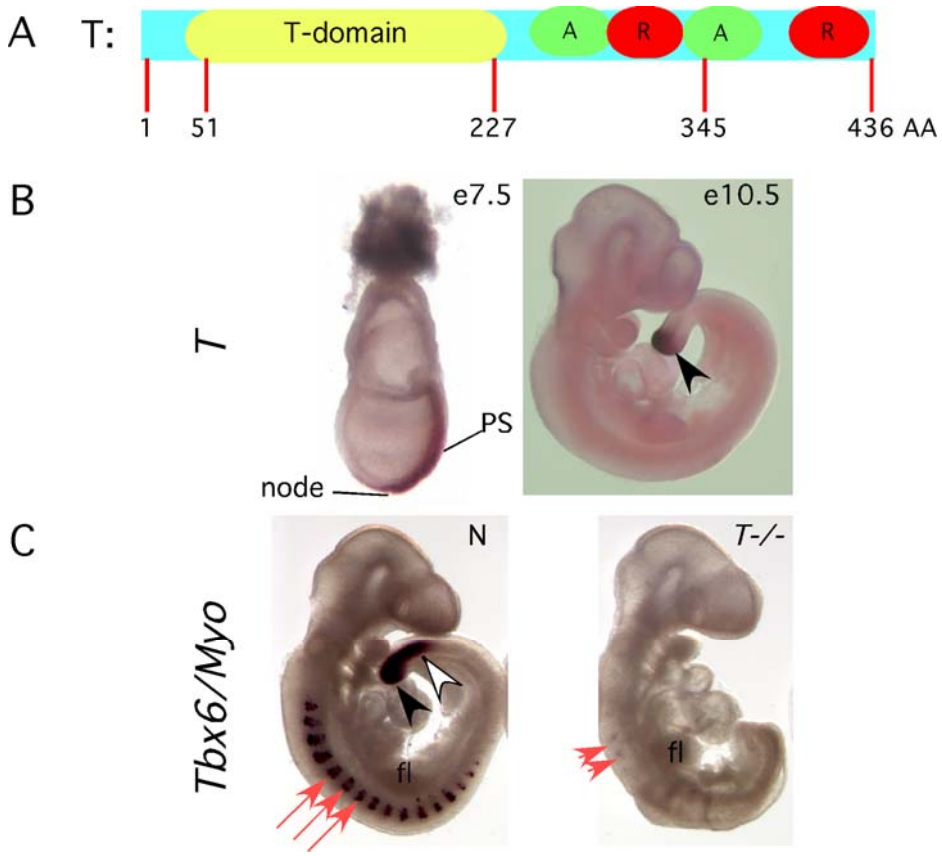

Figure 3: *T* expression patterns and mutant phenotypes

(A) Domain structure of T. Activation (A) and repression (R) domains are depicted (Kispert et al., 1995). Truncation of the protein in the T^{Wis} allele occurs at amino acid 345. (B) *In situ* hybridization at e7.5 reveals *T* expression within the PS and node. At e10.5, *T* is expressed within the tailbud (arrowhead). (C) *In situ* hybridization for *Tbx6* reveals expression in the tailbud (closed arrowhead) and PSM (open arrowhead), and *myogenin* (red arrows) in a normal (N) e10.5 embryo. In a *T* mutant, expression of *Tbx6* is lost. *Myogenin* expression is restricted to two irregular anterior somites (arrowhead) and truncation of the body axis occurs shortly after the forelimb bud (fl).

1.3.2 Biochemical studies reveal that T functions as tissue-specific transcription factor

Landmark biochemical studies revealed that T binds to 24 bp almost perfect palindromic sequence, with a core half-site sequence of 5'-AGGTGT-3', with the TGT triplet providing the majority of the physical contact points. This DNA binding property is conferred by the first 227 amino acids (AA) of the protein (out of 436 AA total), identical to that of the *Xenopus* and zebrafish orthologs (Kispert and Hermann, 1993). Furthermore, T functions as a tissue-specific transcription factor, with two transcriptional activation and two repression domains contained within the C-terminus of the protein (AA 230-436) and a cryptic nuclear localization sequence (NLS) contained between AA 137 and 320 (Kispert et al., 1995) (Figure 3A).

Crystallographic studies of the T-domain (DBD) of Xbra in complex with the 24 bp palindromic sequence revealed T bound as a dimer. Each monomer consisted of a seven-stranded β–barrel, with a novel sequence specific DNA recognition motif in which protein contacts are made both in the major and minor grooves of the DNA, contacting the phosphate backbone over 22 of the 24 bp. Additionally, the T-domain was found to be a monomer in solution, and dimerized only upon contact with DNA (Muller and Herrmann, 1997) (Figure2).

1.3.3 Placing T in a genetic pathway during mesoderm formation

Although much is known about the mutant phenotypes, expression patterns, and biochemical properties of T, surprisingly little is known about the genetic pathways employed by T to elicit such effects. Several direct targets of T orthologs in *Xenopus* and zebrafish have been identified; but only *Dll1* has been suggested to be a direct downstream target of T within the mouse genome (Hofmann et al., 2004). Despite the paucity of known murine T targets, Xbra, the *Xenopus*

ortholog, is known to activate the expression of *eFGF, Xwnt11* and *Brachyury-induced-homeobox* (*Bix4)* in the marginal zone (Casey et al., 1998; Tada et al., 1998; Tada and Smith, 2000). *eFGF* and *Bix4* are required for mesoderm formation and patterning, and *Xwnt11* is required for gastrulation and convergent-extension movements via the non-canonical Wnt pathway. Despite the identification of targets in *Xenopus*, a direct correlation between targets in other species cannot be made, as the mouse ortholog of the Bix/Mix family, *Mml*, which is expressed in the visceral endoderm, has been suggested to be a direct target of another T-box gene, *Eomesodermin* (Russ et al., 2000).

Targets of No-tail (Ntl), the zebrafish homolog of T have been elucidated using a chromatin immunoprecipitation followed by microarray analysis technique (ChIP-chip), utilizing mid-gastrula staged embryos (Morley et al., 2009). This revealed that a homeobox transcription factor *floating head* (*flh*) is a direct target of Ntl in the notochord and is required to specify notochordal fate. Additional targets identified by this method included genes required for specification of posterior fate, muscle cell fate, cell movements, and establishment of left-right asymmetry, with 27% of the genes identified encoding transcription factors including Hox, forkhead, and T-box factors. Further investigation is required to understand the full degree of overlap between regulation of mesoderm formation by *T* orthologs in *Xenopus*, zebrafish, and other vertebrates.

Although there is only one suggested direct target of T in the mouse, the signaling pathways that control expression of *T* are better documented. *wnt3a* is expressed within the PS and is required for proper trunk and tail paraxial mesoderm, such that *wnt3a* mutants only form 7-9 abnormal anterior somites. In *wnt3a* mutants, posterior somites do not form and instead an ectopic neural tube forms ventral to the normal axial neural tube. w*nt3a* embryos, initially

express *T*, but expression is lost by e8.5 (Takada et al., 1994). Elements necessary to drive the expression of *T* in the PS, but not in the node or notochord, lie 500 bp upstream of the start of transcription (Stott et al., 1993). Within this region are two canonical LEF1 binding sites. LEF/TCF transcription factors are the effector molecules of the Wnt signaling pathway, suggesting that wnt3a directly regulates expression of *T* in the PS. (Yamaguchi et al., 1999).

Fgf8 is a secreted signaling molecule that is expressed in the epiblast at e5.75, and subsequently in the PS. *Fgf8* is critical for gastrulation and expression of *T* (Crossley and Martin, 1995). Like the *T* mutant embryos, mesoderm fails to migrate away from the PS in *Fgf8* mutants. A similar phenotype is also observed in *Fgf Receptor-1* (*Fgfr1*) mutants, suggesting that *Fgf8* binds to *Fgfr1* during gastrulation to exert its effect (Ciruna et al., 1997; Yamaguchi et al., 1994). Additionally, *T* is not expressed in *Fgf8* or *Fgfr1* mutants, suggesting that Fgf8 and Fgfr1 are upstream of T and are required for proper expression of *T*.

1.3.4 T^{Wis}, an allele of *T*

Six alleles of *T* have been characterized. Through this analysis, there appears to be a direct correlation between the strength of the allele and the anterior range of the phenotype (Beddington et al., 1992). The T^{Wis} allele arose as a spontaneous insertion of a transposable element into exon 7 that results in transcripts that bypass the exon 7 splice donor site. Although T^{Wis} is transcribed and translated in the same pattern as wild-type *T*, the insertion results in the formation of a primary mRNA product that results in a truncated protein at AA 345 (Kispert et al., 1995; Shedlovsky et al., 1988) (Figure 3A). Up to eight different mutant transcripts are made from the T^{Wis} allele, although no wild-type transcripts are made (Goldin and Papaioannou, 2003). T^{Wis} heterozygous mice display a shortened to no tail phenotype that is highly variable. T^{Wis}

homozygous embryos completely lack somites, an allantois, and notochord, and display a kinked neural tube. In contrast, *T* null embryos form 7-9 abnormal anterior somites, a small, malformed allantois, and lack only posterior notochord (Chesley, 1935). The increased severity of the T^{Wis}/T^{Wis} phenotype as compared to the *T* null phenotype suggests that T^{Wis} acts as a dominant-negative to interfere with a co-expressed T-box transcription factor. We hypothesized that this is T-box 6 (Tbx6). Tbx6 is a T-box transcription factor that is co-expressed with T and is discussed in section 1.4.

1.3.5 Implications for T in human health

A heterozygous 1013C>T missense mutation in human *T* has been linked to congenital vertebral malformations (CVMs) found in three genetically unrelated individuals diagnosed with sacral agenesis, Klippel-Feil syndrome, and multiple cervical and thoracic vertebral malformations. The missense mutation changes a moderately conserved alanine residue to valine (Ghebranious et al., 2008). Although the phenotypes are variable amongst individuals, heterozygous mutations in *T* may lead to CVMs, like congenital scoliosis.

Additionally, T has been linked to some forms of cancer. T can induce tumor cells to undergo an EMT, an important step in metastasis of tumors. Specifically, T induces a down-regulation of E-cadherin, and was found to be up-regulated in a variety of cancerous tumors (Fernando et al., 2010). Secondly, *T* has been identified as the major susceptibility gene for familial chrodomas, a rare bone cancer arising from cells of notochordal origin. Genomic duplications of the region containing T has been isolated within chordoma affected families, and *T* transcripts are specifically up-regulated in chordomas in microarray assays, but not in other related bone cancers (Vujovic et al., 2006; Yang et al., 2009).

Identification of direct downstream targets of T as well as further understanding transcriptional regulation of downstream targets will help to elucidate the etiology of CVMs related to mutations in T, and in human cancers where T is implicated.

1.4 TBX6 IS CRITICAL FOR POSTERIOR PARAXIAL MESODERM FORMATION AND PATTERNING

Tbx6 encodes a T-box transcription factor that is critical in patterning the mesoderm, and is co-expressed with *T* in the PS and tailbud. Unlike *T*, *Tbx6* heterozygous mice are apparently normal. *Tbx6* null embryos do not form paraxial mesoderm posterior to the forelimb bud, suggesting Tbx6 is critical for the specification of this tissue (Chapman and Papaioannou, 1998).

1.4.1 Tbx6 expression patterns and mutant analysis

In the mouse, *Tbx6* transcripts are first detectable at e7.0 within the PS and paraxial mesoderm cells migrating out of the PS (Chapman et al., 1996). At e9.5, *Tbx6* is detected in the PSM and tailbud. Expression of *Tbx6* transcript and protein is sharply down-regulated upon segmentation of the PSM into somites (Chapman et al., 1996; White et al., 2003). Expression in the tailbud persists until axis extension is complete at e13.5. The expression of *T* and *Tbx6* overlap within PS and tailbud (Chapman et al., 1996) (compare Figure 3B and Figure 4B).

Tbx6 heterozygotes are normal and fertile. Mutant analyses of null and hypomorphic alleles in embryos reveal a role for Tbx6 in both patterning and specification of the paraxial mesoderm (Chapman et al., 2003; Chapman and Papaioannou, 1998; White et al., 2003). The

enlarged tailbud phenotype of *Tbx6* null embryos is first apparent at e9.5 (Figure 4D). In addition, null mutant embryos display kinked neural tubes, multiple hematomas and vascular anomalies, which eventually lead to embryonic death by e12.5. Like the *wnt3a* null, the *Tbx6* null forms only 7-9 abnormal anterior somites. Ectopic neural tubes form in place of the paraxial mesoderm posterior to the forelimb bud (Figure 3D). Additionally, PS specific markers such as *T, wnt3a,* and *Fgf8* are all expressed in a *Tbx6* null mutant, but paraxial mesoderm markers such as *Notch1, Dll1,* and *paraxis* are not, suggesting that PS cells are established, but are unable to differentiate into PSM (Chapman and Papaioannou, 1998). The levels of cell proliferation and apoptosis in *Tbx6* null embryos is similar to normal littermates, suggesting that Tbx6 is necessary for cells exiting the PS to adopt a paraxial mesoderm fate (Chapman et al., 2003).

Analysis of chimeric embryos consisting of *Tbx6* null ES and wild-type cells revealed that *Tbx6* mutant cells had wide distribution throughout the embryo but did not contribute to the forming somites, and the central portion of the bulbous tailbud consisted almost entirely of mutant cells (Chapman et al., 2003). This suggests a cell-autonomous requirement for differentiation into paraxial mesoderm, and that cells lacking *Tbx6* may be defective in exiting the PS and instead remain within the tailbud.

Examination of Tbx6 hypomorphs lends evidence for a role in patterning the PAM. *rib-vertebrae* (*Tbx6*^{rv}) is a hypomorphic allele of *Tbx6* that is caused by a 185 bp insertion within the regulatory region of *Tbx6* (Theiler and Varnum, 1985; White et al., 2003). This results in expression of less than heterozygous levels of *Tbx6* that ultimately causes malformations of vertebrae and ribs, a shortened trunk, and kinked tail (Figure 4E). Additional evidence for Tbx6's role in patterning the paraxial mesoderm lies in partial transgene rescues of *Tbx6* null embryos. Null embryos harboring a trangene containing the entire *Tbx6* coding region and

upstream sequences necessary to drive a reporter construct in a *Tbx6* specific pattern results in partial rescue of the *Tbx6* null phenotype (White et al., 2003). Similar to the $Tbx6^{rv}$ hypomorphs, these embryos express less than heterozygous levels of *Tbx6*, and display fusions of the ribs and vertebrae. Additionally, defects in expression of *Myogenin* were noted within partially rescued embryos. *Myogenin* is a marker specific for the myotome compartment of the somite. Altogether, this suggests that multiple somite compartments are affected in *Tbx6* hypomorphs. Severity of the phenotype directly corresponds to level of transgene expression, suggesting dosage of this gene is critical during development (White et al., 2003).

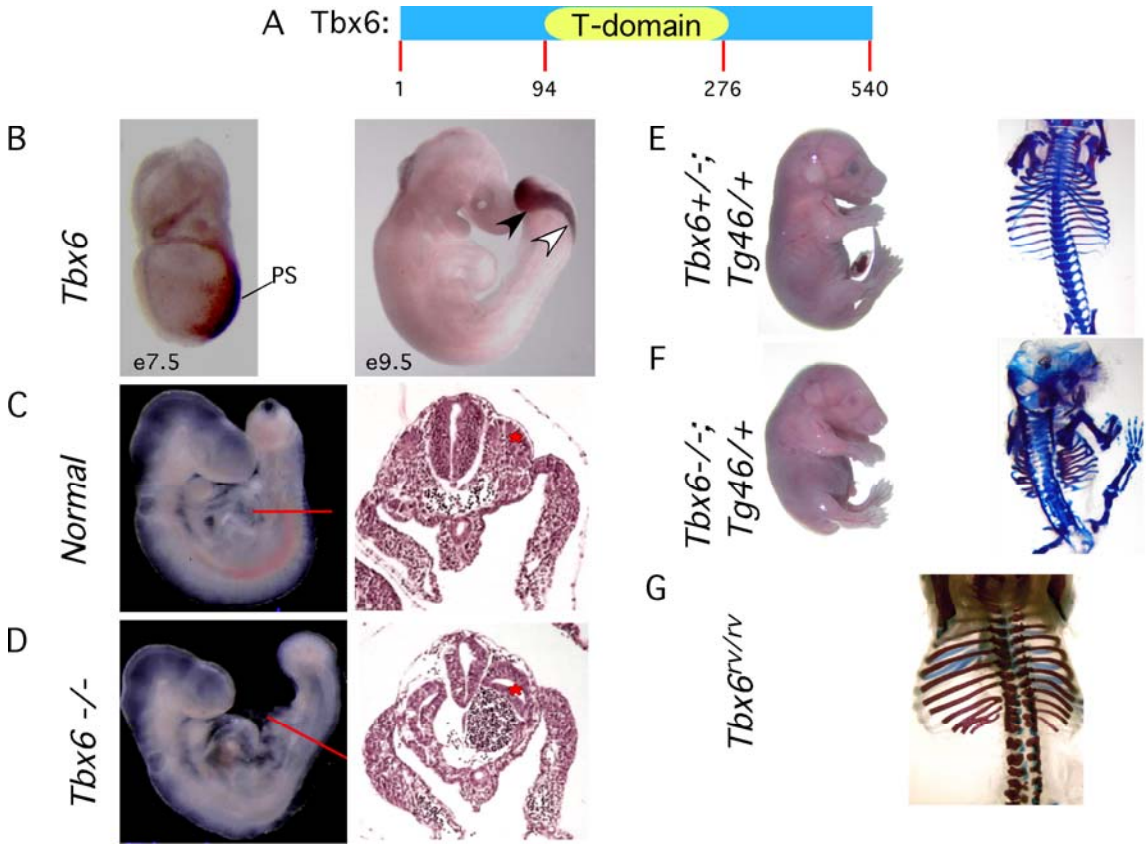

Figure 4: Tbx6 expression patterns and mutant phenotype

(A) Domain structure of Tbx6. (B) *In situ* hybridization showing *Tbx6* expression within the PS and PSM cells lateral to the PS at e7.5, and within the tailbud (closed arrowhead) and PSM (open arrowhead) at e9.5 (Chapman et al., 1996). (C) A normal embryo at e10.5. The red line denotes the level of cross-section in the corresponding histological section. Hemotoxylin and Eosin stained histological section shows epithelial somites (asterisk) flanking either side of the central neural tube (Chapman and Papaioannou, 1998). (D) A bulbous tailbud is apparent in a *Tbx6* null mutant, and histological sections reveal somites are replaced by ectopic neural tubes (asterisk) (Chapman and Papaioannou, 1998). (E,F) Gross morphology of the incomplete rescue of *Tbx6-/-; Tg46/+* as compared to *Tbx6+/-; Tg46/+* reveals shortened tail and axis. Skeletal preparations reveal fused vertebrae and ribs along the entire body axis (White et al., 2003). (G) Skeletal preparation of the Tbx6 hypomorph, *rv/rv*, reveals fusions and malformations of the ribs and vertebrae (White et al., 2003).

1.4.2 Biochemical analyses of Tbx6

EMSA experiments demonstrate that Tbx6 can bind to the same 24 bp palindromic binding sequence and half-site selected by T (White and Chapman, 2005). Tbx6 is able to bind to the palindromic site as a monomer, however upon the addition of increased amounts of Tbx6 relative to the probe, a second shifted band became obvious, suggesting that two Tbx6 monomers can bind simultaneously (White and Chapman, 2005). This is distinct from T which dimerized in the presence of the palindromic sequence (Muller and Herrmann, 1997). Binding as two separate monomers to the palindromic site was also observed for Tbx3 (Coll et al., 2002) (Figure 2). PCR-based binding site selection assays performed with Tbx6 demonstrated a consensus binding sequence for Tbx6 of: 5'-AGGTGTBRNNNN-3' where B is anything but A, and R is a purine. Despite the isolation of the core 5'-AGGTGT-3' sequence common to T-box transcription factors, sequences exactly matching the T half-site were not recovered for Tbx6 (White and Chapman, 2005).

1.4.3 The Notch pathway regulates expression of *Tbx6*

Identification of the *rv* mutation as a *Tbx6* allele highlights the importance of proper regulation of signaling pathways that control expression of *Tbx6* during development. A sequence comparison of *Tbx6* homologs in human, mouse, rat, and dog revealed a conserved region upstream of the start of transcription that contained a binding site for RBP-Jκ, the downstream effector of the Notch signaling pathway (White et al., 2005). Transgenic reporter constructs containing a single point mutation within the RBP-Jκ binding site which abolishes binding

revealed little to no activity of the reporter *in vivo* between e9.5 to e12.5, when the reporter would normally be active (White et al., 2005).

T is expressed earlier than *Tbx6* within the PS, suggesting T may regulate expression of *Tbx6. Tbx6* is initially expressed at e7.0 in a *T* mutant, however as the *T* null phenotype becomes apparent at e8.5 and the PS fails to continue to produce mesoderm, expression of *Tbx6* is lost (Chapman et al., 1996). The presence of T-box binding sites within introns 1 and 5 of *Tbx6* suggests that maintenance of *Tbx6* expression may be dependent on T (Hofmann et al., 2004), although the loss of expression of *Tbx6* in the *T* null also may result from the loss of mesoderm production rather than a direct effect of T on *Tbx6* transcription. It is therefore unresolved whether T directly regulates *Tbx6* expression.

1.4.4 Downstream targets of Tbx6

Four direct downstream targets of Tbx6 are known, including *Dll1*, the ligand for the Notch receptor, placing the Notch signaling pathway both upstream and downstream of Tbx6 (White and Chapman, 2005). The other three targets also encode proteins essential for somitogenesis, and include *Mesoderm Posterior-2,* (*Mesp2*), *Mesogenin1* (*Msgn1),* and *Ripply2.* Interestingly, all of these targets also require input from other signaling molecules for their proper expression, and appear to be a part of a complicated feedback pathway existing within the PSM (Dunty et al., 2008; Hofmann et al., 2004; Wittler et al., 2007; Yasuhiko et al., 2006).

1.4.4.1 *Dll1*

Dll1 is expressed in the PS and PSM, and is subsequently is restricted to the posterior halves of somites (Bettenhausen et al., 1995). The overlapping expression of *Dll1* with *Tbx6,* along with

the absence of *Dll1* in the *Tbx6* mutant, led to the hypothesis that *Dll1* is a direct target of Tbx6 (Chapman et al., 1996). Additionally, *Tbx6*^{rv}/*Tbx6*^{rv} embryos displayed reduced expression of *Dll1* (White et al., 2003). A strong genetic interaction is observed between *Tbx6* and *Dll1*; compound heterozygotes display fusions of the vertebrae that resulted in a kinked tail and occasional missing portions of the proximal ribs (White and Chapman, 2005).

Previous studies identified a 1.5 kb regulatory element from the *Dll1* gene that was capable of driving expression of a minimal promoter-*lacZ* reporter in the PSM beginning at e7.5 (Beckers *et al.*, 2000a). Within this 1.5 kb mesoderm (msd) enhancer are four putative T-box binding sites (BS1-4) located in a 100 bp region; two of which are targets for Tbx6 (White and Chapman, 2005). The enhancer element also contains multiple potential LEF/TCF binding sites that are required for activity of the enhancer element (Hofmann et al., 2004). Loss of either the T-box binding sites or the LEF/TCF binding sites was sufficient to cause loss of expression of a transgenic reporter *in vivo*, suggesting input from both the Wnt signaling pathway and Tbx6 is required for proper expression of *Dll1* and thus Notch signaling (Hofmann et al., 2004).

Altogether, the genetic, transcriptional, and biochemical data indicates that *Dll1* is a direct downstream target of Tbx6 in the PS and PSM.

1.4.4.2 *Mesp2*

The basic helix-loop-helix transcription factor Mesp2 also has a critical role in somitogenesis and establishment of anterior-posterior polarity of the somites (Saga et al., 1997). *Mesp2* is expressed in a dynamic pattern in the anterior PSM, and is down-regulated as the PAM segments. Mesp2 functions to define the anterior border of the forming somite, via Lfng mediated suppression of Notch signaling (Morimoto et al., 2005). *Mesp2* mutant embryos lack

segmented somites that results in fusions of the ribs and vertebrae, a shortened body axis, and perinatal death (Saga et al., 1997).

Transgenic reporter analysis, transcriptional studies, and EMSA experiments identified T-box binding sites essential for expression of Mesp2 (Yasuhiko et al., 2006). Chromatin immunoprecipitation (ChIP) experiments demonstrated that Tbx6 was able to bind to these sites *in vivo* (Yasuhiko et al., 2008). RBP-Jκ binding sites were also found within the *Mesp2* enhancer region, and transcriptional assays demonstrated a synergistic activation of the enhancer upon addition of the NICD and Tbx6. Since the expression of *Mesp2* overlaps only in the anterior limit of Tbx6 expression, this suggests that a threshold amount of Notch signaling in addition to Tbx6 is required for proper *Mesp2* expression (Yasuhiko et al., 2006).

Mesp2 has also been implicated in the ubiquitin-mediated proteosomal degradation of Tbx6. Loss of *Mesp2* results in an expansion of the Tbx6 protein domain, but not *Tbx6* mRNA in the mouse (Oginuma et al., 2008). Furthermore, the T-domain of *Xenopus* Tbx6 directly interacts with the MH2 domain of Smad6, which in turn mediates the ubiquitination and proteosomal degradation of Tbx6 via the E3 ligase, Smurf1 (Smad-ubiquitin regulator factor 1) (Chen et al., 2009). It is unknown how Smurf1 is linked to Mesp2-mediated Tbx6 proteosomal degradation. This suggests that a complicated feedback mechanism exists between Notch signaling, Mesp2, and Tbx6 at the border between the nascent somite and the PSM.

1.4.4.3 *Ripply2*

Work in zebrafish and *Xenopus* identified the Ripply-Bowline-Ledgerline (RBL) family of proteins as being essential for somitogenesis (Kawamura et al., 2005; Kondow et al., 2006). The RBL family of proteins serve as adaptors for the interaction of transcription factors and the Groucho/TLE family of co-repressors. Loss of *Ripply1* in zebrafish leads to improper anterior-

posterior somite patterning, due to an inability to repress *mesp-b*, the *Mesp2* ortholog (Kawamura et al., 2005).

In mouse embryos, *Ripply2* is expressed at e8.5 within the anterior PSM, in a dynamic pattern that exists as either one stripe within S-1, or two stripes, one within S-1 and a weaker stripe in the anterior portion of S0 (Biris et al., 2007). The expression of *Ripply2* is temporally activated in S-1 following *Mesp2* expression. Loss of *Ripply2* function results in persistent expression of *Mesp2*, causing a shortened body axis, rib and vertebral defects, and results in perinatal death (Chan et al., 2007; Dunty et al., 2008). Comparison of the human and mouse enhancer elements of *Ripply2* uncovered conserved putative binding sites for Tbx6 and an E-box binding site specific for Mesp2 (Dunty et al., 2008). Synergistic activation from the enhancer element by Tbx6 and Mesp2 was observed in luciferase transcriptional assays, with the presence of both the E-box and T-box binding sites essential to drive expression from the enhancer element. As was observed in zebrafish, mouse *Ripply2* negatively regulates expression of *Mesp2* in a Groucho-dependent manner (Dunty et al., 2008; Morimoto et al., 2007).

Interestingly, a physical interaction between Tbx6 and mesp-b was necessary for the activation of *Bowline* (*Ripply2* homolog) in *Xenopus* (Hitachi et al., 2008). Whether Mesp2 and Tbx6 directly interact in the mouse is not known. Bowline has also been proposed to mediate a physical interaction between Xgrg-4 (a Groucho/TLE family member) and Tbx6 to negatively regulate expression of *Thylacine-1* (*Thy1)*, an ortholog of *Mesp2*, suggesting that Tbx6 has both activation and repression abilities at the *Thy1* promoter (Kondow et al., 2007). In total, this suggests yet another complicated feedback mechanism occurs between Tbx6, Mesp2, and Ripply2 at the border between the PSM and nascent somite.

1.4.4.4 *Msgn1*

Msgn1 is a basic helix-loop-helix transcription factor that is expressed throughout the PS and PSM, and has a critical role in specification of the PSM. Similar to the *Tbx6* mutant, *Msgn1* null embryos fail to form posterior somites, exhibit a kinked neural tube, an enlarged tailbud, and die *in utero* by e14.5 (Yoon and Wold, 2000). Msgn1 has a critical role in promoting the maturation of the PSM by feeding back to suppress wnt3a signaling, demonstrating another complicated feedback loop during somitogenesis. Msgn1 also acts synergistically with Notch to promote maturation of the PSM via activation of cyclic expression of *Lfng* (Yamaguchi, personal communication). Examination of the enhancer region upstream of *Msgn1* revealed four conserved binding sites for T-box transcription factors and LEF/TCF factors clustered within 1.2 kb of the start of transcription site. Transgenic reporter experiments demonstrated that this region drove expression in the PSM, with mutation of either the T-box binding sites or the LEF/TCF sites abolishing the activity of a transgenic reporter construct *in vivo* (Wittler et al., 2007). This suggests that input from both Tbx6 and the Wnt signaling pathway is required for proper expression of *Msgn1*.

1.4.5 Implications for Tbx6 in human health

Clinical investigation of a 593 kb microdeletion containing *TBX6* has been associated in one patient with mental retardation, developmental delay, hemivertebrae, and rib anomalies (Shimojima et al., 2009). Due to the fact that mice harboring a hypomorphic allele of *Tbx6* display rib and vertebral fusions (White et al., 2005), DNA from this one patient was examined for mutations that may result in a hypomorphic allele, however no mutations within the *TBX6* coding region were identified (Shimojima et al., 2009). The phenotype and associated

microdeletion are suggestive, but not conclusive, of the phenotype being caused by haploinsufficiency of *TBX6*. Because only the coding region was analyzed, it is possible that mutations within the regulatory region have acted to reduce *TBX6* below heterozygous levels.

Currently, no other mutations within *TBX6* are known to cause vertebral disorders. It is possibly that idiopathic cases of congenital vertebral malformations such as spondylocostal and spondlylothoracic dysostosis may be caused by mutations in *TBX6.* Alternatively, mutations in other genes required to regulate levels of *TBX6* could also cause these types of disorders. Given the homozygous null embryonic lethal phenotype observed in the mouse embryo it is likely that similar mutations in humans would result in spontaneous miscarriages.

1.5 AIMS OF DISSERTATION RESEARCH

The objective of my thesis research is to elucidate the role of competition between co-expressed T-box transcription factors. Perceived competition may occur in two manners: first, T-box transcription factors may compete for binding to consensus binding sequences within endogenous enhancers, where the two factors may have inherently different activities. For example, one factor my serve as a transcriptional activator, while another may serve as a repressor. As a second non-exclusive alternative, these factors may compete for binding to a common co-factor. In this way, one factor may reduce the other's transcriptional ability by sequestering a co-factor necessary for its transcriptional ability. Both of these possibilities have been previously observed with other T-box transcription factors, although prior to this work the dynamics between T and Tbx6 within the PS was unknown.

Investigation of this dynamic was accomplished using a combination of tissue culture, biochemical, and *in vivo* genetic approaches, allowing for a more powerful analysis. Data generated *in vivo* lends validity to the *in vitro* results, signifying that the results are most likely not an artifact of the particular assay. The significance of this work extends beyond the obvious further characterization of how T and Tbx6 interact in patterning the mesoderm. We propose that all T-box transcription factors have the ability to compete for downstream targets when co-expressed and that competition may play a role in the phenotypes observed in human birth defects associated with mutations in T-box transcription factors. While this work focuses on the T-box family of transcription factors, it implies that other transcription factors that share a conserved DNA binding domain may also compete for downstream targets *in vivo* when they are co-expressed and that this competition contributes to the overall developmental dynamics of the organism.

2.0 T AND TBX6 COMPETE FOR BINDING AT THE *DLL1* ENHANCER *IN VIVO*

2.1 INTRODUCTION

We hypothesized that T and Tbx6 compete for binding sites in target gene enhancers in the PS and tailbud, and that this competition regulates the PS to PAM transition. Furthermore, we hypothesized that T is responsible for maintaining cells in a PS-like state, while Tbx6 is required for cells to adopt a PAM fate. In this scheme, T and Tbx6 may have both common and unique target genes. We predict that T and Tbx6 have different activities at a subset of common targets due to differing transactivation/repression domains and/or binding affinities.

2.1.1 Genetic interactions between T and Tbx6 suggest that they compete *in vivo*

Understanding the genetic relationship between Tbx6 and T is essential to unraveling how they interact in the formation and patterning of the mesoderm. Although expression of *T* precedes that of *Tbx6* during development, T does not directly regulate initiation of *Tbx6* expression, because *Tbx6* is initially expressed in a *T* null mutant. Expression of *Tbx6* is lost once the *T* mutant phenotype is obvious, suggesting that *T* may be required for the maintenance of *Tbx6* expression, however since *T* mutants fail to maintain a PS, loss of *Tbx6* expression is likely to be an indirect effect of the loss of the PS (Chapman et al., 1996).

Both T and Tbx6 have been implicated in the regulation of *Dll1* expression. It is extremely likely that Tbx6 directly regulates *Dll1* expression *in vivo,* as its mesoderm expression is lost in a *Tbx6* mutant, and a strong genetic interaction is observed in *Tbx6+/-; Dll1+/-* mice, which display kinky tails (Figure 5A) (White et al., 2003). Additionally, Tbx6 is able to bind to two binding sites within the *Dll1* enhancer region in EMSA experiments (White and Chapman, 2005; White et al., 2003). Based on luciferase transcriptional assays, Hofmann et al. suggested that *Dll1* is also a downstream target of T. *Dll1* expression is also lost in a *T* mutant, however the loss of *Tbx6* expression confounds this analysis (Hofmann et al., 2004). Experiments to determine if T is able to physically bind to the identified binding sites within *Dll1* were not performed, leaving the issue of whether *Dll1* is indeed a true target of T unresolved.

Although T and Tbx6 are expressed in a partially overlapping domain within the PS and tailbud, and are able to bind similar sequences *in vitro,* genetic studies have demonstrated that they are not functionally redundant. Heterozygous ES cells with *Tbx6* knocked into the *T* locus ($T^{Tbx6KI}/+$) were injected into a wild-type blastocyst to generate chimeras. Mice with a low percentage of chimerism displayed kinky tails and a shortened trunk, while higher percentage chimeric embryos were not obtained, presumably because chimeras with a high percentage of $T^{Tbx6KI}/+$ cells result in embryonic lethality. Injection of the $T^{Tbx6KI}/+$ ES cells into blastocysts that ubiquitously express GFP confirmed that high percentage chimeras resulted in severe axis truncations leading to embryonic death. This result suggested that Tbx6 is not able to compensate for the loss of one allele of *T*. Since *T* heterozygous mice displayed a short tail phenotype, but were viable, this also suggested that early expression of *Tbx6* within the PS or an increased level of *Tbx6* was deleterious for normal development of the embryo. Phenotypes of chimeric embryos generated with T^{Tbx6KI} ES cells were similar to those generated with *T/T* ES

cells (Beddington et al., 1992). This suggests that excess Tbx6 within the PS is able to compete with T for downstream targets (Chapman, unpublished).

In the converse experiment, *T* was knocked into the *Tbx6* locus (*Tbx6*^{TKI}/+). Again, high percentage chimeras were not obtained, suggesting an increased expression of *T* within the PS and novel expression in the PSM results in embryonic lethality. Chimeras dissected at e10.5 displayed disorganized somites, suggesting expression of *T* within the PSM results in somite anomalies. Chimeras with a low percentage of ES cell contribution (~30%) were viable, and displayed tail kinks. These results also suggest that *T* and *Tbx6* are not functionally redundant and that expression of *T* within the PSM maybe detrimental to the normal development of the embryo (Chapman, unpublished).

Previous genetic analyses with *T* and *Tbx6* were performed with the original *T* mutation that contained a large deletion of several genes, complicating phenotypic analysis of double mutants. To further investigate the genetic relationship between *T* and *Tbx6*, an allele of *T* that deletes the second and third exon ($T^{\Delta 2\text{-}3}$), and therefore the majority of the T-domain, was made. (Kwan, Chapman, and Behringer, unpublished). Embryos that are homozygous for $T^{\Delta 2\text{-}3}$and *Tbx6* reveal that the *T* mutation is epistatic to the *Tbx6* mutation. T^{Wis} homozygotes that were wild-type, heterozygous, or a homozygous null mutant for *Tbx6* revealed no morphological differences between the genotypes; all lacked an allantois and displayed severe axis truncations, indicating that similar to the $T^{\Delta 2\text{-}3}$ allele, T^{Wis} is epistatic to *Tbx6* (Chapman et al., 2003). Interestingly, while the tailbud of *Tbx6-/-* embryos is enlarged, the *Tbx6-/-;* T^{Wis}/+ embryos did not have a bulbous tailbud, suggesting that the dominant-negative nature of the T^{Wis} allele is able to reduce the effectiveness of wild-type *T* allele, thus reducing the tailbud size (Chapman et al., 2003).

An interesting genetic relationship suggestive of competition for binding to common downstream targets exists in *T^{Wis};Tbx6* compound heterozygous mice. While *Tbx6/+* mice are phenotypically normal, and *T^{Wis}/+* mice display a short tail phenotype but no rib or vertebral defects, *Tbx6/+*; *T^{Wis}/+* mice exhibit fusions of the ribs and vertebrae, reminiscent of *$Tbx6^{rv/rv}$* (Figure 5B) (Chapman, unpublished). We hypothesize that this genetic interaction is likely due to the dominant-negative nature of T^{Wis}, which competes with Tbx6 for binding to a common subset of downstream targets shared by Tbx6 and T. In this scheme, T^{Wis} binding to a shared target would result in down-regulation of that particular target, thus phenocopying a Tbx6 hypomorph to elicit this phenotype *in vivo.* These phenotypes are not seen in *Tbx6/+; T/+* embryos.

Additional genetic evidence for a potential competition between T and Tbx6 *in vivo* lies in transgenic embryos (*Tg46*) that over-express Tbx6. The *Tg46* transgene expresses *Tbx6* within its endogenous domain in the PS and PSM at less than heterozygous levels. Homozygosity for the *Tg46* transgene resulted in a truncated axis that displayed a filamentous tail structure, reminiscent of the *T^{Wis}* heterozygous phenotypes (Figure 5C). Most *Tg46/Tg46* embryos die perinatally for unknown reasons, as all internal organs appear normal. The axis truncation phenotype of *Tg46/Tg46* embryos suggests that excess Tbx6 may compete with T in the PS to reduce the ability of T to function in axis extension.

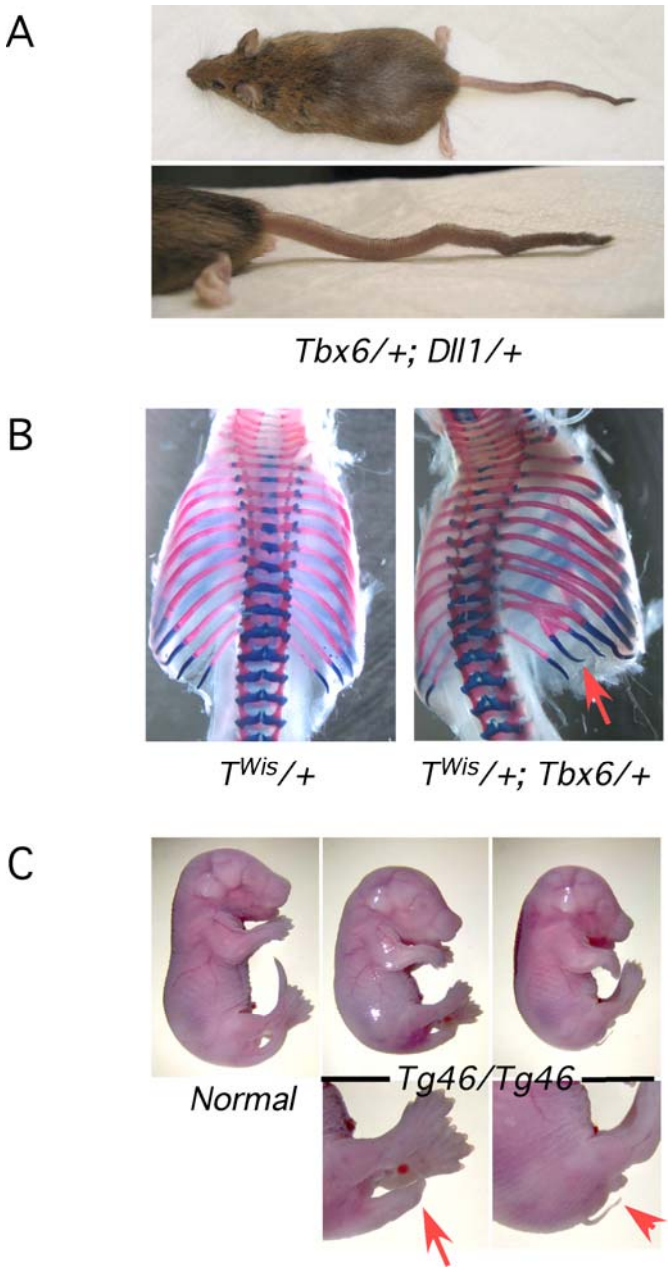

Figure 5: *Tbx6* genetic interactions

(A) A strong genetic interaction is observed in *Tbx6/+; Dll1/+* mice, which display kinky tails, not observed in *Tbx6/+* or *Dll1/+* mice (White and Chapman, 2005). (B) T^{Wis}/+ embryos have a short tail phenotype, but normal ribs and vertebrae, while T^{Wis}/+; *Tbx6/+* embryos display fusions of the ribs (arrow). (C) Embryos that over express Tbx6 within its endogenous domain via the *Tg46* transgene display a truncated body axis, resulting in a shortened tail (arrow) or thin, filamentous tail (arrowhead).

2.1.2 Aims of these studies

The above genetic data suggests T and Tbx6 may compete for a common set of downstream targets *in vivo.* Additionally, both T and Tbx6 have been implicated in regulating expression of *Dll1* (Hofmann et al., 2004; White and Chapman, 2005). I sought to confirm and further characterize this potential competition using a combination of transcriptional and biochemical assays. To begin to unravel these relationships, I performed luciferase transcriptional assays, ChIPs, and quantitative eletrophoretic mobility shift assays (qEMSAs) to determine transcriptional activity and target specificities of T and Tbx6. We chose to use the *Dll1* enhancer region as a representative model system to characterize potential competition between T and Tbx6 for these experiments because *Dll1* is expressed in an overlapping domain (the PS) with both *T* and *Tbx6.* In addition, *Dll1* is a *bona fide* target of Tbx6 with four potential T-box binding sites within the enhancer region, two that Tbx6 can directly bind (White and Chapman, 2005).

2.2 T AND TBX6 DIFFERENTIALLY ACTIVATE ENHANCERS IN LUCIFERASE ASSAYS

It has been previously reported that both T and Tbx6 serve as transcriptional activators. We sought to determine if these two proteins were able to activate transcription at the same level using several enhancers cloned into pGL4.10M luciferase vector (Promega). These include: the 24 bp palindromic T-bind element cloned upstream of the β–globin minimal promoter (T^{bind}–luc), the ~200 bp region of *Dll1-msd* containing four T-box binding sites cloned upstream of the β–globin minimal promoter (*Dll1-msd*-luc), and a 300 bp region of the *Mesp2* enhancer

and promoter (*Mesp2P/E*-luc). I subsequently constructed several versions of Tbx6 and T mammalian expression constructs to characterize the activity of T and Tbx6 at these enhancers. All experiments were performed in human epithelial kidney (HEK) 293T cells, a common cell line for performing luciferase assays with T and Tbx6, as a true "primitive streak" cell line does not exist.

2.2.1 Determining relative levels of T and Tbx6 expression from mammalian expression vectors

HEK293T cells were transfected with an equivalent amount of pCS2mt-Tbx6, pCS3mt-T, or pCS3mt-T- and pCS3mt-Tbx6- Gal4 activation domain (Gal4AD) fusion proteins. To interpret the relative activity of the luciferase reporter, it is essential to know how much protein was made from each construct in HEK293T cells. The relative levels of protein expressed from each construct were determined by Western blotting using an anti-myc antibody (Figure 6A). The N-terminal myc epitope tag allows for comparison of expression levels of these constructs. Lysates were prepared from transfected cells, and equal amounts of total protein (as determined by Bradford dye assay) were loaded onto a 7.5% SDS-PAGE gel. Western blotting was performed using the 9E10 myc antibody. Subsequently, the blot was stripped and re-probed with an anti-tubulin antibody to serve as a loading control. Use of ImageGauge to quantitate the results revealed that myc-Tbx6, myc-T, and myc-Gal4AD-T produce approximately equivalent amounts of protein, while myc-Gal4AD-Tbx6 produced at least 4.5-fold less protein (Figure 6B). The difference in expression level of myc-Gal4AD-Tbx6 may be more dramatic because the other bands may have reached maximum pixel intensity prior to the myc-Gal4AD-Tbx6 band becoming apparent.

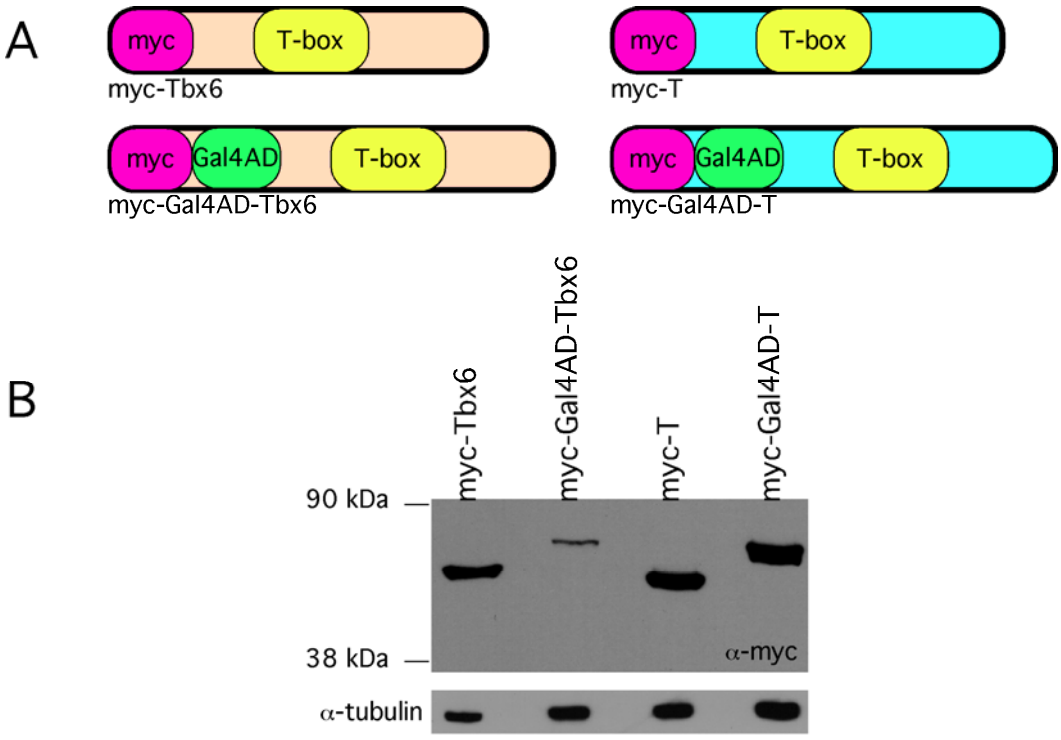

Figure 6: Western blotting reveals relative levels of T and Tbx6 fusion proteins in transfected cells

(A) Depiction of Tbx6 and T constructs for mammalian expression. (B) Western blotting of HEK293T cells transfected with equivalent amounts of each expression construct reveals that myc-Gal4AD-Tbx6 produces approximately 4.5 times less protein than the other constructs, which produce equivalent amounts of protein.

2.2.2 T and Tbx6 act as transcriptional activators

To determine how T and Tbx6 affect transcription, T^{bind}-luc was co-transfected into HEK293T cells with equivalent amounts of myc-Tbx6, myc-T, myc-Gal4AD-T, or myc-Gal4AD-Tbx6. pRL-CMV (Promega), which constitutively expresses Renilla luciferase, was co-transfected to serve as an internal control to normalize for differing levels of transfection or number of cells. Transfections were performed in triplicate in a 96-well plate, and repeated once. Results

depicted are the average of two independent experiments. Relative luciferase units (RLUs) and standard error were calculated over the six data points.

Myc-Tbx6 and myc-T were able to activate transcription weakly from the T^{bind} enhancer, activating at approximately 5.8 and 7.5 fold over background, respectively. While full-length myc-T and myc-Tbx6 activate transcription weakly at approximately equivalent levels, myc-Gal4AD-T activated transcription at 30.1 fold, myc-Gal4AD-Tbx6 activated transcription at 14.4 fold (Figure 7). It should be noted that this difference in activity with the Gal4AD fusion constructs most likely reflects the different expression levels of myc-Gal4AD-Tbx6 and myc-Gal4AD-T protein produced (Figure 6).

Equal levels of activation from myc-Tbx6 and myc-T at this synthetic enhancer was not surprising, as both T and Tbx6 are able to bind to the enhancer, with T binding as a dimer across the two half-sites (Muller and Herrmann, 1997), and Tbx6 binding as two monomers to the two half-sites (White and Chapman, 2005), although the relative affinities for these sites are unknown.

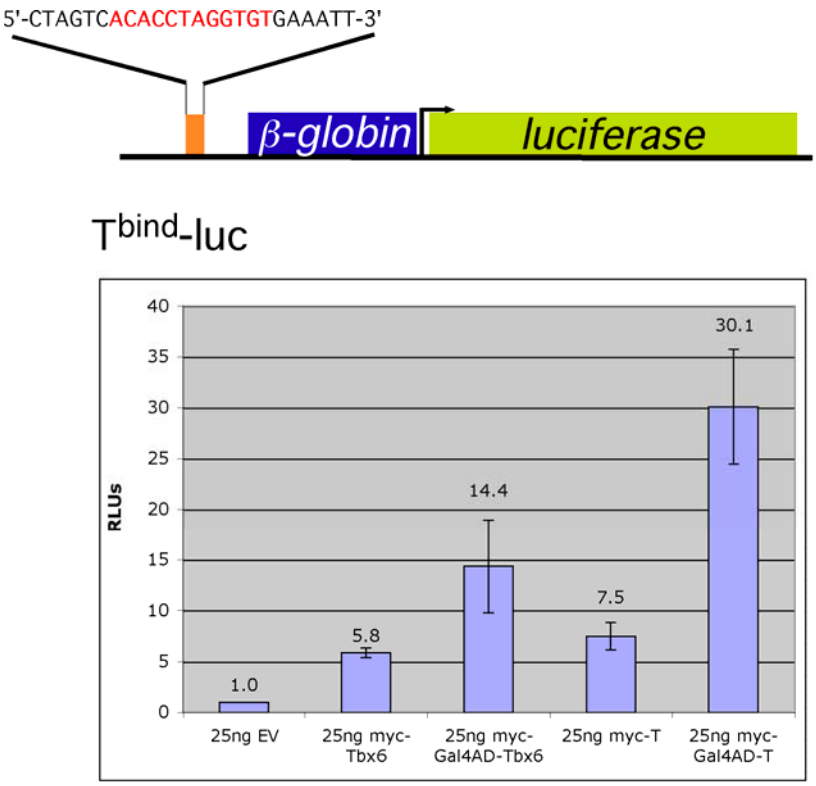

Figure 7: T and Tbx6 act as transcriptional activators

Graphical analysis of relative luciferase units (RLUs) produced from 25ng of indicated protein expression vector co-transfected with 25ng of T^{bind}-luc. The empty protein expression vector (EV) negative control was set to 1.

2.2.3 T and Tbx6 activate transcription from the *Dll1-msd* enhancer at different levels

To further investigate the transcriptional role of T and Tbx6 *in vitro*, we chose to use the *Dll1-msd* enhancer, which contains four T-box binding sites clustered within 100 bp upstream of the β-globin minimal promoter and luciferase reporter (Figure 8A). Contrary to the results obtained with the T-bind enhancer, myc-T and myc-Tbx6 activated at very different levels from the *Dll1-msd* enhancer. Myc-Tbx6 activated at approximately 261-fold over background, while myc-T activated at a 10-fold lower level of 22-fold over background. Myc-Gal4AD-Tbx6 and myc-

Gal4AD-T also showed different activation abilities, with myc-Gal4AD-Tbx6 activating at 2132 fold over background and myc-Gal4AD-T activating at 414 fold (Figure 8B). It is interesting to note that myc-Gal4AD-Tbx6 activates at a higher level than myc-Gal4AD-T when equivalent amounts of expression plasmid was transfected, even though less protein is produced from this construct.

It is not surprising that both T and Tbx6 were able to activate transcription from this enhancer, as *Dll1* is expressed in the overlapping domain with *T* and *Tbx6* (the PS and tailbud), and both T and Tbx6 have been suggested to regulate expression of *Dll1* (Hofmann et al., 2004; White and Chapman, 2005). Our data shows that although both T and Tbx6 are able to activate transcription from *Dll1-msd*, Tbx6 serves as a better activator of transcription than T at this enhancer. Differential activation by T and Tbx6 could occur for several reasons that are not necessarily mutually exclusive. First, Tbx6 may simply be a stronger activator of transcription than T. Secondly, T may require a co-factor for more robust activity that may not be expressed in HEK293T cells. Finally, it is unclear from these experiments whether T is binding to the same binding sites as Tbx6 within the *Dll1-msd* enhancer region, or different ones. T may bind to fewer sites of have a lower affinity for the sites in the *Dll1-msd* as compared to Tbx6. These possibilities are further addressed in section 2.4.

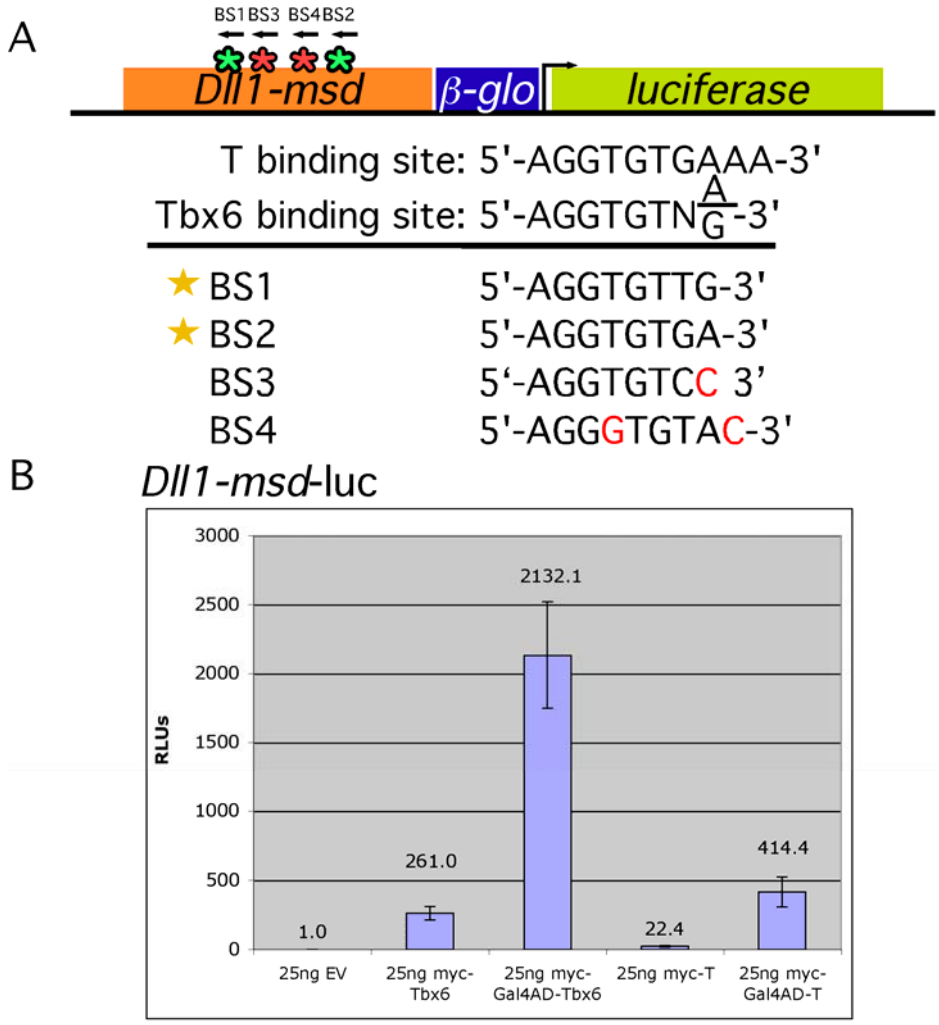

Figure 8: Tbx6 and T activate at different levels from the *Dll1-msd* enhancer

(A) Dll1-*msd* consists of a 200 bp region of the *Dll1* enhancer that contains four T-box binding sites (BS1-4), two of which Tbx6 binds to (green asterisks), and two of which Tbx6 is unable to bind (red asterisks). BS3 and BS4 contain deviations from Tbx6's preferred binding sequence (red letters). The *Dll1-msd* was cloned upstream of the β-globin minimal promoter and luciferase reporter for use in luciferase assays. (B) Graphical analysis of relative luciferase units (RLUs) produced from 25ng of indicated protein expression vector co-transfected with 25ng of *Dll1-msd*-luc. Empty protein expression vector (EV) negative control set to 1.

2.2.4 T and Tbx6 differentially activate from the *Mesp2* enhancer

Mesp2 is a confirmed downstream target of Tbx6, and is expressed within the anterior portion of the PSM as the somite segments (Yasuhiko et al., 2006). Although *Dll1* is expressed in an overlapping domain with *T* and *Tbx6*, *Mesp2* is co-expressed only with *Tbx6.* I therefore sought to determine how T and Tbx6 acted at this enhancer. The *Mesp2E/P* enhancer contains four putative T-box binding sites within 300 bp upstream of the start of transcription, and the endogenous promoter sequences (Figure 9A).

These studies revealed that myc-T activated transcription from this construct at a 10-fold lower level than myc-Tbx6 (71.6 and 713.6 fold activation levels, respectively), similar to the activation levels from the *Dll1-msd* enhancer. Myc-Gal4AD-Tbx6 activated transcription 3146.7 fold while myc-Gal4AD-T activated transcription at 1322.3 fold (Figure 9B). This trend was very similar to that observed with at the *Dll1-msd* enhancer, which is surprising since *Mesp2* is most likely exclusively a Tbx6 target, as T is not expressed in the same domain as *Mesp2.* Activation of the *Mesp2* enhancer by myc-T may not reflect a physiologically relevant event, since *T* is not expressed in the anterior PSM. However, T may bind the *Mesp2* enhancer in the PS/tailbud and serve to block Tbx6 from binding. It is likely that Tbx6's activation of the reporter construct reflects the endogenous situation, as chromatin immunoprecipitation experiments have demonstrated that Tbx6 binds to the *Mesp2* enhancer *in vivo* (Yasuhiko et al., 2006).

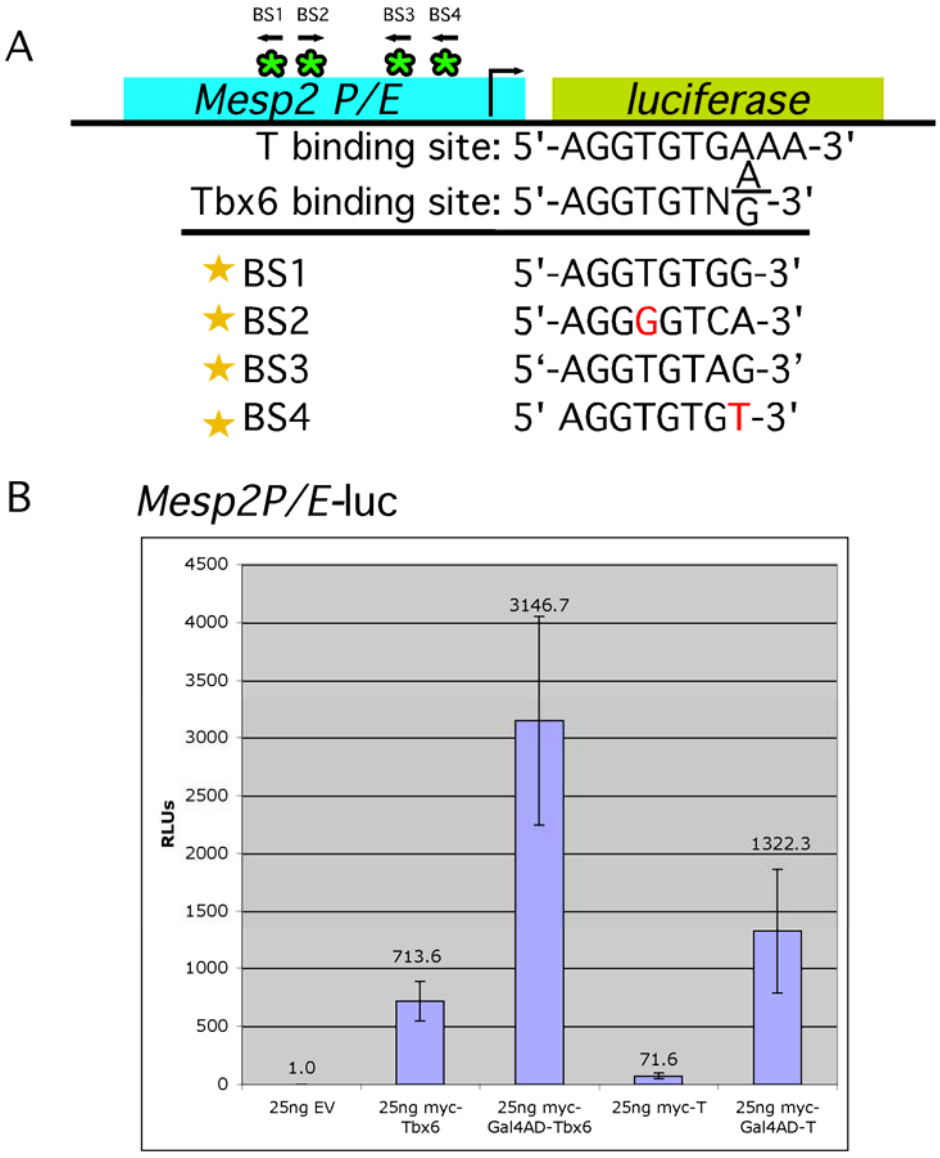

Figure 9: T and Tbx6 activate transcription from the *Mesp2* promoter/enhancer

(A) A 300 base-pair region of the *Mesp2* promoter/enhancer (P/E) containing four T-box binding sites (BS1-4) was cloned upstream of the luciferase reporter. Bases that deviate from Tbx6's consensus are red. Tbx6 can to bind to all four binding sites in EMSA experiments (Yasuhiko et al., 2006; Yasuhiko et al., 2008). (B) Graphical analysis of relative luciferase units (RLUs) produced from 25ng of indicated protein expression vector co-transfected with 25ng of *Mesp2P/E*-luc. Empty protein expression vector (EV) negative control set to 1.

2.3 T REDUCES TBX6'S TRANSCRIPTIONAL ABILITY

The above luciferase assays demonstrate that T and Tbx6 differentially activate transcription from the *Dll1-msd* enhancer and the *Mesp2P/E* reporter constructs, but not from the T^{bind} palindromic enhancer. We therefore hypothesized that addition of T to a constant amount of Tbx6 would compete for binding sites within the enhancers and effectively reduce the activity of Tbx6 from the *Dll1-msd* and *Mesp2P/E* reporter constructs, but not from the T^{bind} enhancer. Competition luciferase assays were performed to test this hypothesis, whereby increasing amounts of myc-T were added to constant amounts of myc-Tbx6 and the relative luciferase reporter activity was measured. To determine if this was a feasible method for observing competition between T and Tbx6, we first constructed a dominant-negative version of Tbx6 that should be able to compete with wild-type Tbx6 in luciferase assays to test the efficacy of our competition luciferase assays.

2.3.1 Dominant-negative Tbx6 is nuclear localized and able to bind DNA

To first test the efficacy of competition luciferase experiments, we designed a putative dominant-negative version of Tbx6. Most T-box transcription factors contain their activation/repression domains within the C-terminus of the protein (Conlon et al., 2001). We therefore created a version of Tbx6 with a C-terminal truncation after the T-domain, with an N-terminal myc-tag and nuclear localization sequence (NLS), myc-$Tbx6^{NLS\Delta C}$ (Figure 10A). We searched for canonical NLS sequences within Tbx6, but were unable to identify any. Without an added NLS myc-$Tbx6^{\Delta C}$ did not consistently localize to the nucleus (data not shown), implying that Tbx6's nuclear localization sequence may lie within the C-terminus of the protein. In contrast,

HEK293T cells transfected with myc-Tbx6$^{NLS\Delta C}$ and processed for immunofluorescence with anti-myc and our N-terminal anti-Tbx6 antibody revealed co-localization of the myc-Tbx6$^{NLS\Delta C}$ protein and the TO-PRO3 nuclear stain in 100% of the cells, indicating that myc-Tbx6$^{NLS\Delta C}$ is nuclear localized (Figure 10A).

Western blotting performed using HEK293T cells transfected with myc-Tbx6$^{NLS\Delta C}$ revealed that the myc-Tbx6$^{NLS\Delta C}$ protein had a higher molecular weight than the predicted size of 47kDa. Sequencing of the myc-Tbx6$^{NLS\Delta C}$ construct confirmed proper cloning and presence of several stop codons at the C-terminal end. Additionally, Western blotting of cells transfected with myc-Tbx6$^{NLS\Delta C}$ as compared to myc-Tbx6$^{\Delta C}$ (lacking the NLS) revealed that the myc-Tbx6$^{\Delta C}$ protein ran much closer to the expected size, indicating the charged nature of the NLS (consisting of mainly lysines and arginines) may retard proper migration within the SDS-PAGE (data not shown). Although myc-Tbx6$^{NLS\Delta C}$ may run much higher than expected on a Western blot, we are confident that we are producing the appropriate protein within the cells.

We predicted that myc-Tbx6$^{NLS\Delta C}$ would still be able to bind DNA, as it retains its T-box DNA binding domain. To test this, I prepared nuclear extracts from cells transfected with myc-Tbx6, myc-Tbx6$^{NLS\Delta C}$, or untransfected cells. These nuclear extracts were used in EMSA experiments with radiolabeled T^{bind}, T^{half} (T-half site binding sequence), or T^{mut}, a mutated verson of T^{bind}. Both myc-Tbx6$^{NLS\Delta C}$ and full-length myc-Tbx6 were able to bind to T^{bind} and T^{half}, and only weakly to T^{mut}. Super-shift with anti-Tbx6 antibody further demonstrated that the presence of myc-Tbx6 and myc- Tbx6$^{NLS\Delta C}$ within the nuclear lysates were causing the observed shifting of the radioactive probe (Figure 10C). Binding to T^{mut} was likely due to the presence of one intact binding site within the palindrome (see section 6.4 for sequences of oligonucleotides used).

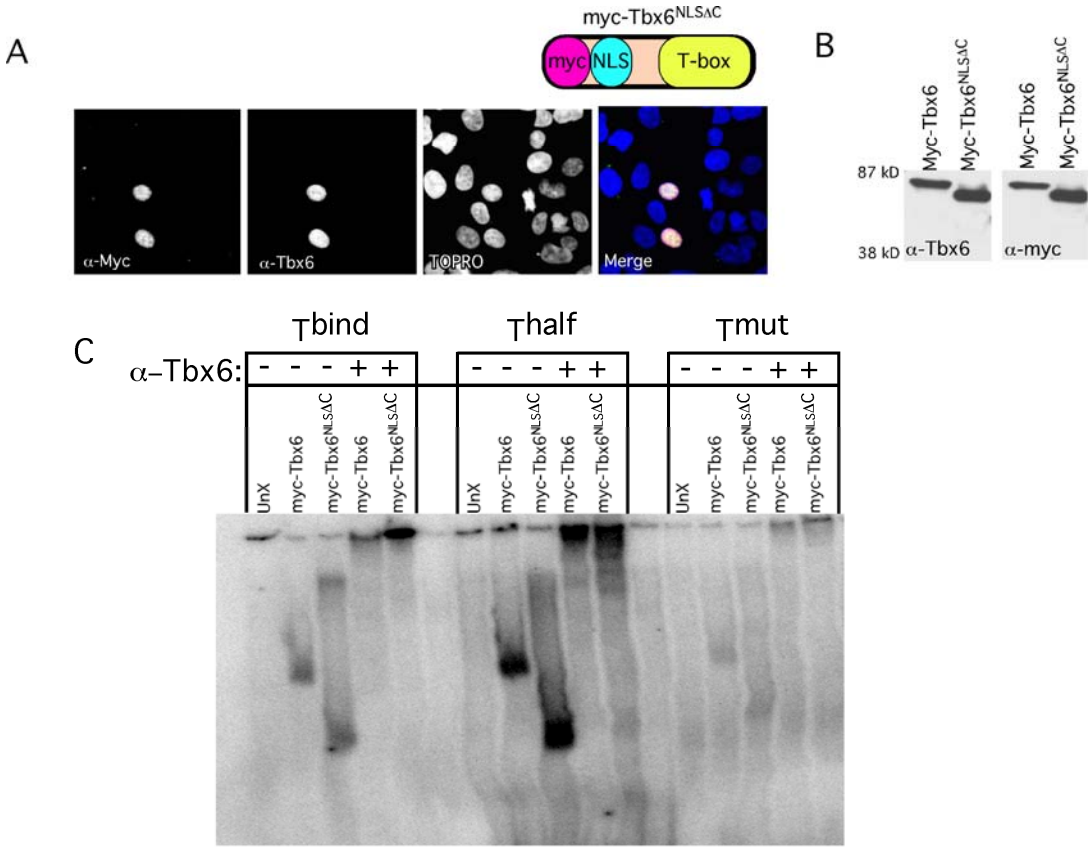

Figure 10: Myc-Tbx6$^{NLS\Delta C}$ is localized to the nucleus and can bind to DNA.

(A) The myc-tagged Tbx6$^{NLS\Delta C}$ with an N-terminal nuclear localization sequence (NLS) was transfected into HEK293T cells. Immunofluorescence was performed with anti-Tbx6 and anti-myc antibodies. Tbx6$^{NLS\Delta C}$co-localized with the TOPRO DNA stain in the nucleus. (B) Western blotting reveals that myc-Tbx6$^{NLS\Delta C}$ runs at a higher molecular weight than the expected size of 47 kDa. The proteins were detected with both anti-Tbx6 and anti-myc antibodies. (C) Nuclear extracts from HEK293T cells were used in EMSAs with radiolabeled T^{bind} (palindromic T site), T^{half} (T half-site), and a mutated version of T^{bind}, T^{mut}. Both myc-Tbx6 and myc-Tbx6$^{NLS\Delta C}$ were able to bind to the palindromic T^{bind} and half-site (T^{half}), and weakly to the mutated site. A shift was not observed using nuclear extracts from untransfected (UnX) cells. Super-shifts with anti-Tbx6 antibody further demonstrated that the presence of myc-Tbx6 and myc-Tbx6$^{NLS\Delta C}$ within the nuclear lysates were causing the observed shifting of the radioactive probe.

2.3.2 A dominant-negative version of Tbx6 decreases the transcriptional ability of full-length Tbx6

Part of our goal was to create a dominant-negative version of Tbx6 that was still able to bind to DNA, but not transactivate. Since many T-box transcription factors contain their activation and/or repression domains within the C-terminus of the protein (Conlon et al., 2001), we therefore hypothesized that the activation domain of Tbx6 would be also contained within its C-terminus. To confirm this, I tested the transcriptional activity of myc-Tbx6$^{NLS\Delta C}$ in luciferase assays at the *Dll1-msd* enhancer. Consistent with previous observations for other T-box factors, myc-Tbx6$^{NLS\Delta C}$ activated at a mere four-fold over background as compared to myc-Tbx6 and myc-Gal4AD-Tbx6 which activated at 1541-fold and 3931-fold over background, respectively, thus demonstrating a loss of transcriptional activity by myc-Tbx6$^{NLS\Delta C}$. Addition of increasing amounts of myc-Tbx6$^{NLS\Delta C}$ (range of 5ng-50ng) to a constant 25ng of myc-Tbx6 or myc-Gal4AD-Tbx6 resulted in a linear decrease in RLUs, suggesting that myc-Tbx6$^{NLS\Delta C}$ is able to compete for binding at the *Dll1-msd* enhancer (Figure 11). Although it appears myc-Tbx6$^{NLS\Delta C}$ is better at competing with myc-Gal4AD-Tbx6, myc-Gal4AD-Tbx6 is produced at ~4 fold lower than myc-Tbx6 and myc-Tbx6$^{NLS\Delta C}$, which most likely accounts for the observed effect (see Figure 6B, 10B).

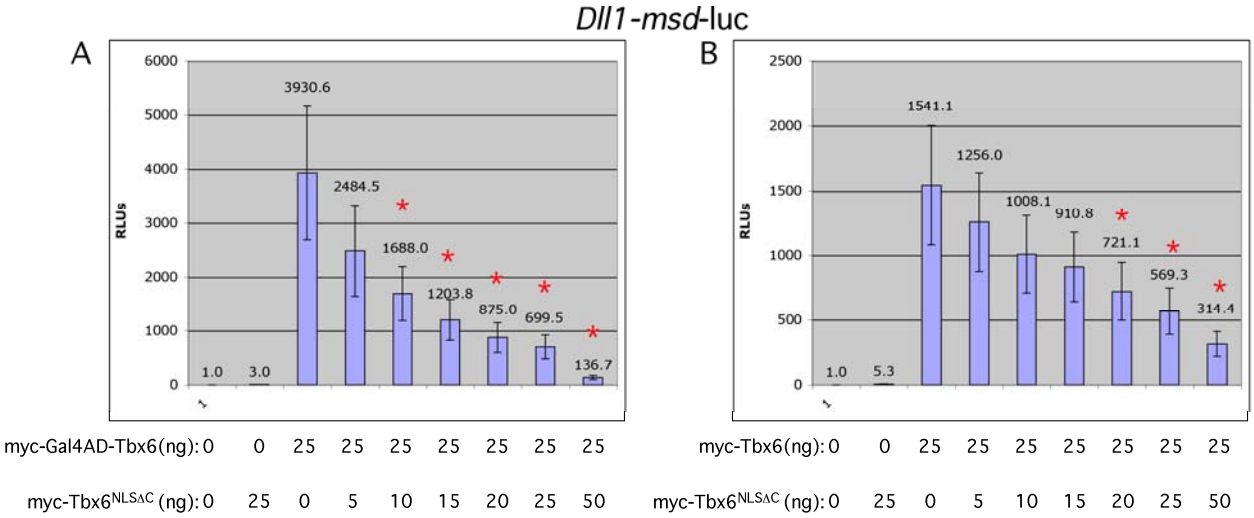

Figure 11: Increasing amounts of myc-Tbx6$^{NLS\Delta C}$ lowers the transcriptional activity of myc-Tbx6 and myc-Gal4AD-Tbx6.

(A) Competition luciferase assay adding increasing amount of myc-Tbx6$^{NLS\Delta C}$ to constant amounts of myc-Tbx6 (A) or myc-Gal4AD-Tbx6 (B). 25ng *Dll1-msd*-luc co-transfected for all data points. Red asterisks above bars indicate $p<0.05$.

2.3.3 T decreases Tbx6's transcriptional activity at *Dll1-msd* and *Mesp2P/E*, but not at T^{bind}

Previously described genetic data (section 2.1.1) led us to hypothesize that T competes for binding at the enhancers of common target genes *in vivo*. Data obtained with myc-Tbx6$^{NLS\Delta C}$ demonstrated the efficacy of our competition luciferase assays, allowing us to take a similar approach to determine if T could compete with Tbx6 at *Dll1-msd*, *Mesp2P/E*, and T^{bind} to effectively lower the transcriptional activity of Tbx6. Competition luciferase assays were performed as described above whereby increasing amounts of myc-T (5ng-50ng range) were added to a constant amount of myc-Tbx6 or myc-Gal4AD-Tbx6 using the *Dll1-msd*, *Mesp2P/E*, or T^{bind} luciferase constructs. Addition of increasing amounts of myc-T to a steady amount of

myc-Tbx6 or myc-Gal4AD-Tbx6 resulted in a decrease in luciferase units from the *Dll1-msd* and *Mesp2P/E* luciferase reporters, but not the T^{bind} reporter (Figure 12). Myc-T may not effectively compete to lower luciferase units at the T^{bind} enhancer simply because the activity levels of myc-T and myc-Tbx6 at this enhancer are approximately equivalent. At T^{bind}, a somewhat additive effect is noted, whereby a combination of 25ng of myc-T and 25ng of Tbx6 resulted in 9.6 fold activation over background, which is slightly less than the sum of the 5.4 and 5.8 fold activation of 25ng of myc-T and myc-Tbx6 on their own, respectively (Figure 12E). To determine if myc-T could reduce the activity level of myc-Gal4AD-Tbx6 at the T^{bind} enhancer, increasing amounts of myc-T were added to a constant amount of myc-Gal4AD-Tbx6. Increasing amounts of myc-T reduced the RLUs from the T^{bind} enhancer (Figure 12F), suggesting that since the activity level of myc-T is considerably lower than either myc-Tbx6 or myc-Gal4AD-Tbx6 at the *Dll1-msd* and *Mesp2P/E* enhancers, increasing amounts of T could lower Tbx6 transcription ability simply by displacing either myc-Tbx6 or myc-Gal4AD-Tbx6 from the enhancer. Again, in this case, the dramatic drop in luciferase units observed when 5ng of myc-T is added as a competitor to myc-Gal4AD-Tbx6 is likely due to the low levels produced by the myc-Gal4AD-Tbx6 plasmid.

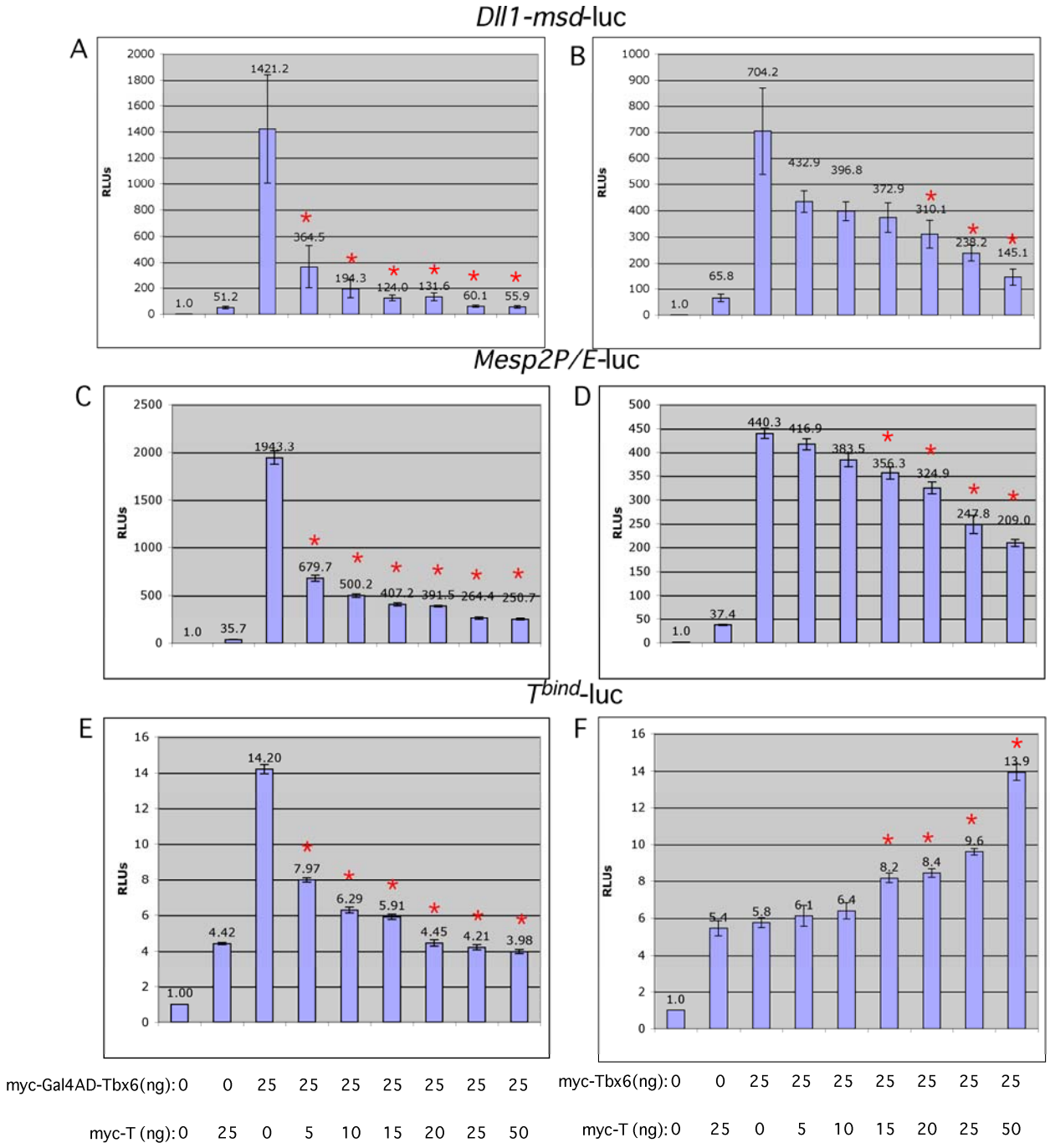

Figure 12: myc-T competes with myc-Tbx6 and myc-Gal4AD-Tbx6 at *Dll1-msd* and *Mesp2P/E* but not T^{bind}.

(A) Competition luciferase assay adding increasing amounts of myc-T to a constant amount of myc-Gal4AD-Tbx6 (A,C,E) or myc-Tbx6 (B,D,F) to 25ng of *Dll1-msd*-luc (A,B), *Mesp2P/E*-luc (C,D) or T^{bind}-luc (E,F). Red asterisks above bars indicate $p<0.05$.

2.3.4 T^{Wis} competes with both Tbx6 and T

T^{Wis} encodes a truncated T protein (Figure 13A), which is thought to act as a dominant-negative since it can still bind to DNA but cannot transactivate (the *T^{Wis}* allele is discussed in more detail in section 1.3.4). We therefore PCR amplified the region of *T* encoding the T^{Wis} protein and cloned it into the pCS3mT vector, allowing for expression of a myc-tagged fusion protein in mammalian cells. Immunofluorescence performed with HEK293T cells transfected with the pCS3-myc-T^{Wis} construct and stained with anti-T and anti-myc antibodies revealed that the myc-T^{Wis} protein is nuclear localized (Figure 13B). Western blotting of equivalent amounts of total protein from lysates of transfected HEK293T cells with an anti-myc and anti-T antibodies revealed the myc-T^{Wis} construct produced a protein that ran at a slightly higher molecular weight than the predicted size (~50 kDa) (Figure 13C). Despite this, we are confident we are making the appropriate protein as the construct has been verified by sequencing.

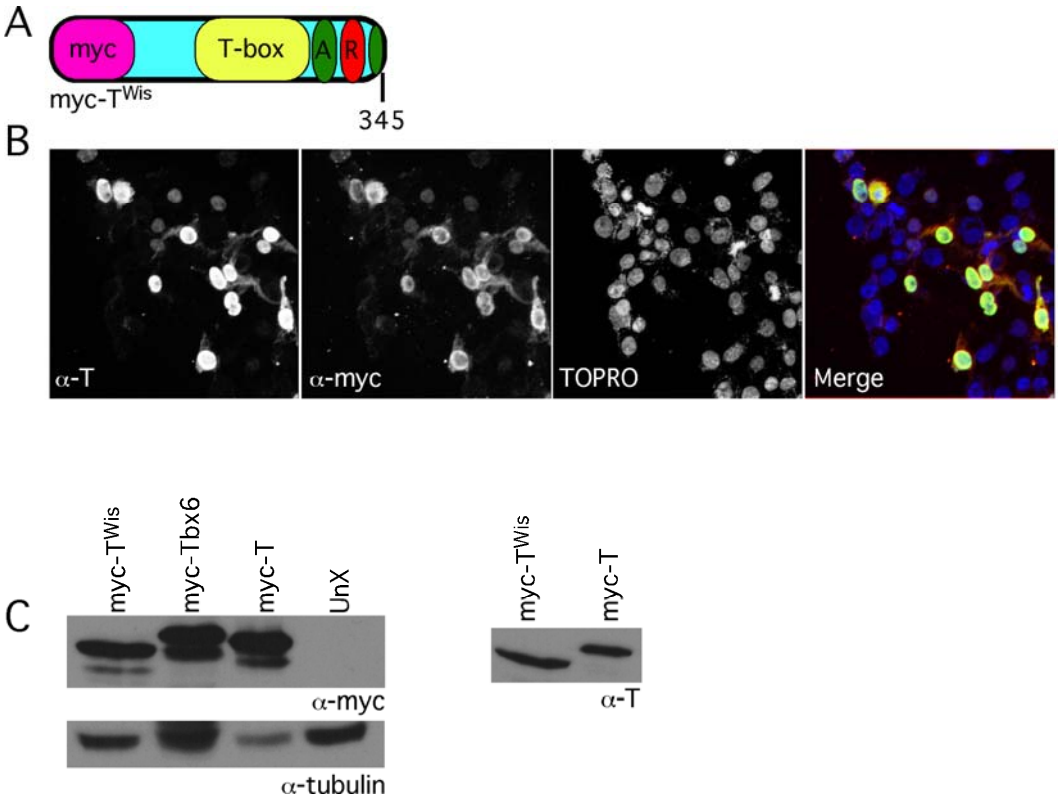

Figure 13: myc-T^{Wis} is nuclear localized in transfected HEK293T cells.

(A) Depiction of the myc-T^{Wis} construct. The T^{Wis} protein contains a stop codon after amino acid 345. (B) Immunofluorescence of transiently transfected HEK293T cells and stained with anti-T (green) and anti-myc (red) antibodies reveals co-localization of myc-T^{Wis} with the nuclear stain, TOPRO (blue). (C) Western blotting of lysates prepared from transiently transfected HEK293T cells. The myc-T^{Wis} protein can be detected with both an anti-myc and anti-T antibody. UnX = untransfected.

To determine whether T^{Wis} can indeed act as a dominant-negative by interfering with T and Tbx6's ability to bind enhancers, we performed competition luciferase assays using the *Mesp2P/E* luciferase construct. On its own, myc-T^{Wis} did not significantly activate transcription, (2.5-fold over background) as compared to 139.2 and 49.2 for myc-Tbx6 and myc-T, respectively (Figure 14). Addition of increasing amount of myc-T^{Wis} to constant amounts of either myc-T (Figure 14A) or myc-Tbx6 (Figure 14B) effectively decreased the activity level from the *Mesp2P/E* enhancer.

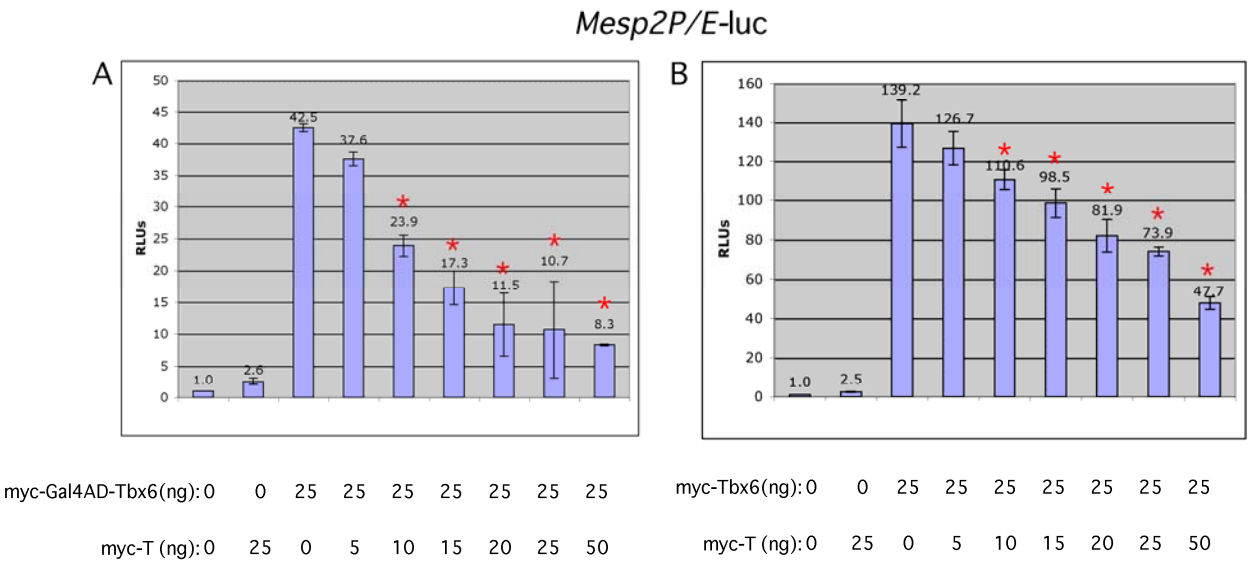

Figure 14: myc-T^{Wis} is unable to activate transcription from *Mesp2P/E* and lowers transcriptional activity of myc-T and myc-Tbx6.

(A) Competition luciferase assay adding in increasing amounts of myc-T^{Wis} to a constant amount of myc-Gal4AD-Tbx6 (A) or myc-Tbx6 (B) to 25ng of the *Mesp2P/E*-luc. Red asterisks above bars indicate $p<0.05$.

2.4 CHROMATIN IMMUNOPRECIPITATION EXPERIMENTS DEMONSTRATE BOTH T AND TBX6 BIND TO THE *DLL1-MSD* ENHANCER *IN VIVO*

The above luciferase assays suggest that T and Tbx6 can compete at the *Dll1-msd* and *Mesp2P/E* enhancers. There are several possible ways that T and Tbx6 can compete at the *Dll1-msd* and *Mesp2P/E* enhancers including at the level of DNA binding or by competing for common co-factors. We hypothesized that the observed competition at least partially occurs at the level of DNA binding. To test this hypothesis, we performed chromatin immunoprecipitation (ChIP) to determine whether T and Tbx6 are bound to the *Dll1-msd* enhancer *in vivo*.

2.4.1 Tbx6 binds at the *Dll1-msd* enhancer in cultured cells

As a preliminary experiment, fibroblasts were derived from our three-component embryos that inducibly express myc-Tbx6 (see Chapter 3 for full details on three-component mice). The fibroblasts were grown in the presence of doxycycline (DOX) to induce expression of myc-Tbx6, while cells grown without DOX (and thus not expressing myc-Tbx6) served as a negative control. Immunoprecipitations were performed with the cross-linked DNA/proteins and our N-terminal anti-Tbx6 antibody, as well as no antibody controls. After purification of bound fragments, radioactive PCR was performed with primers designed to amplify a 358 base-pair region flanking all four T-box binding sites in the *Dll1* enhancer. Primers for a genomic region of *Wnt7a* served as a negative control and should amplifiy a 208 bp fragment in input chromatin only. ChIP in this cell culture based assay revealed that Tbx6 was bound to the *Dll1-msd* region, but not to *Wnt7a* (Figure 15). This suggested that Tbx6 could bind to the *Dll1-msd* enhancer *in vivo*, although over-expression of myc-Tbx6 in this cell culture system may allow for Tbx6 to bind to sites it would not normally bind to under physiological conditions.

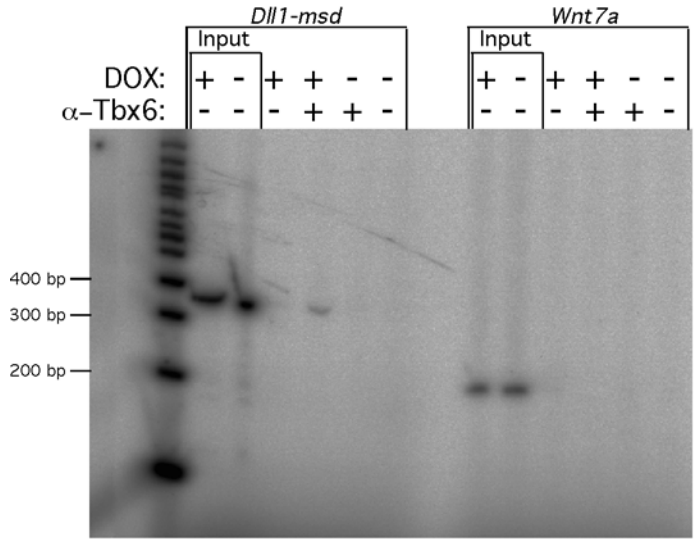

Figure 15: ChIP demonstrates Tbx6 binds to the *Dll1-msd* enhancer region in cultured cells

ChIP from fibroblasts inducibly expressing myc-Tbx6 reveals that myc-Tbx6 bound to the *Dll1-msd* enhancer region, but not to an unrelated control genomic region (*Wnt7a*). Amplification of the *Dll1-msd* enhancer region is observed only in cells treated with doxycycline (DOX) to induce expression of myc-Tbx6 and anti-Tbx6 antibody.

2.4.2 T binds to the *Dll1-msd* in transfected tissue culture cells

NIH3T3 cells transfected with pCS3T.10, served as source of cells expressing T, with untransfected cells serving as a negative control. NIH3T3 cells were used in these experiments because they are a readily transfectable mouse fibroblastic cell line. Immunoprecipitations were performed with anti-T and radioactive PCR was performed as described above. ChIP in this cell culture based assay revealed that T was also able to bind within the *Dll1-msd* region, but not within *Wnt7a*, the unrelated genomic control (Figure 16). This suggests that T is able to bind to the *Dll1-msd* enhancer *in vivo.* However, over expression may result in promiscuous binding in this experiment.

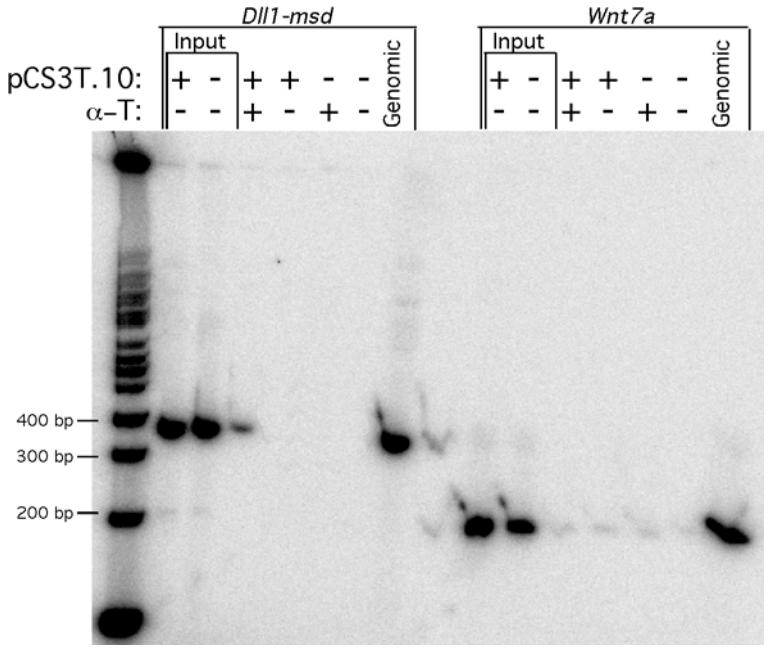

Figure 16: ChIP demonstrates T can bind to the *Dll1-msd* region

ChIP was performed using T antibody and chromatin prepared from NIH3T3 cells transfected with pCS3T.10. Untransfected cells served as a negative control. Amplification of the *Dll1-msd* enhancer region is observed only in cells transfected with pCS3T.10 and anti-T antibody.

2.4.3 ChIP experiments using tailbuds confirm T and Tbx6 bind to the *Dll1-msd in vivo*

Over expression of T and Tbx6 in the above cell culture ChIP experiments may result in promiscuous binding. Therefore, we sought to repeat these experiments using dissected tissue expressing endogenous T and Tbx6 as a source for our ChIP experiments. Use of an endogenous tissue allows us to conclusively determine whether or not both T and Tbx6 bind to the *Dll1-msd* region *in vivo*. Tailbud tissue, which expresses T and Tbx6, was dissected from approximately fifty e10.5 wild-type embryos and used as a source of chromatin for ChIP assays. An equivalent amount of somitic tissue, which does not express Tbx6 or T, was used as a negative control. Immunoprecipitations were performed as described above. Use of endogenous tailbud tissue

revealed that both T and Tbx6 bound to the *Dll1-msd* under physiological conditions during mouse development (Figure 17).

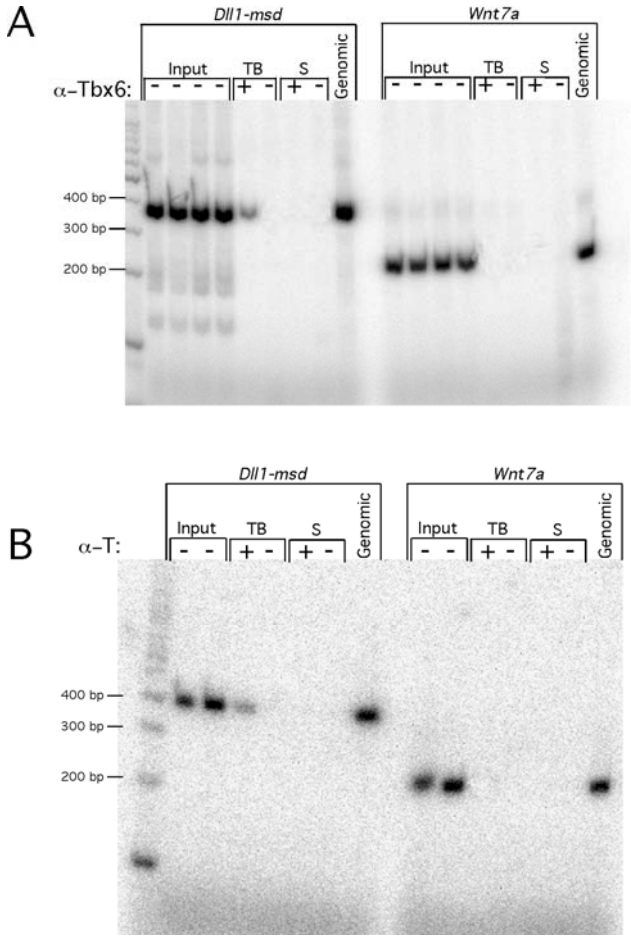

Figure 17: ChIPs performed with embryonic dissected tissues demonstrate that both Tbx6 and T bind to the *Dll1-msd* enhancer *in vivo.*

Chromatin isolated from e10.5 tailbud tissue (TB), and somitic tissue (S) was used for ChIP. A PCR amplified fragment corresponding to *Dll1-msd* was observed in ChIPs performed with TB tissue with anti-Tbx6 (A) and anti-T antibody (B) but not in somitic tissue or when antibody was not added. The *Wnt7a* fragment was only amplified in input samples and not under any other conditions. Genomic DNA was used as a positive control for the PCR reaction.

2.5 T AND TBX6 HAVE DIFFERENT AFFINITIES FOR THE BINDING SITES WITHIN *DLL1-MSD*

As both T and Tbx6 bind within the *Dll1-msd* region, we questioned if T and Tbx6 were able to bind to the same sites, thus revealing a mechanism for competition. It has been previously shown that Tbx6 binds to BS1 and BS2, but not BS3 or BS4 (White and Chapman, 2005). We hypothesized that T could bind to at least one of the same binding sites as Tbx6, thus resulting in the observed competition in our luciferase assays. To test this hypothesis, I performed EMSA experiments using radiolabeled oligonucleotides corresponding to each of the four *Dll1-msd* binding sites and purified recombinant DBDs of Tbx6 and T. These purified proteins allowed us to determine the affinity of each protein for the individual binding sites.

2.5.1 Tbx6-DBD and T-DBD have different binding preferences within *Dll1-msd*

EMSAs were performed by adding 100ng (400 nM final concentration) of purified His-Tbx6-DBD or His-T-DBD was added to radioactively end-labeled double-stranded oligonucleotides (40 μM final concentration) corresponding to each of the four T-box binding sites within *Dll1-msd* plus ~8-10 flanking nucleotides (Figure 18A). As expected, Tbx6-DBD bound to both BS1 and BS2. In contrast, T-DBD bound to BS2, but only weakly if at all to BS1. Neither Tbx6-DBD nor T-DBD bound to BS3 or BS4 (Figure 18B). Because the oligonucleotide probe was included in excess, the shifted band observed most likely corresponds to one monomer of His-Tbx6-DBD or T-DBD bound. A second, higher molecular weight band (indicating possible dimerization of the DBDs) was not observed under the conditions tested in this experiment.

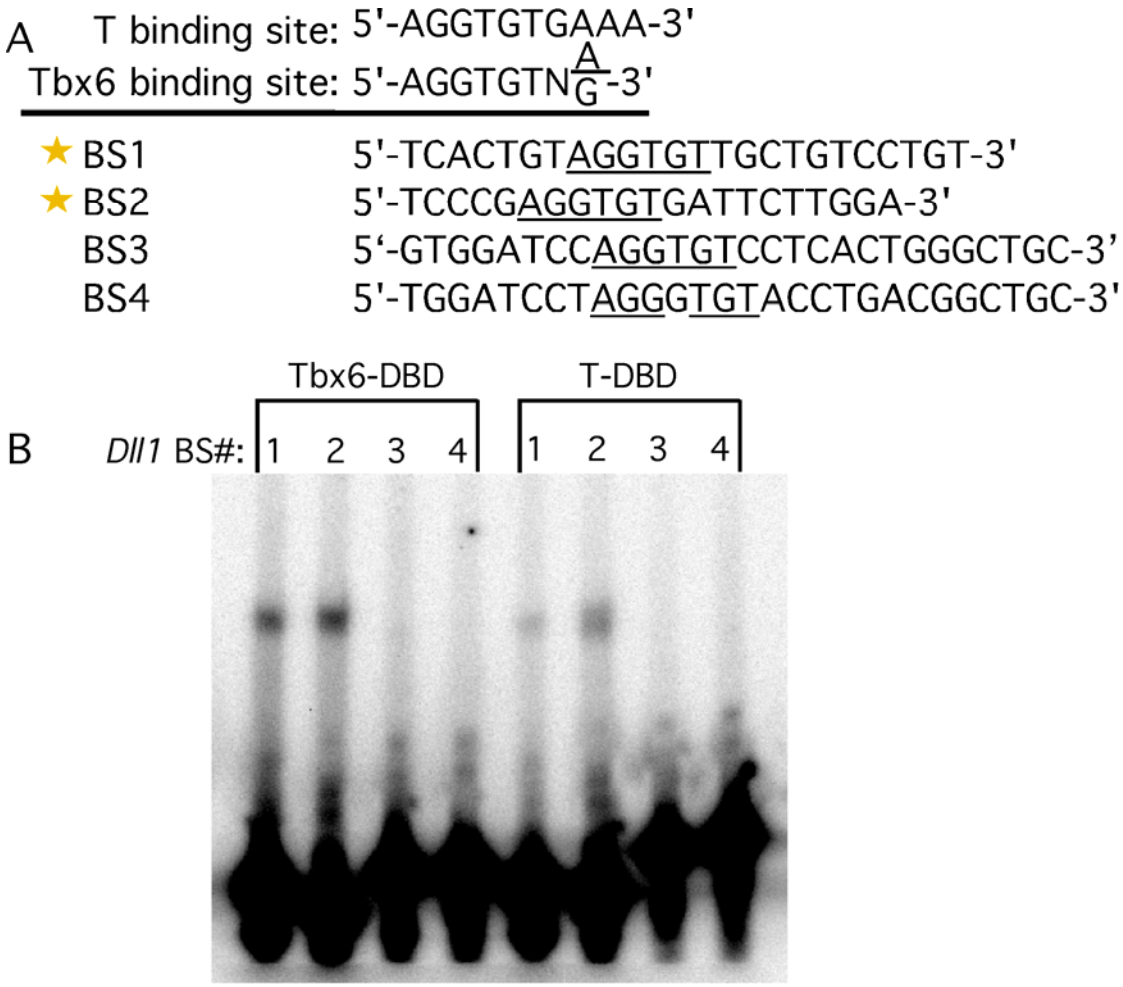

Figure 18: His-Tbx6-DBD and His-T-DBD have different binding preferences within *Dll1-msd*

(A) Tbx6 and T's binding preferences and corresponding sequences of BS1-4 in *Dll1-msd* used for EMSA. Putative T-box binding sites are underlined. (B) His-Tbx6-DBD or T-DBD was added to double-stranded radiolabeled probes corresponding to BS1-4 and reactions were allowed to reach equilibrium. His-Tbx6-DBD binds BS1 and BS2, but not BS3 or BS4. His-T weakly bound to BS1, bound better to BS2, and did not bind BS3 or BS4.

2.5.2 Tbx6-DBD and T-DBD have different affinities for *Dll1-msd* BS2

We sought to determine the binding affinities of Tbx6-DBD for BS1 and BS2 and T-DBD for BS2. We employed a quantitative EMSA approach whereby increasing amount of Tbx6-DBD

(range: 2.1 x 10^{-8} – 2.1 x 10^{-5}M) or T-DBD (range: 4.0 x 10^{-6} – 2.4 x 10^{-5}M) were added to a constant, limiting amount of labeled BS1-4 oligonucleotides (10pM). Since the DNA concentrations were negligible compared to the protein, the protein concentration required to bind half the DNA can be taken as an approximation of the disassociation constant, K_d (Harada et al., 1994). The percentage of DNA bound was plotted versus the concentration of each DBD, and the best fit of the data was determined to be the three-parameter Hill equation. Fitting of the data and determination of the Hill co-efficient and K_d was performed with Sigmaplot.

Strong cooperativity was observed for both Tbx6-DBD and T-DBD, as determined by a Hill co-efficient value greater than one. The Hill co-efficient of Tbx6-DBD was 2.97 and 3.18 at BS1 and BS2, respectively (Figure 19A,B). The Hill co-efficient of T-DBD was 14.42 for BS2 (Figure 19C). This cooperativity is most likely the result of the increased protein concentrations causing dimerization and non-specific contacts of a second monomer with nucleotides outside of the core 5'-AGGTGT-3' sequences. The T-DBD has previously been shown to exist as a monomer in solution, but bind to the palindromic binding sequence as a dimer (Muller and Herrmann, 1997). Although T-DBD can bind the palindromic binding sequence as a dimer, under conditions of lower protein concentrations it can also bind as a monomer (Kispert and Hermann, 1993).

Ultimately, we sought to determine whether differences in the binding affinities between His-Tbx6-DBD and T-DBD for *Dll1-msd* BS2 could explain the different levels of activation between myc-Tbx6 and myc-T at the *Dll1-msd* enhancer in our luciferase assays. The K_d of His-Tbx6-DBD for BS1 and BS2 were similar, at 1.53μM and 1.30μM, respectively (Figure 19A,B). The T-DBD had a ten-fold lower affinity for BS2, at 13.88μM (Figure 19C). The binding affinity of T-DBD for BS1 could not be measured, as our protein preparation did not allow for

high enough concentrations to be obtained in order to achieve enough data points to fit to a curve.

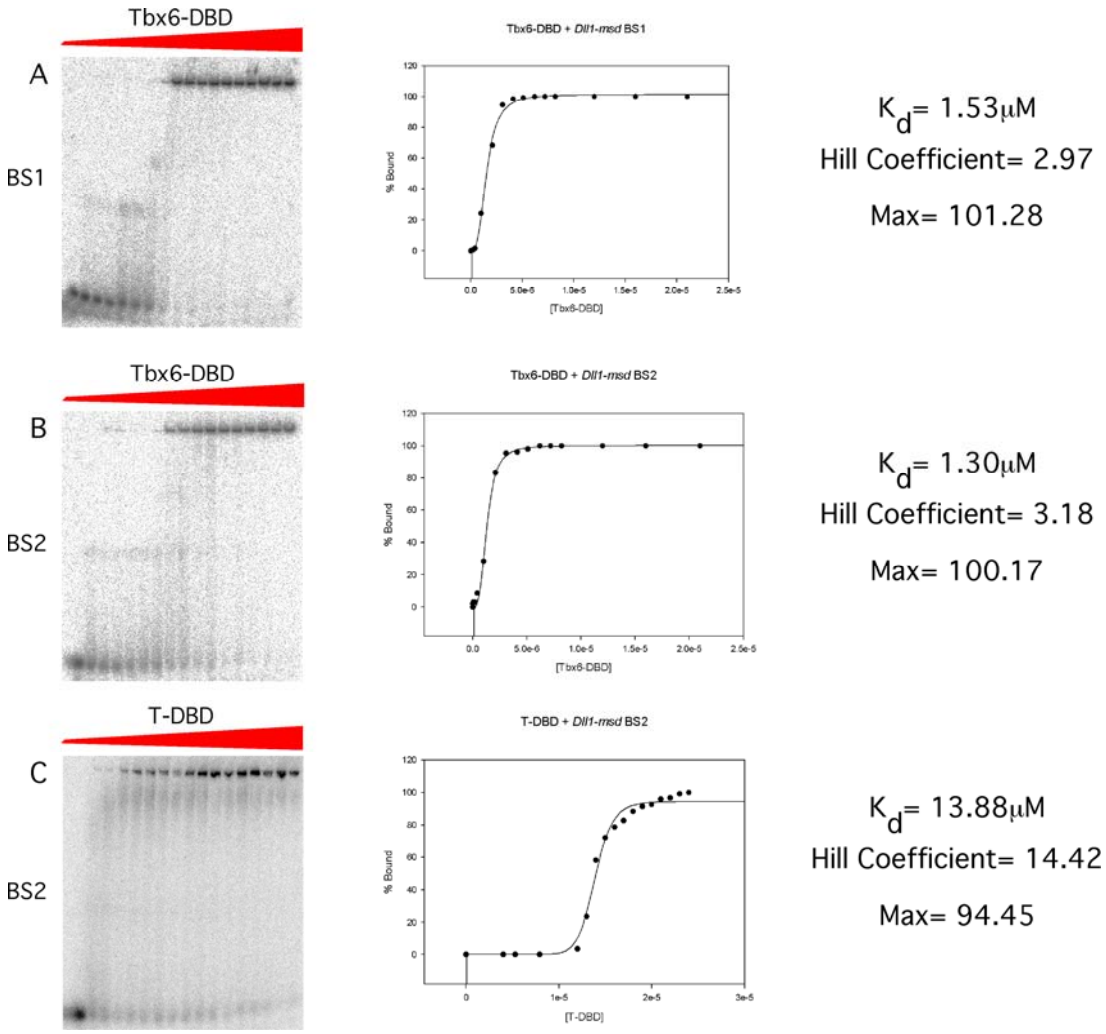

Figure 19: Tbx6-DBD and T-DBD have different affinities for *Dll1-msd* BS2.

Increasing amount of His-Tbx6-DBD (range 0.21nM– 2.1 μM) or T-DBD (range: 4.0μM – 2.4μM) was added to a constant 10pM of double-stranded labeled oligonucleotide corresponding to *Dll1-msd* BS1 or BS2. Percentage DNA bound versus concentration of protein was plotted and fitted to a three-parameter Hill equation to determine binding affinity (K_d), Hill co-efficient, and maximum percentage bound (max). (A, B) His-Tbx6-DBD has a similar binding affinity for BS1 and BS2 at 1.53μM and 1.30μM, respectively. A Hill co-efficient of ~3 indicates cooperativity of binding, and that the shifted band most likely represents a dimer. (C) T-DBD binds BS2 with a ten-fold lower affinity than Tbx6-DBD, at 13.88μM, although exhibits stronger cooperativity of binding with a Hill co-efficient of 14.42.

2.6 DISCUSSION

Establishment of the luciferase assay protocol opened up a wealth of possibilities to quickly test hypotheses about how T and Tbx6 could be interacting *in vivo* to generate the observed phenotypes when levels of T or Tbx6 are altered, or between different alleles of T and Tbx6. The combination of our luciferase assays with biochemical techniques such as EMSAs and ChIPs have generated a powerful way to investigate target gene selection and the transcriptional control of these target genes by T and Tbx6.

2.6.1 T and Tbx6 compete for a common set of downstream targets

Results from our ChIP experiments using embryonic tissues as the source of endogenously expressed T and Tbx6 confirm that *Dll1* is a downstream target of both. It is essential to perform these experiments with endogenously expressing tissue, as over-expression of transcription factors in cultured cells can lead to false positive results due to non-specific binding. Use of endogenous tissue allows us a much greater confidence in determining whether T and Tbx6 are present at the enhancers of endogenous targets. Although it is possible that either factor could be interacting indirectly with *Dll1-msd*, we believe this interaction is direct as our EMSA data indicated that both T-DBD and Tbx6-DBD could bind to *Dll1-msd* BS2. We have focused on *Dll1* as a model of competition because it is a confirmed downstream target of both T and Tbx6 and is co-expressed in the tailbud with both T-box factors. It is plausible that other common downstream targets of both transcription factors exist, although none are known to date. Future strategies to determine more common and unique targets of both T and Tbx6 are discussed in section 5.2.3.

Results of our luciferase assays indicate that although both T and Tbx6 can activate transcription from the *Dll1-msd* enhancer, they activate at ten-fold different levels. This differential level of activation was also observed with the *Mesp2P/E* reporter, suggesting that T and Tbx6 may have inherently different transcriptional activities. T was also able to compete with Tbx6 at both enhancers in competition luciferase assays to effectively decrease the transcription activity of Tbx6 when T and Tbx6 were co-expressed (Figure 12B,D). Contrary to this, both T and Tbx6 weakly activated transcription from the synthetic T^{bind} enhancer at approximately equivalent levels and act in an additive fashion in competition luciferase assays.

As the T^{bind} sequence represents an 'optimal' binding site for both T and Tbx6, this suggests that the context of the binding sites and flanking sequences may influence the ability of T or Tbx6 to bind efficiently. Preferences for different flanking sequences surrounding the core 5'-AGGTGT-3' sequence has been previously noted (Conlon et al., 2001). Despite the apparent differential preferences for nucleotides outside of the core sequence, examination of the two published crystal structures of the T-domain of T and Tbx3 in complex with the palindromic binding sequence reveals that the amino acids that make specific contacts with the bases are highly conserved across all T-box proteins, as well as a high degree of structural conservation of the T-domain (Coll et al., 2002; Muller and Herrmann, 1997). This raises the question of how exactly differential specificity for nucleotides outside of the core sequence arises. It is possible that addition of the N- and C-termini could potentially add the complexity of interacting proteins to the mix, as the published crystal structures were obtained with only the T-domain in contact with the palindromic binding sequence. Alternatively, slight differences in structure between the published crystal structures and what occurs *in vivo* may account for differential selection of targets. It is also important to note that although the palindromic binding sequence was selected

by T in PCR-based binding site selection experiments and was used to generate the crystal structures of the T-domain of T and Tbx3 (Coll et al., 2002; Muller and Herrmann, 1997), the palindromic sequence does not exist *in vivo,* and therefore represents an artificial condition.

Although examination of structural data on the T-domain reveals that it is unclear why or how T-box proteins differentially select target genes, our EMSA data suggested that T-DBD and Tbx6-DBD have differential ability to bind to T-box binding sites within the *Dll1-msd* enhancer. Tbx6-DBD bound with similar affinity to both *Dll1-msd* BS1 and BS2. In contrast, T-DBD bound weakly to BS2 weakly, if at all, to BS1. These results were not entirely surprising since PCR-based binding site selection data indicated that Tbx6 prefers any nucleotide and then a purine, respectively, in positions +1 and +2 following the core 5'-AGGTGT-3'sequence (White and Chapman, 2005). In contrast, T prefers 5'-GAAA-3' following the core sequence (Kispert and Hermann, 1993). Examination of the structural model of T and Tbx3 T-domains revealed that the majority of the contacts outside of the core sequence occur on the opposite strand to the core sequence (Coll et al., 2002; Muller and Herrmann, 1997).

For T, BS1 contains an unfavorable TG sequence in the +1 and +2 positions following the core. Based on PCR-based binding site data, this TG dinucleotide following BS1 would still allow Tbx6 to bind, which is consistent with our EMSA results. In contrast, BS2 contains a GA dinucleotide in the +1 and +2 positions, allowing for binding by both Tbx6 and T. Neither T nor Tbx6 bound BS3 despite containing the core 5'-AGGTGT-3' sequence. However, BS3 has a CC dinucleotide in the +1 and +2 positions, which is unfavorable for both T and Tbx6 based on PCR-based binding site selection data. Altogether, these data suggest that sequences outside of the core are indeed important. BS4 consists of an unfavorable 5'-AGGGTGTG-3' sequence, which does not match the core sequence because the core TGT triplet is separated by an extra G.

The majority of DNA contacts occur at the TGT triplet via a conserved arginine residue (Muller and Herrmann, 1997). It is therefore not surprising that neither T nor Tbx6 binds to BS4. Although the EMSA experiments give us an idea of which binding sites Tbx6-DBD and T-DBD prefer, they do not allow us to examine potential cooperative binding of multiple sites. Future directions to address cooperative binding of multiple sites are addressed in section 5.2.2.

Although *Dll1-msd* BS2 is an exact match for both T and Tbx6, T-DBD binds with a ten-fold lower affinity than Tbx6-DBD (K_d= 13.88μM as compared to 1.3μM). As there is currently no structural data for Tbx6-DBD, reasons for the decreased affinity of T-DBD for BS2 relative to Tbx6-DBD are unknown. Based on the high degree of similarity between the structure of the T-domain of Tbx3 and T, and the highly conserved nature of amino acids that make direct contact with the bases, we would expect that the structure of Tbx6-DBD to be very similar to the T-DBD structure.

Overall, this data indicated that there might be a common subset of targets that are shared by T and Tbx6 within the PS and tailbud. An example of this is *Dll1*, where Tbx6-DBD can bind two of the four T-box binding sites (BS1 and BS2), while T-DBD is only able to bind BS2 with a reasonable affinity and BS1 with a much lower affinity. On its own, BS1 would not appear to contribute significantly to T's activity at *Dll1-msd*, although it is possible that T may bind cooperatively to BS1 after binding BS2 *in vivo*. In addition, T-DBD binds with a ten-fold lower affinity to BS2 than does Tbx6-DBD. These differences in binding preferences between Tbx6 and T may translate into differences in activity level from the *Dll1-msd* enhancer in our luciferase assays. As a non-exclusive alternative, T and Tbx6 may share a common, unknown co-factor necessary for their transcriptional ability. In this scenario, addition of increasing amount of T relative to Tbx6 would serve to sequester the common co-factor from Tbx6 to T,

thus reducing its transcriptional ability. To date, there are no known co-factors of murine Tbx6 or T, making this a difficult hypothesis to investigate further at this time. It is certainly possible that competition for a common co-factor may in combination with competition for binding to the enhancer lead to the *in vitro* behavior and *in vivo* phenotypes that we observe. In the case of *Dll1*, we have shown that both factors can to bind to the enhancer region *in vivo*, and that competition for binding to the enhancer region itself partially, if not entirely, can explain our competition luciferase data. This signifies that competition for binding to the enhancer region of common target genes is a potential mechanism for T-box factor regulation of target gene expression.

We predict that T and Tbx6 share at least some common targets in their common expression domain, the PS/tailbud. The lower activity of T as compared to Tbx6 at a common subset of targets may serve to attenuate a Tbx6-based transcriptional program *in vivo*. We hypothesize that T is required to maintain cells in a PS-like state and continue to generate mesoderm to extend the body axis, while Tbx6 is required to push cells towards a PAM fate. Therefore, changes in intracellular concentrations of Tbx6 and T relative to each other can influence cell fate decisions and result in the observed phenotypes. This is especially apparent in our *Tg46/Tg46* embryos that over-express *Tbx6* within its endogenous domain, and display truncated axis similar to *T* heterozygotes. Here, we hypothesize that the increased levels of Tbx6 relative to T results in an increased Tbx6 transcriptional program pushing cells towards a PAM fate relative to maintenance of a PS-fate. Additionally, the bulbous tailbud phenotype observed in *Tbx6* null embryos may be the result of an dramatically increased level of T and absence of Tbx6 resulting in an excess of cells maintained in a PS-like state. This is further supported by the reduction of the bulbous tailbud phenotype in *T/+; Tbx6-/-* embryos. Overall, the observed

competition between T and Tbx6 for a common subset of genes may regulate the transition between PS and PAM fate *in vivo.*

2.6.2 T^{Wis} acts as a dominant-negative and competes with both T and Tbx6

T^{Wis} is an allele of T that produces a protein capable of binding DNA but not transactivating, and has been proposed to act as a dominant-negative by preventing related proteins from binding similar targets (Kispert et al., 1995; Shedlovsky et al., 1988). Luciferase assays performed with myc-T^{Wis} at the *Mesp2P/E* reporter verified that myc-T^{Wis} did not activate transcription. When tested in competition luciferase assays, myc-T^{Wis} effectively lowers the transcriptional activity of myc-T from the *Mesp2P/E* enhancer. Although it is expected that a dominant-negative allele of *T* such as *T^{Wis}* would be able to compete with the full-length T, the increased severity of the *T^{Wis}/T^{Wis}* phenotype (no somites formed) as compared to the *T* null phenotype (0-5 abnormal, anterior somites formed) suggested that T^{Wis} can also compete with a related transcription factor (likely Tbx6) for binding to a common subset of downstream targets. Competition luciferase assays performed with increasing amounts of myc-T^{Wis} and a constant amount of myc-Tbx6 demonstrated that myc-T^{Wis} could also lower the transcription activity of myc-Tbx6 from the *Mesp2P/E* reporter. These competition luciferase assays support the genetic interaction observed in *T^{Wis}/+; Tbx6/+* mice, which display fusions of the ribs and vertebrae that are not observed in *Tbx6/+* or *T^{Wis}/+* mice, which are similar to the Tbx6 hypomorphic phenotype. Importantly, this genetic interaction is not observed in *T/+; Tbx6/+* mice, *T^{Wis}/+; Dll1/+* mice or in *T^{Wis}/+; wnt3a/+* mice (data not shown), indicating this interaction is specific to Tbx6. Alternatively, T^{Wis} may function as a neomorph, gaining activity at genes that T does not usually affect. The binding preferences and affinities of T^{Wis} relative to T are also not known.

In conclusion, the genetic interactions observed when relative levels of T and Tbx6 are altered in T^{Wis}/+; *Tbx6*/+ mice support our hypothesis that these two T-box transcription factors compete for a common subset of downstream targets *in vivo.* This implies that the developing embryo is exquisitely sensitive to relative levels of T and Tbx6. Our hypothesis is further supported by our competition luciferase data and biochemical data, suggesting that *Dll1* is one common target of both T and Tbx6 in the PS and tailbud where they have different activity levels. In our luciferase assays with the *Dll1-msd* enhancer, we observed a ten-fold difference in transcriptional activity between Tbx6 and T. This difference in activity may serve to attenuate a high transcriptional level of *Dll1* within the tailbud, allowing for tightly regulated expression levels both spatially and temporally. Identification of other common targets of T and Tbx6 awaits ChIP followed by next-generation sequencing (ChIP-seq) experiments with our protocols for use of dissected embryonic tissue.

3.0 GENERATION OF *DLL1-MSD-CRE* TRANSGENIC MICE

3.1 INTRODUCTION

To study PAM formation in the mouse, transgenic lines that can be used to either selectively delete or express genes of interest in the paraxial mesoderm are required. To develop a reagent for studying PAM in the mouse, we cloned the *Dll1 mesoderm (msd)* enhancer upstream of a minimal promoter and Cre recombinase to generate *Dll1-msd Cre* transgenic mouse lines. The *Dll1* enhancer is a good candidate for driving ectopic expression within the PAM due to its expression within the PSM and the posterior halves of the somites (Bettenhausen *et al.*, 1995). Previous studies identified a 1.5 kb regulatory element from the *Dll1* gene that was capable of driving expression of a minimal promoter-*lacZ* reporter in the PSM beginning at e7.5 (Beckers *et al.*, 2000a). This 1.5 kb mesoderm (msd) enhancer element contains binding sites for Tbx6 and LEF/TCF (Wnt signaling pathway) transcription factors (Hofmann *et al.*, 2004; White and Chapman, 2005). Genetic evidence along with EMSA and transcriptional assays demonstrate the importance of both Tbx6 and Wnt signaling in controlling expression from the *Dll-msd* enhancer (Beckers et al., 2000b; Hofmann et al., 2004; White and Chapman, 2005; White et al., 2003).

3.1.1 Aims of these studies

Our lab is interested in the specification and patterning of the PAM. In the mouse, knockout studies can reveal the requirement of genes in specific processes with the caveat that usually only the first critical role of the gene product is revealed due to early embryonic lethality. The use of different Cre recombinase transgenic lines to create tissue-specific knockouts has thus become critical for the study of the roles of gene products in a specific tissue. In addition to the use of knockouts, Cre recombinase transgenic lines can be used to selectively activate gene expression in transgenic systems. We sought to investigate the phenotypic consequences of mis-expression or deletion of a gene of interest in the PAM in the mouse. We established *Dll1-msd Cre* transgenic lines and thoroughly characterized one line, *Dll1-msd Cre Tg33*, by describing the domain of Cre recombinase activity during mouse embryogenesis. It is notoriously difficult to achieve high transfection efficiency in primary fibroblasts and therefore expression of a gene of interest in a large number of primary fibroblasts was not previously possible. We used this *Dll1-msd Cre* line together with a conditional reverse tet-transactivator transgenic line to generate fibroblasts for inducible expression of myc-tagged full length Tbx6 in mesoderm-derived cells. We tested the efficacy of this 3-component scheme to generate fibroblasts that inducibly express myc-tagged, full-length Tbx6 in a significant proportion of fibroblasts (70-75%) for use in chromatin immunoprecipitation studies.

3.2 EXPRESSION OF *DLL1-MSD CRE IN VIVO*

To examine the spatial and temporal activity of the *Dll1-msd* Cre recombinase in this *Dll1-msd Cre Tg33* line, transgenic mice were mated to *ROSA26-lacZ* reporter mice in which the expression of *lacZ* is silenced until the floxed stop of transcription is removed by Cre recombinase activity (Soriano, 1999). Embryos from these matings were dissected and stained for β-galactosidase activity. Once the stop of transcription is removed, the *lacZ* reporter is expressed in the cells expressing the Cre recombinase and all of their descendants (Figure 20).

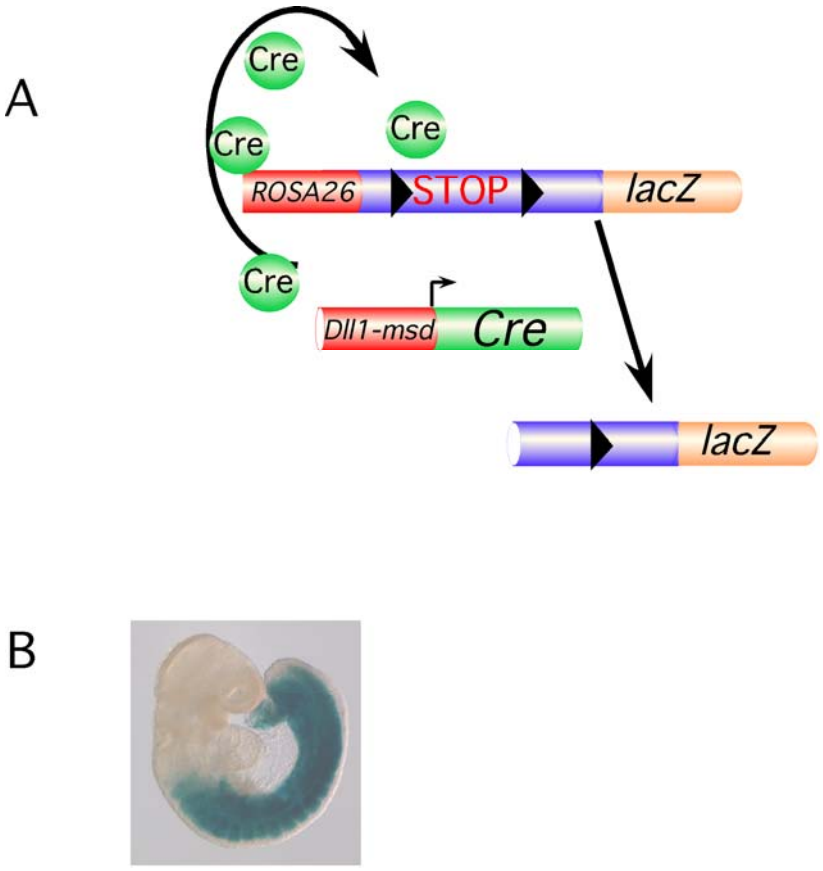

Figure 20: Schematic of *Dll1-msd Cre; ROSA26-LacZ* reporter

(A) Expression of *Cre recombinase* from the *Dll1-msd* enhancer beginning at e7.0 results in the excision of the floxed stop of transcription cassette upstream of the *lacZ* reporter inserted at the ubiquitously expressed *ROSA26* locus, allowing for expression of *lacZ.* (B) Examination of *lacZ* expression domain at e9.5 via β-galactosidase staining of *Dll1-msd Cre/+; ROSA26-rtTA/+* embryo reveals the initial domain of Cre recombinase activity and all their derivatives.

At e7.0, very few cells in the embryonic portion of the embryo stained positive for β-galactosidase activity. Between e7.5 and e8.5, β-galactosidase activity was apparent in the posterior of the embryo in the mesoderm, consistent with the appearance of PAM, although robust staining was not observed until the headfold stage. At this stage β-galactosidase activity was found in the PAM and its derivatives, but was not observed in the node or allantois (Figure

21A-C). By e8.5, both PSM and somites stained positive, while the tip of the tail, which contains the PS, did not stain (Figure 21D). By e9.5 and at later stages, staining was more widespread and was found in the lateral plate (limb buds) and intermediate mesoderm, in addition to the somites and PSM. Staining, however, was not detected in the heart, notochord, neural tissue, surface ectoderm or endoderm, as expected (Figure 21E,F). At later stages, staining was also observed in the vasculature and mesenchyme of the lungs, pancreas and other gut-derived organs, but not in ectoderm or endoderm derived tissue (Figure 21G-I). In summary, the *Dll1-msd Cre* transgene results in recombination beginning at e7.0 in the PAM and later in its derivatives, in particular the somites. At later stages, reporter gene activity was not limited to PAM-derived tissues, but was also observed in a wide-variety of mesoderm derivatives. Apoptosis induced by high levels of Cre recombinase has been reported for some Cre transgenic lines (Naiche et al., 2005a). Although we did not directly assay for apoptosis, we have not observed an obvious increase in cell death in hemizygous or in homozygous animals, which are both viable and fertile.

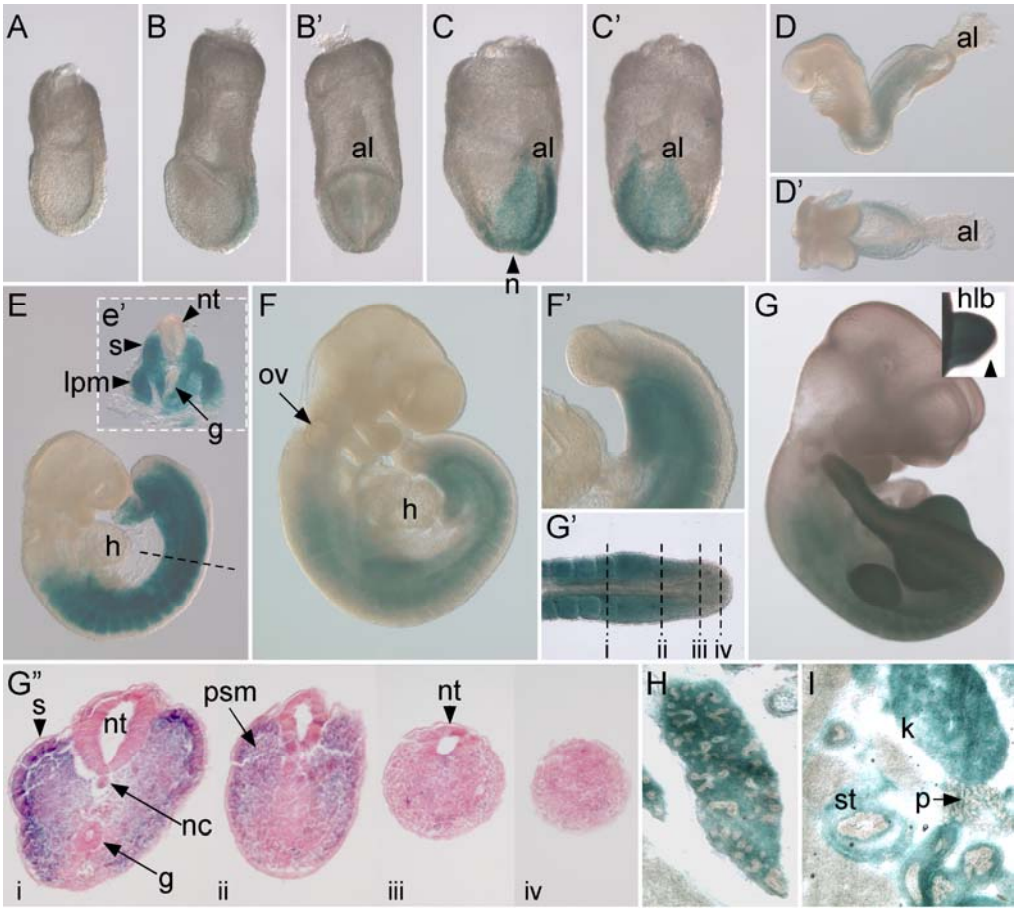

Figure 21: *Dll1-msd Cre* recombinase activity.

Activity of Cre recombinase driven by the *Dll1-msd* enhancer was assayed by staining for β-galactosidase activity in *Dll1-msd Cre/+; ROSA26-lacZ/+* embryos dissected between e7.0-14.5 and stained for β-galactosidase activity. At e7.0 (A) to early e7.5 (B, posterior view B'), few cells in the posterior of the embryo were stained. By the early headfold stage (C, posterior view C'), staining was detected in the posterior of the embryos and extended laterally, but was not observed in the node (n). At e8.5 (D, side view and D', dorsal view), staining was detected throughout the PSM and somitic mesoderm, but not in the allantois (al) or at the tip of the tail where the PS resides. (E) e9.5 embryo (~16 somites) stained overnight for β-galactosidase activity showed staining in the posterior half of the embryo but not in the heart (h) or the tail tip containing the PS. A thick section through the trunk region (dotted line in the embryo shows the approximate plane of section) shows staining in the somitic (s) and lateral plate mesoderm (lpm), but not in the gut (g), neural tube (nt) or surface ectoderm. Similarly, a later stage e9.5 embryo (F, F' high magnification of the tail) had staining through all the somites, PSM and forelimb bud (flb), but no staining

in the PS. (G) By e10.5 staining is obvious through much of the posterior half of the embryos including the fore- and hindlimb buds (hlb). The inset shows the staining in the mesenchymal layer of the hlb but not the surface ectoderm (arrowhead). (G') High magnification of an e10.5 tail showing β-galactosidase staining in the PSM and somites along with approximate planes of sections (dotted lines) for those shown in G", which were counterstained with eosin. The tip of the tail (iii and iv) is largely devoid of stained cells, whereas cells in the PSM (ii) and somite (J) are stained. (H-I) Cryosections of an e14.5 embryo stained for β-galactosidase activity revealed staining in the mesenchyme of the lung (H), pancreas and stomach (panel I), but not in the endoderm components of these organs. Staining was also observed throughout the kidney (I). Abbreviations used: al, allantois; flb, forelimb bud; g, gut; h, heart; hf, headfolds; hlb, hindlimb bud; k, kidney; lpm, lateral plate mesoderm; nt, neural tube; nc, notochord; n, node; psm, presomitic mesoderm; p, pancreas; ov, otic vesicle; s, somite; st, stomach.

3.3 GENERATION *OF DLL1-MSD CRE; ROSA26-RTTA* EMBRYONIC FIBROBLASTS

The widespread expression of the *lacZ* reporter in mesoderm derivatives prompted us to test whether we could use the *Dll1-msd Cre* transgenic line to drive mesoderm-specific expression of a conditional reverse tet-transactivator (rtTA) in embryos and in fibroblasts derived from these embryos. The reverse tet-transactivator activates transcription from tet-responsive elements (TRE) in the presence of doxycycline (DOX), a tetracycline derivative, and therefore can be used to control the expression of a gene of interest that is cloned downstream of the TRE. To do this, embryos were dissected from crosses of *Dll1-msd-Cre/+* mice with mice homozygous for *ROSA26-rtTA EGFP*. Similar to the *lacZ* reporter line, expression of the *rtTA* and *EGFP*, whose expression is linked to rtTA via an IRES, are silenced until excision of the floxed stop of transcription by Cre recombinase. Embryos were dissected at e13.5 and *Dll1-msd Cre/+;*

ROSA26-rtTA EGFP/+ embryos were used to prepare fibroblasts. In the first set of experiments, the head and liver were removed, and fibroblasts were made from the remaining portion of the embryo. In the second set of experiments, neural tissue, heart and lungs were also removed in an attempt to increase the proportion of cells expressing rtTA (Figure 22A). Fibroblasts were expanded and at the fourth passage plated onto coverslips. In both sets of fibroblast preparations, cells were stained with the nuclear stain TO-PRO 3 to obtain the total number of cells, and either examined for EGFP directly or stained with anti-GFP to determine the percentage of cells expressing EGFP. Fibroblasts derived from embryos with only the head and liver removed had an average of 72.4% EGFP positive cells (n=216 total cells counted). Fibroblasts derived from embryos in which heart, lungs, and neural tube were also removed had an average of 70.1% EGFP positive cells (n=239 total cells counted). Visualizing EGFP directly (above results) or via anti-GFP immunofluorescence gave similar results with approximately 74.7% (n=234) of the population being EGFP positive (Figure 22B). Overall, this indicates that between 70-75% of fibroblasts isolated by this method are EGFP and therefore rtTA positive thus demonstrating the feasibility of using the *Dll1-msd Cre* mice for generating mesoderm-derived embryonic fibroblasts that express the reverse tet-transactivator in a high percentage of cells.

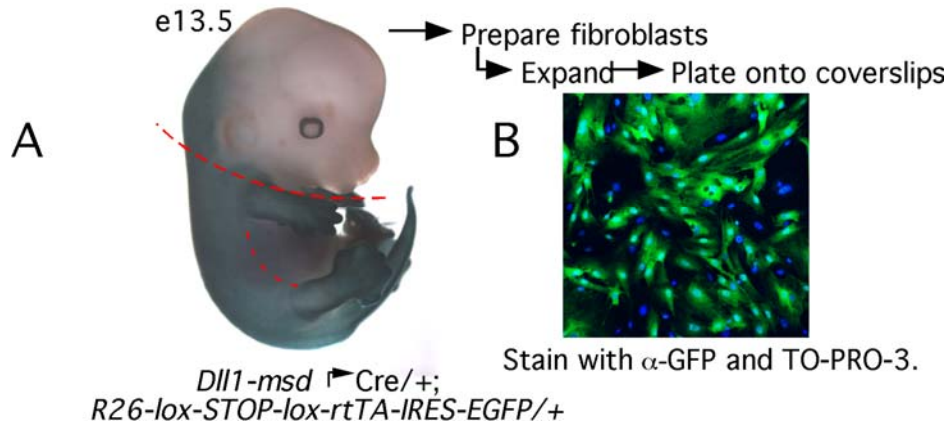

Figure 22: Derivation of *Dll1-msd Cre rtTA* fibroblasts.

Embryos isolated at e13.5 from crosses of *Dll1-msd Cre/+* and *ROSA26-rtTA EGFP (R26-rtTA)* transgenic mice were used to generate embryonic fibroblasts. (A) β-galactosidase stained *Dll1-msd Cre/+; ROSA26-lacZ* embryo at e13.5. Cells that have undergone a Cre recombination event along with their progeny are stained blue. The EGFP serves as a reporter of Cre recombination since its expression is linked to rtTA via an IRES. To derive fibroblasts, the head and liver were removed (red dotted lines) and in some cases the heart, lungs, and neural tube, to create the cell suspension. (B) Fibroblasts derived from *Dl11-msd Cre/+; ROSA26-rtTA EGFP/+* embryos were plated on coverslips and stained with anti-GFP (green) and TOPRO nuclear stain (blue). GFP positive fibroblasts represent those cells that are expressing rtTA (70-75% of the total population), while TO-PRO3 marks the nucleus of all cells.

3.4 GENERATION OF *TRE:MYC-TBX6* FIBROBLASTS

Since a high percentage of fibroblasts were positive for the recombination event we decided to test whether they were also useful to inducibly express *myc-Tbx6* in a mesodermally derived embryonic fibroblast line. In experiments described in Chapter 4, we generated another transgenic line of mice, *Tetracycline-responsive-element (TRE):Myc-Tbx6*, which express a full-

length Myc-tagged version of *Tbx6* upon addition of the tetracycline (Tet)-transactivator (TA). To initially test whether the *TRE:Myc-Tbx6* embryos are responsive to the Tet-TA and do not express Myc-Tbx6 in the absence of the Tet-TA, we generated fibroblasts from e13.5 *TRE:Myc-Tbx6* embryos. These fibroblasts were transfected with the tet-TA expression plasmid and processed for either immunofluorescence or Western blotting. Immunofluorescent staining of transfected fibroblasts revealed a low percentage (~10%) of fibroblasts that expressed a nuclear Myc-Tbx6 (Figure 23A). The low percentage of transfected cells is most likely because primary cell lines are notoriously hard to transfect with high efficiency. Nuclear localization is important because this is critical for Myc-Tbx6 to elicit a transcriptional effect in experiments described in Chapter 4. Western blotting revealed that myc-Tbx6 was produced only in the presence of the Tet-TA, and that the protein was the correct size (70 kDa) (Figure 23B).

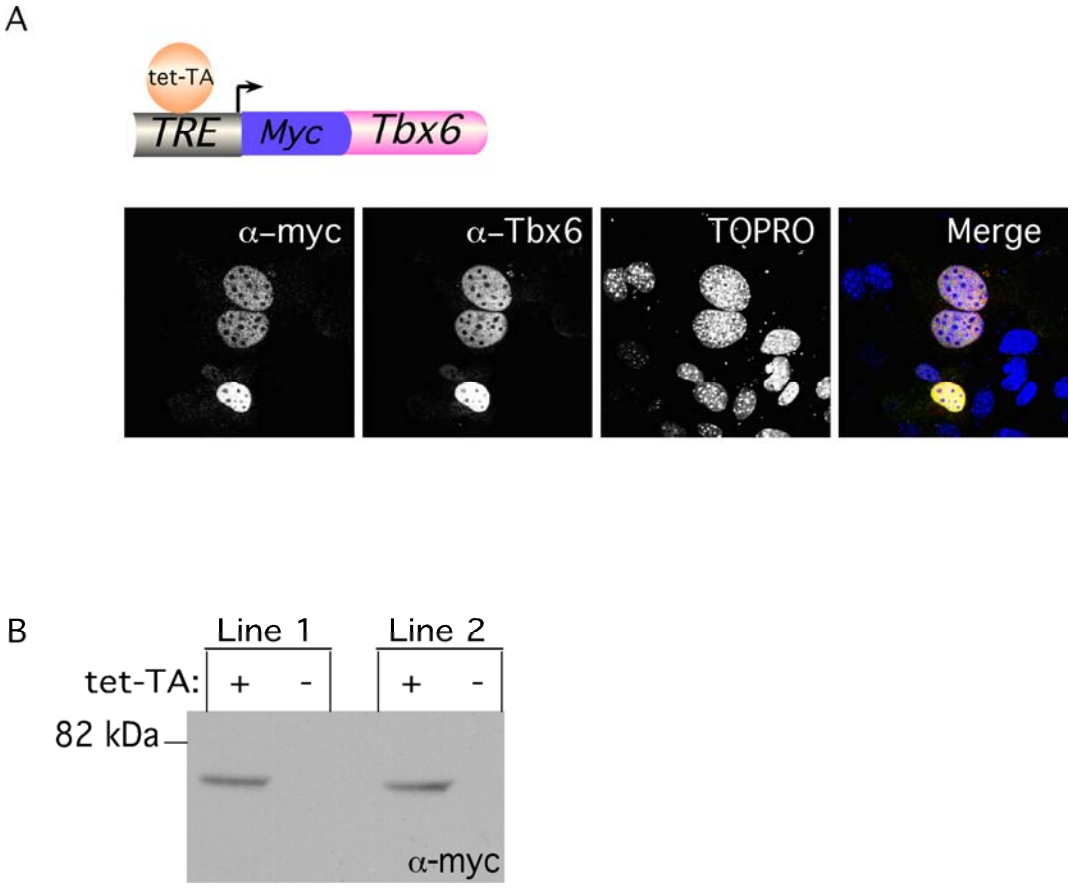

Figure 23: Generation of *TRE:Myc-Tbx6* fibroblasts

(A) Fibroblasts derived from *TRE:Myc-Tbx6/+* embryos were transfected with a tet-TA expression plasmid, and processed for immunofluorescence and stained with anti-myc (green), anti-Tbx6 (red) and TO-PRO3 (blue). The merge panel reveals nuclear localized Myc-Tbx6 staining in ~10% of cells. (B) Western blotting of two separate *TRE:Myc-Tbx6* fibroblast lines revealed that Myc-Tbx6 is expressed only when the tet-TA is transfected.

3.5 GENERATION OF *DLL1-MSD CRE; ROSA26-RTTA; TRE:MYC-TBX6* FIBROBLASTS

Lastly, we desired to create a system by which we could inducibly express a gene of interest in mesodermal fibroblasts. To generate these fibroblasts, we dissected embryos carrying all three

transgenes: *Dll1-msd Cre; ROSA26-rtTA;* and *TRE:Myc-Tbx6* (Figure 24A). In this system, rtTA is inactive until DOX is added to the culture media, therefore these cells were cultured either in the presence or absence of DOX for 36 hours, before performing immunofluorescence and Western blotting experiments. Immunofluorescence experiments revealed approximately 50% of the cells expressed a nuclear Myc-Tbx6 (Figure 24B), although approximately 70-75% of cells expressed EGFP (and thus rtTA). All cells expressing Myc-Tbx6 also expressed EGFP, demonstrating that the expression of *TRE:Myc-Tbx6* was dependent on rtTA and DOX. Western blotting revealed proper expression of Myc-Tbx6 in the presence of DOX, but not in its absence (Figure 24C).

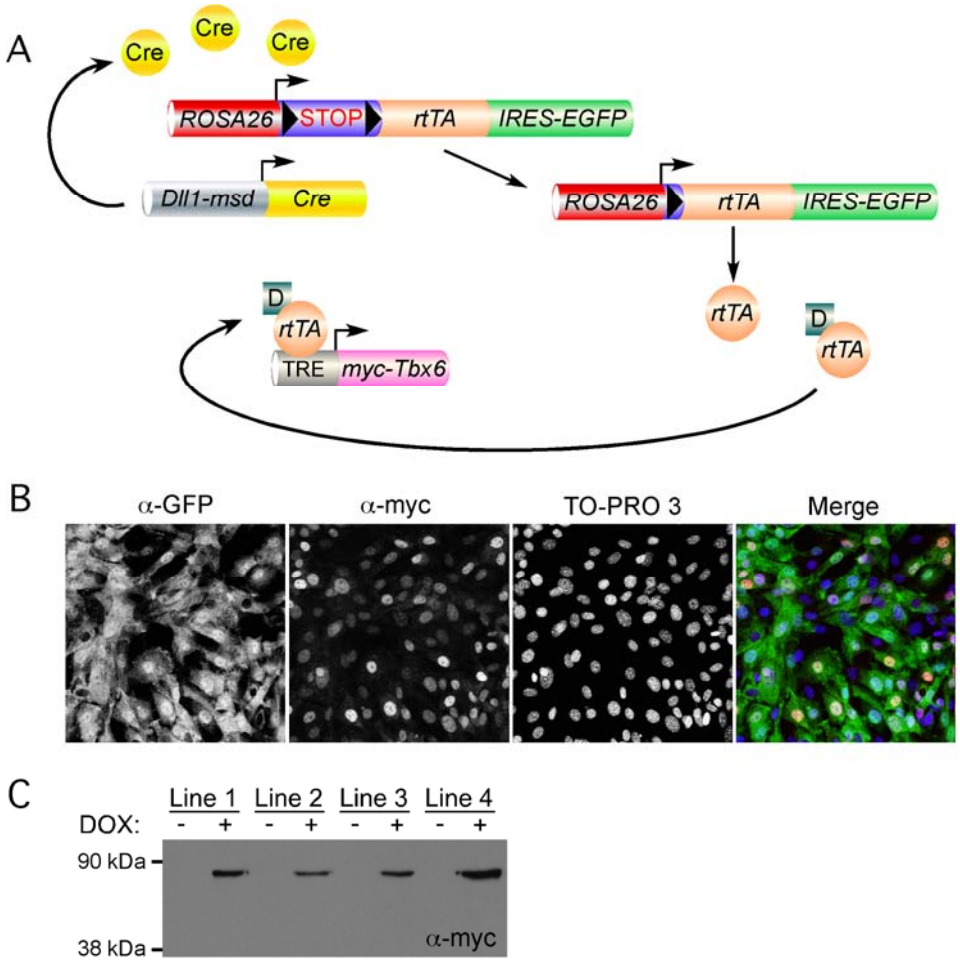

Figure 24: Generation of mesodermal fibroblasts that inducibly express *Myc-Tbx6*

(A) Schematic representing *Dll1-msd:Cre; ROSA26-rtTA-EGFP; TRE:Myc-Tbx6*-based gene expression in fibroblasts. Expression of Cre recombinase in mesoderm derivatives excises the floxed stop (STOP) of transcription cassette, allowing expression of *rtTA-IRES-EGFP*. The loxP sites are represented by black triangles. rtTA binds to the Tet-responsive element (TRE) but is inactive until DOX (D, teal box) is added to tissue culture media. Once activated, rtTA-DOX drives expression of *myc-Tbx6*. (B) Fibroblasts derived from *Dll1-msd:Cre; ROSA26-rtTA-EGFP; TRE:Myc-Tbx6* embryos plated onto coverslips and stained with anti-GFP (green), anti-myc (red) and the nuclear stain TO-PRO 3 (blue). Individual and merged channels are shown. Approximately 50% of the fibroblasts inducibly express Tbx6, while 70-75% express EGFP. (C) Fibroblast cell lines derived from four different transgenic embryos were cultured in the absence or presence of DOX, and processed for Western blotting using the anti-myc antibody to detect the fusion protein. Expression of myc-Tbx6 was observed only with addition of DOX.

3.6 DISCUSSION

As assayed by the *ROSA26-lacZ* reporter, *Dll1-msd Cre* transgenic mice possess Cre recombinase activity in the PSM and its derivatives, as well as other mesoderm derivatives. As this enhancer element has previously been shown to be expressed only in PAM (Beckers *et al.*, 2000a), expression of the reporter in other mesoderm tissues, namely lateral plate and intermediate mesoderm, suggests that a common precursor to the lateral plate, intermediate and somitic mesoderm underwent a recombination event. Cre activity was not observed in the PS, suggesting that activity in the lateral plate and intermediate mesoderm was not due to a PS recombination event. Alternatively, the specific integration site of this transgene could be altering its expression, thus allowing more widespread expression of Cre recombinase in other mesoderm derivatives. Although we initially obtained two *Dll1-msd Cre* transgenic lines that had similar reporter gene expression domains, we only thoroughly characterized one of the lines, as the second line was subsequently lost. Interestingly, reporter gene expression was not observed in all mesoderm, as the notochord (axial mesoderm) and cardiac mesoderm were not stained. Furthermore, although activity of Cre in this transgenic line is not limited to PAM, it is limited to mesoderm derivatives, as the neural tube, surface ectoderm, and gut did not stain with β-galactosidase.

Other Cre recombinase lines that are currently available to study mesoderm formation in the mouse include the *T-Cre* line (Perantoni *et al.*, 2005) and the *Meox1-Cre* line (Jukkola *et al.*, 2005). The *T-Cre* line is active beginning at e7.5 in the PS and migrating mesoderm. By e8.5, reporter gene expression was observed in all mesodermal lineages including the extraembryonic mesoderm, the allantois, in addition to the notochord, floorplate of the neural tube, part of the heart, intermediate, and lateral plate mesoderm (Perantoni *et al.*, 2005). Recombination driven

by *T-Cre* is therefore more extensive than we observed for the *Dll1-msd-Cre*. *Meox1-Cre* is a knock-in allele that shows reporter gene expression in the somitic mesoderm but this is limited to the formed somites and does not appear to be active prior to segmentation of the PAM. Later reporter gene expression was also observed in the somitic derivatives, including the dermis of the back, myotome and capillaries in the neural tube (Jukkola *et al.*, 2005). Altogether these comparisons demonstrate that although the *Dll1-msd Cre* shows extensive recombination in many mesodermal cell types, it is less extensive than the *T-Cre* and is active in the PSM, where the *Meox1-Cre* is not. Therefore the *Dll1-msd Cre* line provides a unique tool for researchers who desire to study mesoderm formation in the mouse.

Seventy to 75% of the fibroblasts derived from the posterior portions of *Dll1-msd Cre, ROSA26-rtTA EGFP* embryos were EGFP positive and were therefore descendants of cells that had experienced a Cre recombination event. In an attempt to increase the percentage of cells expressing rtTA/EGFP, we dissected away the neural tube, heart and lungs in addition to the head and liver. This extra step did not improve isolation of rtTA/EGFP expressing cells suggesting that those mesoderm-derived cells exposed to Cre recombinase and their progeny, which includes paraxial, lateral plate and intermediate mesoderm, as well as cells of the vascular cells, mesenchymal cells of the lung and other gut-derived organs represent a high percentage of cells to start, are selectively retained in these cultures, or proliferate at a higher rate than the other cell types.

The high percentage of cells positive for the recombination event in the expanded fibroblast pool support the use of *Dll1-msd Cre* and *ROSA26-rtTA EGFP* mice with transgenic mice harboring a *TRE* driving expression of a gene of interest to generate embryonic fibroblast cells from mesoderm-derived tissues. The rtTA is inactive until DOX is added to the culture

media; therefore expression of the gene of interest is inducible. These fibroblasts may be more similar to embryonic mesoderm cells than existing cell lines, for example NIH/3T3, COS-7, 293T and MDCK, which are often deemed irrelevant to embryonic mesoderm studies. In addition, it is notoriously difficult to achieve high transfection efficiency in primary fibroblasts and therefore expression of a gene of interest in a large number of primary fibroblasts is not possible. We have used this scheme to perform ChIPs (see section 2.4.1), and believe it is useful when the transcription factor of interest is normally expressed in a limited region of the embryo, for example the PSM. These cell lines may also be useful for studies to identify interacting proteins and protein modifications in a more relevant embryonic mesoderm cell type.

It is interesting to note that although 70-75% of fibroblasts generated express EGFP (thus rtTA), only 50% of the cells express Myc-Tbx6 in *Dll1-msd Cre; ROSA26-rtTA; TRE:Myc-Tbx6* fibroblasts. All cells expressing Myc-Tbx6 also express EGFP, demonstrating the need for rtTA for expression of Myc-Tbx6, although not all cells expressing EGFP express Myc-Tbx6. This may represent a delay in expression of *Myc-Tbx6* after initial expression of rtTA, an insufficient concentration of DOX in the culture media, or partial silencing of the transgene.

Altogether our results show that the *Dll1-msd Cre* transgenic line will drive Cre expression in the PAM, as well as other mesoderm derivatives and is therefore a useful reagent for studying not only PAM development but also lateral plate and intermediate mesoderm. A detailed analysis of the phenotypic consequences of mis-expression of *myc-Tbx6* in the PAM appears in Chapter 4, where we further demonstrate the feasibility of using this transgenic line to drive expression of a gene of interest in embryonic mesoderm *in vivo*.

4.0 MIS-EXPRESSION OF TBX6 RESULTS IN TBX18 AND TBX15 NULL-LIKE PHENOTYPES

4.1 INTRODUCTION

We developed a system to misexpress full-length myc-tagged Tbx6 in a variety of mesoderm derivatives during development using a 3-component system to gain both spatial and temporal control of expression within the mouse embryo. This system consists of three separate transgenic lines. The first is *Cre recombinase* under the control of the *Dll1-msd* enhancer. As described in Chapter 3, this enhancer drives expression of *Cre recombinase* within the PSM and all downstream derivatives, in addition to other mesoderm derivatives including the lateral plate and intermediate mesoderm (Wehn et al., 2009). The second transgenic line consists of reverse tetracycline-transactivator (rtTA) inserted at the *ROSA26* locus downstream of a floxed stop of transcription cassette. rtTA is only active when it associates with DOX (Belteki et al., 2005). The third transgenic line contains a full-length myc-tagged Tbx6 under the control of the tetracycline-responsive element (*TRE:myc-Tbx6*).

4.1.1 Aims of these studies

We sought to determine the phenotypic and molecular consequences of ectopic *Tbx6* expression within the segmented PAM and its derivatives, where it is normally not expressed. Using a 3-component transgenic system we drove ectopic expression of *Tbx6* in the formed somites and limb buds. Maintenance of *Tbx6* expression within the segmented PAM resulted in distinct skeletal phenotypes of the ribs, vertebral column, and appendicular skeleton resembling that of *Tbx18* and *Tbx15* null embryos. We hypothesize that these phenotypes arise due to competition between the ectopically expressed Tbx6 with endogenous Tbx18 and Tbx15 at the binding sites of target genes. We further support this hypothesis *in vitro* with luciferase transcriptional assays.

4.2 EXPRESSION OF MYC-TBX6 IN 3-COMPONENT EMBRYOS

To generate embryos containing all three transgenes, we mated *TRE:myc-Tbx6/+; rtTA/rtTA* mice to mice homozygous for the *Dll1-msd:Cre* transgene. Pregnant mice were administered DOX in their drinking water beginning at e6.5 and embryos were dissected from e10.5 to e13.5. At e10.5, whole mount *in situ* hybridization (WISH) revealed the expression domain of both endogenous and ectopic *Tbx6* mRNA. Endogenous *Tbx6* is limited to the PS and PSM, located in the tip of the tail (Figure 25A). In addition to this expression domain, 3-component embryos had mosaic expression of *Tbx6* in the formed somites and limb buds (Figure 25A'-A"). Interestingly, somites in 3-component embryos appeared morphologically normal. Whole-mount antibody staining for Tbx6 at e10.5 revealed a wide gap between the tailbud and the ectopic expression domain (Figure 25B-B"). Robust expression was not observed until somite 4 (S4),

with S1 considered the most recently formed somite. To determine relative levels of endogenous and ectopic Tbx6 protein in these embryos, we dissected tailbuds, somites 1-3, somites 4-5, and limb buds from e10.5 embryos and used these for Western blot analysis. The myc-Tbx6 protein is larger than the endogenous Tbx6 and thus endogenous and ectopic Tbx6 protein can be distinguished by Western blotting using a Tbx6 antibody (Figure 25D). Endogenous Tbx6 protein was detected in isolated tailbuds, however myc-Tbx6 was absent. Low but detectable levels of myc-Tbx6 were found in somites 1-3, which corresponds to the gap observed in the whole-mount antibody staining. Somites 4-5 expressed increased levels of myc-Tbx6, and limb buds showed robust expression.

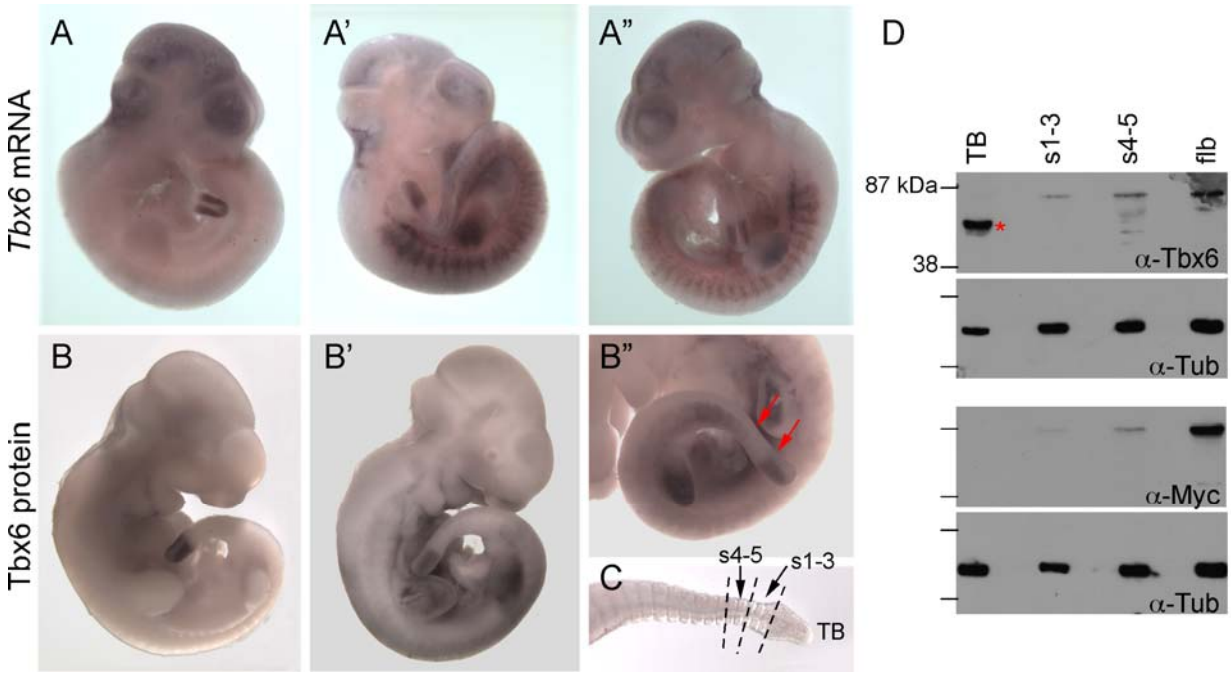

Figure 25: Mosaic expression of myc-Tbx6 in 3-component embryos.

WISH of *Tbx6* expression in a control (A) and 3-component (A'-A") e10.5 embryos. *Tbx6* expression in the control embryo is restricted to the tailbud and PSM (staining in the head is trapped color reactants). (A'-A") Two different 3-component embryos with *Tbx6* transcripts in the tailbud and mosaic *Tbx6 mRNA* expression throughout the somitic mesoderm and limb buds. Whole mount antibody staining using a Tbx6 antibody in control (B) and 3-component (B'-B") e10.5 embryos. Anti-Tbx6 staining revealed Tbx6 protein localization in the tailbud (B), while ectopic myc-Tbx6 protein mimics the mRNA expression in 3-component embryos (B'-B"). (b") A wide gap in Tbx6 endogenous and ectopic expression can be seen between the tailbud and somite 4. Somites 1-3 lie between the red arrows. (C) Tail region from an e10.5 embryo showing the planes of dissection of tissues used for Western blot analysis shown in panel D: tailbud (TB), somites 1-3 (S1-3), somites 4-6 (S4-6) and the limb bud (LB). (D) Westerns blotted with either anti-Tbx6 (α-Tbx6), anti-myc (α-myc), or α-Tubulin antibodies. Tubulin served as a loading control. Endogenous Tbx6, which is a smaller molecular weight (red asterisk), was only detected in tailbud tissue, while ectopic myc-Tbx6 was found at low levels in S1-3, higher levels in S4-6, and was robust within the limb buds. Ectopic Myc-Tbx6 was not detected in the tailbud tissue.

4.3 3-COMPONENT EMBRYOS DISPLAY AXIAL AND APPENDICULAR SKELETAL DEFECTS

To examine the consequences of ectopic myc-Tbx6 expression in the formed somite, pregnant females were treated with DOX at e6.5 and embryos were dissected at e13.5 and examined for both gross and skeletal morphology. Despite the apparent normal segmentation of the PAM, 3-component embryos exhibit striking skeletal defects at e13.5. Although the phenotypes were somewhat variable, common defects were noted within the vertebral column and appendicular skeleton. Of the thirty-four 3-component embryos examined, 100% had vertebral abnormalities including fusion and/or malformation of the atlas and axis, the most anterior somitic derivatives, and fusions of the vertebral pedicles in the anterior of the embryo (Figure 26C). The most severely affected embryos displayed fusions along the entire body axis (85%). Additionally, the distal portions of the ribs were often missing or underdeveloped (91%). Beyond e13.5, 3-component embryos become edemic, and died shortly thereafter *in utero* which precluded analysis at later developmental stages. To increase the level and presumably the number of cells that express myc-Tbx6, we generated embryos that were hemizygous for the *Dll1-msd:Cre* and *rtTA* transgenes, but homozyogous for the *TRE:myc-Tbx6* transgene. These homozygous embryos had more severe phenotypes than the hemizygotes; specifically pedicle morphology was affected along the entire axis and vertebral bodies were also malformed (Figure 26D-D').

Gross morphological examination of 3-component embryos revealed underdeveloped limbs when compared to control littermates (Figure 26A). By e13.5 control littermates display distinguishable digits and cartilage condensation within the limb bud, whereas 3-component embryos still had paddle-shaped limbs without distinguishable digit condensations (94%). Skeletal analysis of these limbs revealed shortened and malformed humerus, radius and ulna

(100%), as well as femur, tibia and fibula (97%), and occasionally the distal limb elements were missing altogether (Figure 26F). 3-component embryos also had hypoplastic scapulae (91%), which contained an ectopic foramen within the center of the blade in 21% of the embryos.

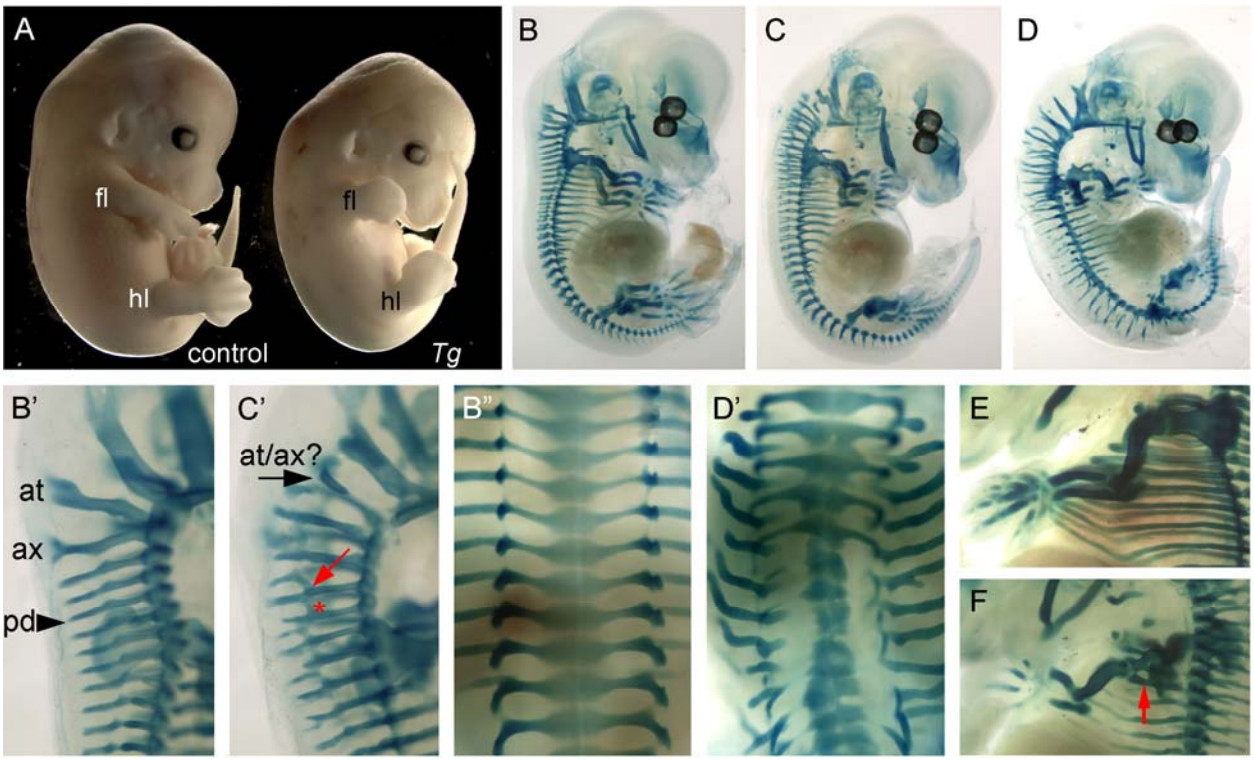

Figure 26: Three-component embryos display limb and vertebral anomalies.

Gross morphology and skeletal preparations of control and 3-component (*Tg*) embryos dissected at e13.5. (A) 3-component embryos were smaller than normal littermates, and had paddle-shaped limb buds, whereas control littermates had begun to form distinct digits (forelimb, fl; hindlimb, hl). Alcian blue-stained skeletons of control (B, C) and 3-component embryos (C, D, F). Panel C shows an embryo hemizygous for *TRE:myc-Tbx6*, while the embryo in panel D is homozygous for the *TRE:myc-Tbx6* transgene. Primed panels show a higher magnification of the embryos in panels B-D. The atlas (at) and axis (ax) are clearly distinct structures in the control embryo (B, B'). (C and D) 3-component embryos displayed fusions and malformations of the atlas and axis (at/ax, C') and expansion and fusions of the vertebral pedicles (red arrow and asterisk, C'). Vertebral body morphology was also affected in 3-component embryos, with more severe phenotypes observed in the embryo homozygous for the *TRE:myc-Tbx6* transgene (D'). Skeletal preparations of the forelimbs revealed shortened and malformed forelimb bones in the 3-component (F) embryos compared to the littermate control (E). An ectopic foramen in the scapula (arrow) was evident in 21% of 3-component embryos (F).

4.4 OVEREXPRESSION OF TBX6 PHENOCOPIES *TBX18* AND *TBX15* NULL EMBRYOS

Ectopic expression of myc-Tbx6 within the limbs and somitic tissue results in severe skeletal phenotypes. One explanation for these phenotypes is that misexpression of Tbx6 drives expression of its own downstream targets and it is the ectopic expression of these targets that causes the observed phenotypes. We therefore examined the expression of the four confirmed downstream targets of Tbx6: *Dll1, Mesp2, Msgn1,* and *Ripply2* (Hitachi et al., 2008; White and Chapman, 2005; Wittler et al., 2007; Yasuhiko et al., 2006). Expression of these targets at e10.5 was found only in their endogenous expression domain (Figure 27 and data not shown), suggesting that phenotypes were not due to misexpression of Tbx6 target genes. It is important to note that even in the limb buds where the level of myc-Tbx6 was highest, there was no ectopic expression of Tbx6 target genes.

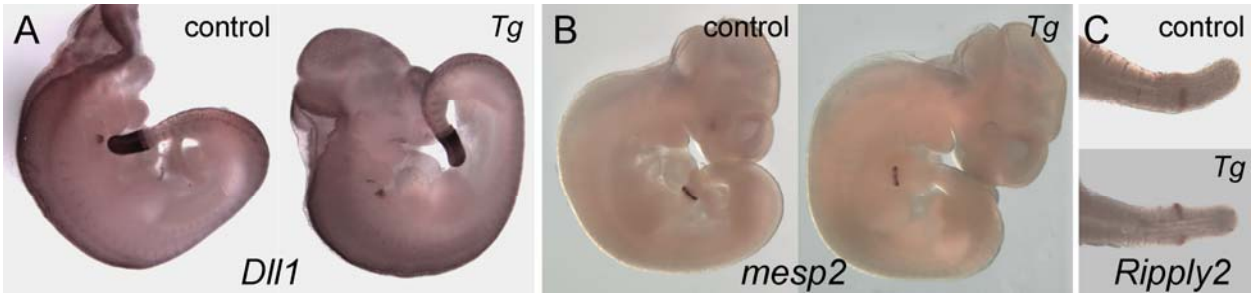

Figure 27: Tbx6 downstream target gene expression in control and 3-component embryos is indistinguishable.

The expression of Tbx6 target genes in 3-component embryos (*Tg*) is indistinguishable from control littermates at e10.5. (A) *Dll1* is expressed in the tailbud, PSM, caudal portion of each somite, and in neural tissue. *mesp2* (B) and *Ripply2* (C) are both expressed at the anterior somitic boundary. There is no ectopic expression of *Dll1* or *mesp2* in the limb buds where the highest levels of ectopic myc-Tbx6 are found.

To further investigate the cause of the observed phenotypes, we examined other T-box transcription factors that are expressed within the somites and limb buds, hypothesizing that ectopic Tbx6 could either directly alter the expression of these T-box factors or compete for binding at their downstream targets, as has been observed for other T-box transcription factors (Goering et al., 2003; Habets et al., 2002). The phenotypes observed in 3-component embryos closely resemble *Tbx15* and *Tbx18* null embryos, as well as embryos in which Tbx5 function is removed after initiation of the limb bud. *Tbx15* null embryos show abnormalities within the atlas and axis (Singh et al., 2005). *Tbx18* null embryos also display phenotypes that overlap with our 3-component embryos, specifically caudalization of the somites, which leads to expanded and fused pedicles of the vertebrae (Bussen et al., 2004). Late removal of Tbx5 results in a scapular foramen, and shortening and malformations of the forelimbs (Hasson et al., 2007).

Tbx5 and Tbx4 initiate and maintain outgrowth of the fore- and hindlimb, respectively, via a feedback loop where direct activation of *Fgf10* expression within the lateral plate

mesenchyme maintains *Fgf8* expression in the overlying apical ectodermal ridge (AER) (Agarwal et al., 2003). Loss of *Tbx5* results in loss of *Fgf10* and consequently loss of *Fgf8* expression in the AER thus halting limb outgrowth (Naiche and Papaioannou, 2003; Rallis et al., 2003; Takeuchi et al., 2003). We examined *Fgf8* expression at e10.5 in 3-component embryos to indirectly assess Tbx5 function. *Fgf8* was expressed in the AER of both control and 3-component embryos (Figure 28A), suggesting that misexpression of Tbx6 does not inhibit this function of Tbx5. While *Tbx18* null embryos do not display limb defects, *Tbx15* null embryos display a delay in endochondral bone formation in the limbs with alterations in bone shape and size. *Tbx15* null embryos also exhibit hypoplastic scapulae with ectopic foramen (Singh et al., 2005), similar to phenotypes observed in the 3-component embryos. Altogether this suggests that Tbx15 may be a primary T-box transcription factor affected by ectopic expression of Tbx6 in the limbs.

Misexpression of myc-Tbx6 in the PAM and limb buds phenocopies aspects of the *Tbx18* and *Tbx15* null embryos. It is conceivable that in these embryos ectopic Tbx6 results in the down-regulation of *Tbx15* and *Tbx18* expression and this down-regulation results in the observed phenotypes. To test this, we examined *Tbx15* and *Tbx18* expression in 3-component embryos at e10.5, focusing on changes in limb and somitic expression domains that might be altered due to ectopic Tbx6 expression. At this stage, *Tbx15* is normally expressed within the limb buds, while *Tbx18* is expressed in the rostral portion of each somite, in addition to the limb buds (Agulnik et al., 1998; Kraus et al., 2001; Singh et al., 2005). Expression of *Tbx15* and *Tbx18* appeared unaffected in our 3-component embryos (Figure 28B-C), suggesting that ectopic Tbx6 is not causing these phenotypes by down-regulating *Tbx15* or *Tbx18* expression.

The vertebral defects in *Tbx18* null embryos are thought to arise from a failure to maintain rostral somite identity, with the hypothesis that the more caudal cells eventually invade the rostral region (Bussen et al., 2004). Expression of *uncx4.1*, which is normally restricted to the caudal portion of each somite, expands slightly into the rostral portion of the *Tbx18* null somites. This expansion occurs only in the more anterior and thus more mature somites, suggesting that Tbx18 is not required for the initial repression of *uncx4.1* expression, but is instead required to maintain rostral-caudal patterning. In *Tbx6* 3-component embryos *uncx4.1* expression was still confined to the caudal portion of each somite, suggesting that rostral-caudal somite identity is initially established in these embryos. Interestingly, *uncx4.1* staining appeared darker suggesting that it was expressed higher in the 3-component embryos compared to control littermates (Figure 28D). Somitic defects in the *Tbx18* nulls are limited to the sclerotome compartment as expression of *myogenin* within the myotome was unaltered (Bussen et al., 2004). Similarly, *myogenin* expression in the 3-component embryos was unaltered (Figure 28E).

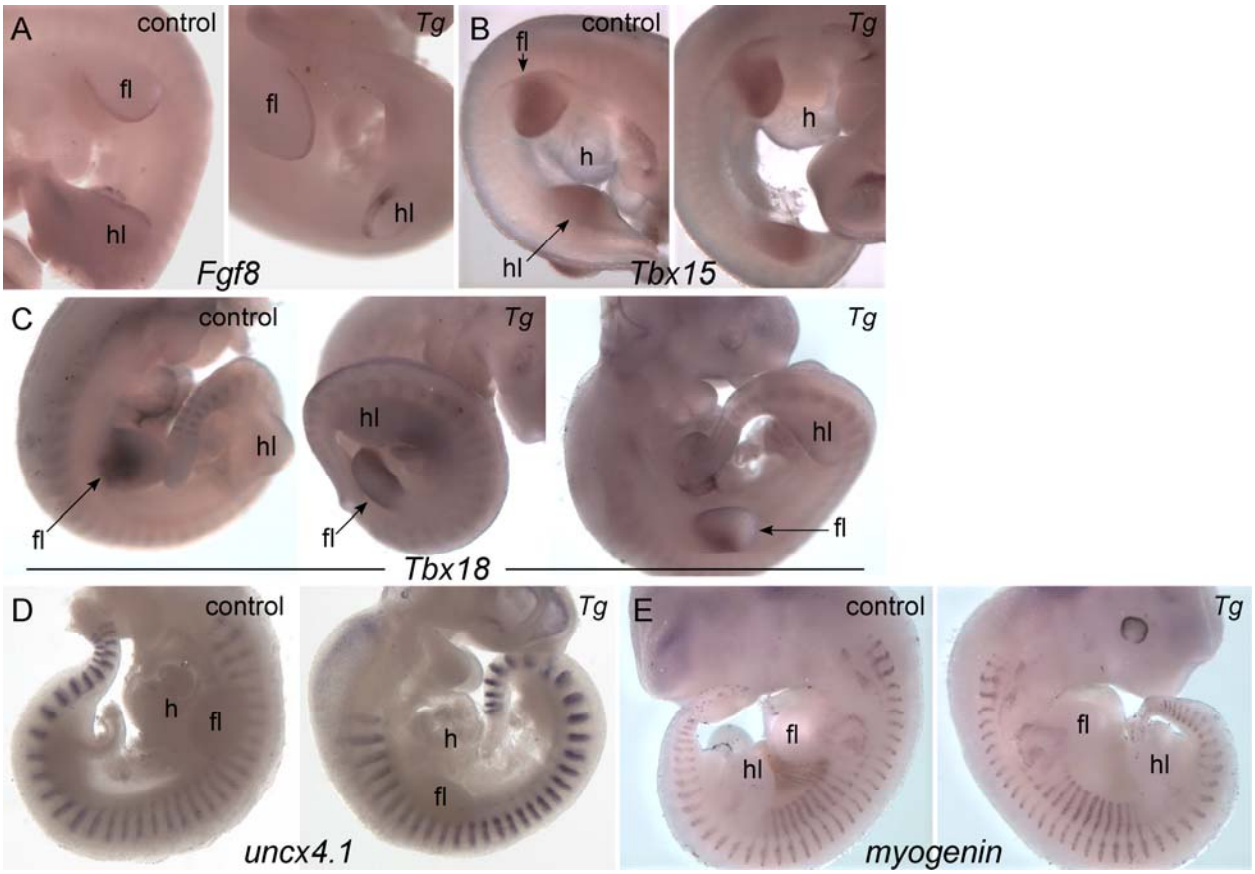

Figure 28: Spatial expression of limb and somitic markers is maintained in 3-component embryos

(A) *Fgf8* expression in the AER of control and 3-component (*Tg*) embryonic limb buds are indistinguishable. (B) *Tbx15* is expressed within the fore- and hindlimb buds of control embryos. (C) *Tbx18* is expressed within the limb buds and the anterior of each somite. There are no observable differences in *Tbx15* or *Tbx18* expression levels or pattern between control and 3-component embryos in the limb buds or somites. (E) *uncx4.1* is expressed within the posterior portion of each somite in both control and 3-component embryos, however expression levels are higher in 3-component embryos. (F) *myogenin* expression in the myotome compartment of the somites was similar in both control and in 3-component embryos. Abbreviations used: fl, forelimb bud; hl, hindlimb bud; h, heart.

4.5 MYC-TBX6 COMPETES WITH TBX18 AT THE *ANF* AND *DLL1* ENHANCERS IN LUCIFERASE ASSAYS

Both *Tbx15* and *Tbx18* encode transcriptional repressors. We hypothesize that ectopic Tbx6 in 3-component embryos interferes with Tbx15 and Tbx18 function at the level of binding to the enhancer of downstream targets. To test this, we first cloned full-length Tbx15 and Tbx18 into the pCS3mT expression vector to produce myc-tagged fusion proteins. Constructs were transfected into COS-7 cells and protein size and expression levels were verified by Western blotting. While the myc-Tbx15 and myc-Tbx18 proteins were of the predicted molecular weights (Figure 29A), myc-Tbx18 was produced at approximately two-fold greater than myc-Tbx6, and myc-Tbx15 was produced at 1.7-fold less than myc-Tbx6 when equivalent amounts of expression plasmids were transfected.

Dll1 is a *bona fide* target of Tbx6 and is also thought to be a target of Tbx18 in the somites. In the segmented PAM, *Dll1* expression is confined to the posterior of each somite via repression by Tbx18 in the anterior compartment (Bussen et al., 2004). We performed transcriptional assays using *Dll1-msd*-luc (see section 2.2.3). Our luciferase assays revealed that on its own myc-Tbx6 activates transcription from the *Dll1-msd* enhancer (Figure 29B). The RLUs for myc-Tbx18 and myc-Tbx15 alone were less than background, suggesting that in our assay system they functioned as repressors (Figure 29B-C). To test whether Tbx15 and Tbx18 could compete with Tbx6, increasing amounts of myc-Tbx18 or myc-Tbx15 expression constructs were added to a constant amount of myc-Tbx6. This resulted in a decrease in luciferase activity (Figure 29B-C).

We further confirmed these results using a second known target of Tbx18, *atrial natriuretic factor* (*ANF),* which is expressed during cardiac morphogenesis. The *ANF* enhancer

contains two consensus T-box binding sites in inverted orientation relative to each other (Brown et al., 2005). Although *ANF* is not an endogenous target of Tbx6, myc-Tbx6 weakly activated transcription from this enhancer (Figure 29E). Neither Tbx18 nor Tbx15 had significant activity at the *ANF* on their own (Figure 29E and Figure 30A). Addition of increasing amounts of myc-Tbx18 lowered the transcriptional activity of myc-Tbx6 from this promoter as well (Figure 29E). Curiously, adding increasing amounts of myc-Tbx15 did not cause a linear decrease in RLUs, and only repressed myc-Tbx6 mediated transcription a maximum of 2-fold (Figure 30A). Altogether, these results suggest that a competition between T-box factors occurs for binding sites found in enhancers of target genes, however the robustness of competition is likely to be T-box factor dependent.

To further examine whether this competition occurred at the level of DNA binding, we used the DNA binding domains of Tbx15 and Tbx18 fused to an N-terminal myc tag and nuclear localization sequence (NLS) in our luciferase assays. Interestingly, the DBD alone of Tbx15 and Tbx18 were sufficient to compete with Tbx6, further supporting that the apparent competition between Tbx6 and Tbx15 or Tbx18 occurs at the level of DNA binding (Figure 29F and 30B; Tbx15DBD at *Dll1-msd*).

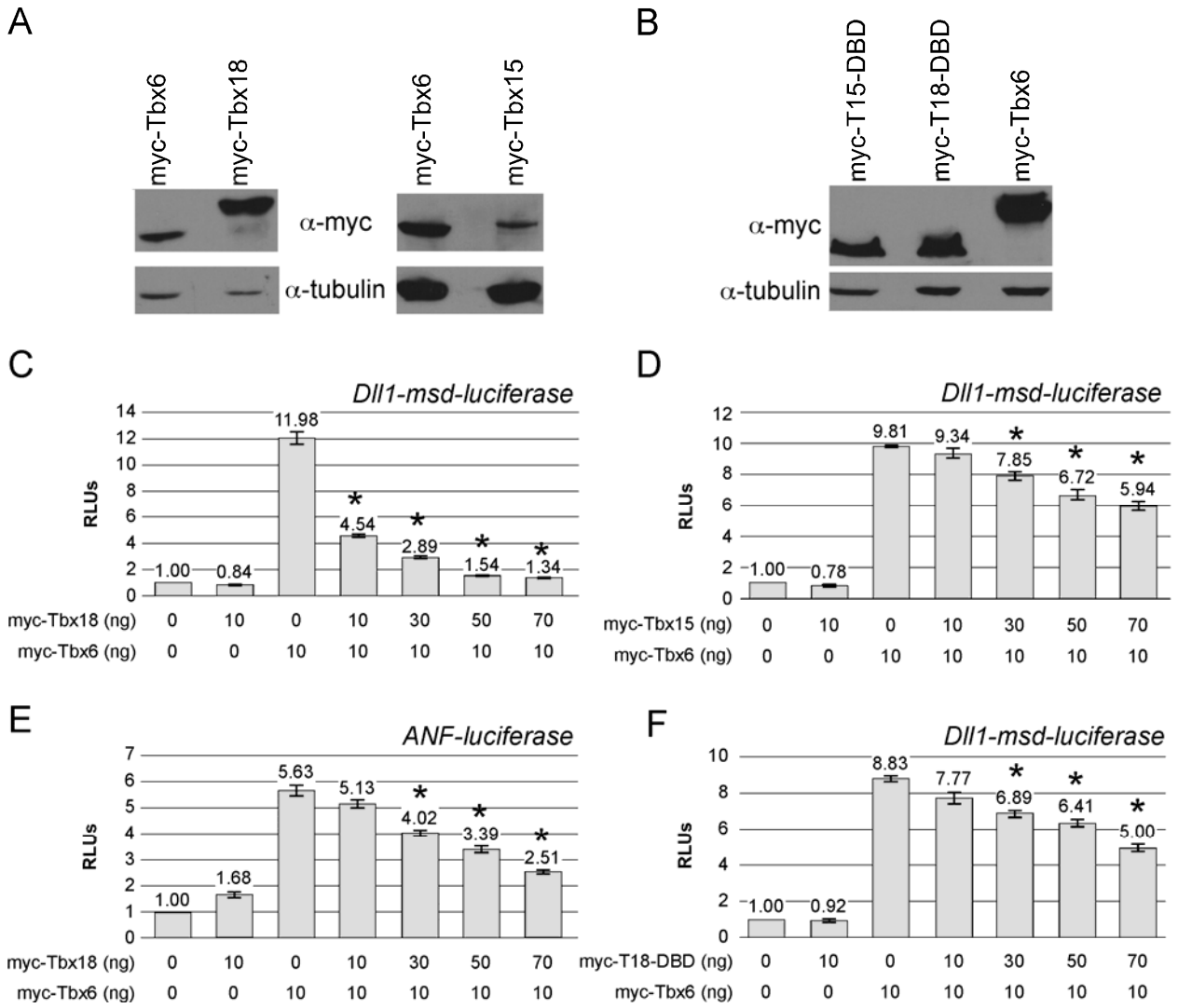

Figure 29: Luciferase assays reveal that Tbx15 and Tbx18 compete with Tbx6 at multiple enhancers.

(A) Lysates from 293T cells transfected with equal amounts of myc-Tbx6, -Tbx15 or -Tbx18 expression vectors were Western blotted with an anti-myc antibody. Twice as much myc-Tbx18 was produced as compared to myc-Tbx6, while 1.7-fold less myc-Tbx15 was produced as compared to myc-Tbx6. (B) Western blotting of lysates from 293T cells transfected with equal amounts of myc-Tbx6, myc-Tbx15-DNA binding domain (myc-T15-DBD) or myc-Tbx18-DNA binding domain (myc-T18-DBD). myc-T15-DBD and myc-T18-DBD were produced at approximately the same level, while myc-Tbx6 was produced at approximately 37% more than either myc-T15-DBD or myc-T18-DBD. (C-D) myc-Tbx18 and myc-Tbx15 slightly repressed transcription from the *Dll1-msd* enhancer, while myc-Tbx6 activated transcription. Addition of increasing amounts of myc-Tbx18 or myc-Tbx15 to a constant amount of myc-Tbx6 lowered

overall RLUs. (E) myc-Tbx6 activated low levels of transcription from the enhancer of the Tbx18 target *ANF*, while myc-Tbx18 did not significantly activate transcription. Addition of increasing amounts of myc-Tbx18 to a constant amount of myc-Tbx6 lowered the overall RLUs from the *ANF* enhancer. (f) The myc-Tbx18-DBD had no significant effect on luciferase expression from the Dll1-msd enhancer. Addition of increasing amounts of myc-Tbx18-DBD to a constant amount of myc-Tbx6 led to a linear decrease in luciferase activity. Asteriks indicate $p<0.05$.

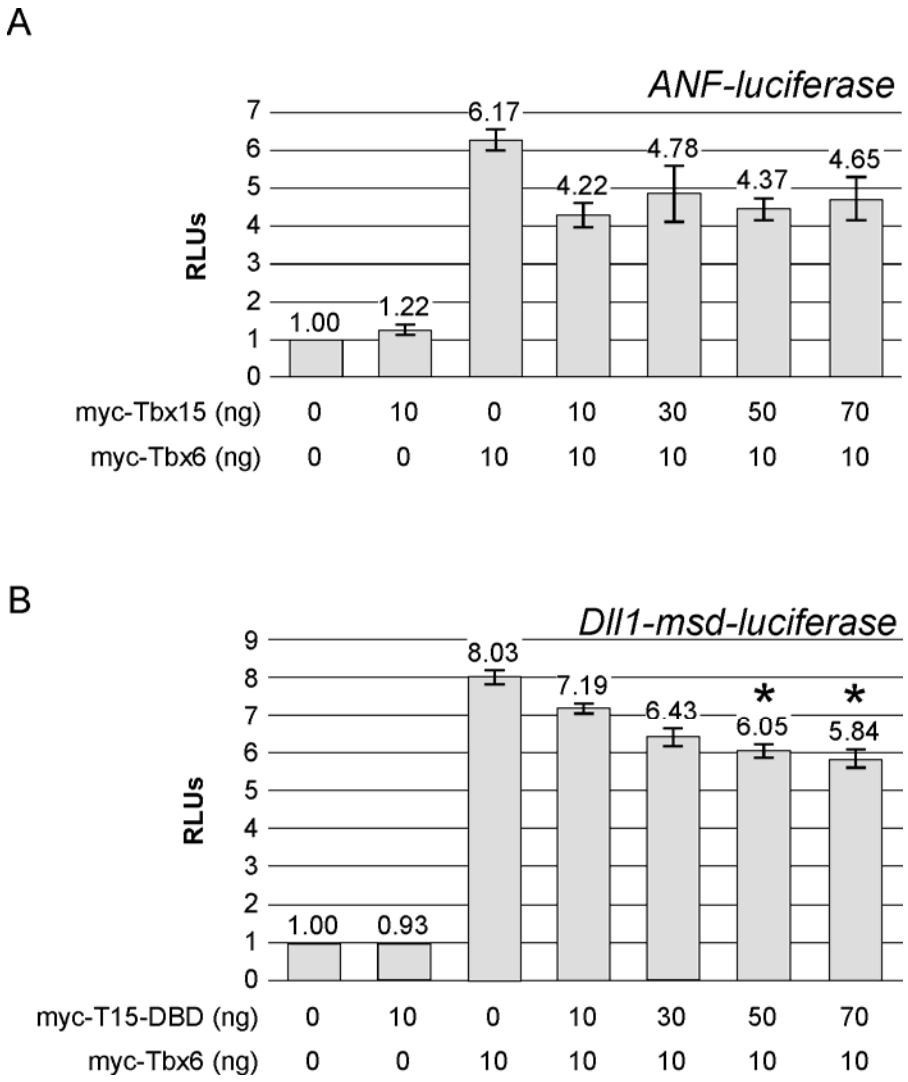

Figure 30: myc-Tbx15 and myc-Tbx15-DBD weakly repress myc-Tbx6 mediated transcription

(A) myc-Tbx6 activated low levels of transcription from the *ANF* enhancer, while myc-Tbx15 had no significant effect on transcription from the *ANF* enhancer. Addition of increasing amounts myc-Tbx15 to a constant amount of myc-Tbx6 had little effect on the overall RLUs, and only maximally lowered the RLU 2-fold. (B) The myc-Tbx15-DBD had no significant effect on luciferase expression from the *Dll1-msd* enhancer. Addition of increasing amounts of myc-Tbx15-DBD to a constant amount of myc-Tbx6 led to a linear decrease in luciferase activity. Asteriks indicate $p<0.05$.

4.6 DISCUSSION

We investigated the phenotypic consequences of ectopically expressing a full-length, myc-tagged Tbx6 in the segmented PAM and LPM. This resulted in axial skeletal phenotypes including fusion and/or malformations of the atlas and axis, fusions of the vertebral pedicles, and underdeveloped distal portions of the ribs. Limb phenotypes included delayed digit formation, shortened and malformed fore- and hindlimb bones. Underdevelopment of the ribs and digits may reflect a developmental delay of 3-component embryos, as they were noticeably smaller than the control littermates at e13.5. Additionally, these embryos also displayed hypoplastic scapulae, which contained an ectopic foramen within the center of the blade 21% of the time. The scapula is derived from both lateral plate and somitic tissue (Huang et al., 2000b), both of which expressed myc-Tbx6 in our 3-component embryos. The observed phenotypes were highly variable most likely due to the high degree of mosaic myc-Tbx6 expression observed by both WISH and whole mount antibody staining. By e13.5, embryos became edemic, precluding analysis at later developmental stages. The cause of the edema and embryonic death has yet to be investigated, although it is interesting to note that the *Dll1-msd:Cre* drives expression within the vasculature (Wehn et al., 2009), which may contribute to the edema.

To understand the mechanisms underlying the observed phenotypes, we first examined the expression of the Tbx6 target genes *Dll1, Mesp2, Ripply2* and *Msgn1*. As none of these were misexpressed in 3-component embryos, we concluded that the phenotypes did not result from up-regulation of Tbx6 targets. It is not surprising that Tbx6 target genes were not ectopically expressed in these embryos since their expression is not solely dependent on Tbx6. In addition to Tbx6, *Mesp2* expression is also dependent on Notch and FGF signaling (Oginuma et al., 2008; Yasuhiko et al., 2006). Expression of both *Dll1* and *Msgn1* requires cooperation between Tbx6

and LEF/TCF transcription factors, and hence Wnt signaling (Hofmann et al., 2004; Wittler et al., 2007). Regulation of *Ripply2* expression requires both Tbx6 and Mesp2 (Dunty et al., 2008; Hitachi et al., 2008). This complex regulatory network ensures that target genes are expressed in the correct time and place during development, effectively prohibiting ectopic expression in the absence of the other necessary factors.

Since ectopic expression of Tbx6 did not induce ectopic expression of its own downstream targets, the question of how these phenotypes arose remained. Four T-box transcription factors (*Tbx2, Tbx15, Tbx18, Tbx22,* all transcriptional repressors) are expressed in the somites. Tbx2 function is required for formation of the intersomitic blood vessels (Harrelson and Papaioannou, 2006), and therefore does not contribute to the patterning of the vertebral column or ribs. Similarly, *Tbx22* null embryos do not exhibit vertebral or rib defects, but rather display a cleft-palate phenotype (Bush et al., 2002). Strikingly, the observed phenotypes of 3-component embryos resembled those of *Tbx15* and *Tbx18* null embryos. We showed that *Tbx15* and *Tbx18* expression domains were not altered in our 3-component embryos, indicating that ectopic myc-Tbx6 did not down-regulate their expression to generate the observed phenotypes. Tbx15 and Tbx18 are both transcriptional repressors; however there are no known targets of Tbx15, and *ANF* and *Dll1* are the only known targets of Tbx18. Although Tbx18 is thought to directly repress *Dll1* expression in the rostral region of each somite, thereby limiting its expression to the caudal portion, the loss of Tbx18 does not result in the expansion of *Dll1* expression (Bussen et al., 2004). Instead, ectopic expression of Tbx18 in the somites leads to the eventual reduction in *Dll1* expression in the caudal somitic compartment. We did not observe a detectable expansion of *Dll1* expression in our Tbx6 3-component embryos.

Bussen and colleagues (2004) showed that Tbx18 is not responsible for the initial R-C patterning of the somites; instead, it appears to be required for maintenance of R-C patterning. To this end, loss of Tbx18 results in the eventual expansion of the *uncx4.1* expression domain (a caudal somite marker) into the rostral region, and the subsequent fusion of pedicle regions of the vertebrae, which is indicative of a loss of rostral sclerotome identity. This expansion is subtle and the authors suggest that it is due to caudal somitic cells expressing *uncx4.1* migrating into the rostral region. In our experiments, ectopic myc-Tbx6 also results in the expansion and fusion of pedicles, but we did not observe a detectable expansion of *uncx4.1* transcripts into the rostral region of the somite. Instead, we observed an up-regulation of *uncx4.1* expression in the caudal halves of the somites. It is conceivable that this may lead to the eventual expansion and fusion of the pedicles. The mosaic expression of myc-Tbx6 in the 3-component embryos may contribute to differences in phenotypes or intensity of phenotypes when compared to the Tbx18 null embryo. Although marker gene expression patterns were not strikingly different in our 3-component embryos compared to controls, the phenotypic consequences of misexpression were apparent. Due to the similarity of phenotypes observed in the *Tbx15* and *Tbx18* null embryos with that of our 3-component embryos, we proposed that a competition between the ectopically expressed myc-Tbx6 and endogenous Tbx15 and Tbx18 exists. In addition, alignment of the DBD of Tbx6, Tbx15, and Tbx18 reveals the high level of conservation between the three T-domains (more than 54% of residues are completely conserved between all three factors), indicating the feasibility of competition between Tbx6 and Tbx15 or Tbx18 (Figure 31).

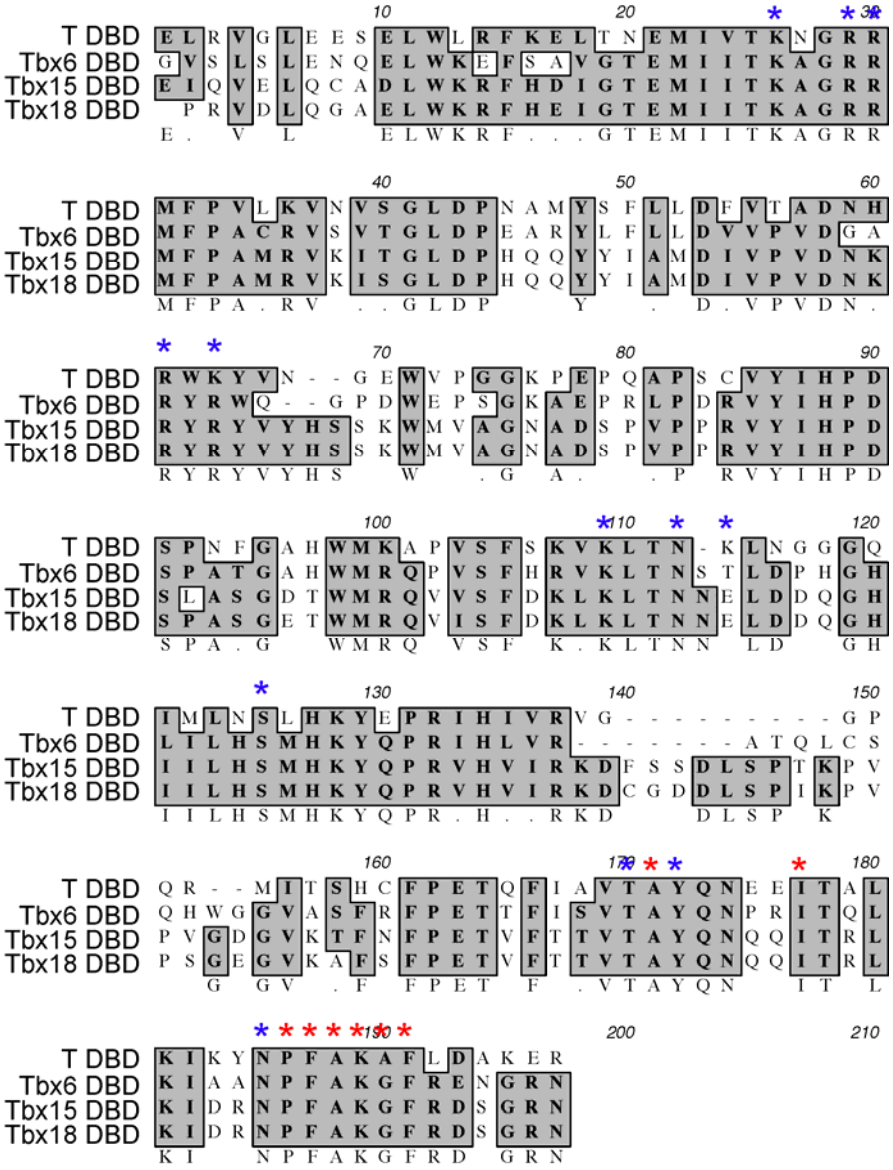

Figure 31: Alignment of the DNA binding domain of Tbx6, Tbx18, and Tbx15

Aligning the DNA binding domain of T, Tbx6, Tbx15 and Tbx18 reveals a high level of conservation. Greater than 54% of residues are completely conserved between all four factors. Residues marked with a red asterik make direct contact with the DNA base, while those marked with a blue asterik make contact with the phosphate backbone, based on the published crystal structure of the T-DBD (Muller and Herrmann, 1997).

Tbx15 and *Tbx18* are co-expressed within the proximal portion of the limb bud mesenchyme (Bussen et al., 2004; Singh et al., 2005). Several T-box factors are also expressed in the limb buds, including *Tbx2*, *Tbx3*, *Tbx4* (hindlimb only), *Tbx5* (forelimb only), *Tbx15*, and *Tbx18*. *Tbx2* and *Tbx3* are expressed within a strip of cells at the anterior and posterior margins of the limb bud and are involved in posterior digit identity (Davenport et al., 2003; Suzuki et al., 2004). The lack of limb abnormalities in *Tbx18* null embryos could stem from functional redundancy within this expression domain. Ectopic Tbx6 expression in our 3-component embryos results in limb phenotypes that are similar to a loss of *Tbx15*, namely changes in the size and shape of the limb bones and a hole in the scapula blade. Tbx15 functions as a transcriptional repressor, however direct targets of Tbx15 are currently unknown. The hole or foramen in the scapula is common to both null mutations in *Tbx15* and conditional removal of Tbx5 after the initiation of limb bud outgrowth (Hasson et al., 2007). *Fgf10* is a direct target of Tbx5 in limb bud mesenchyme, and functions to maintain *Fgf8* expression in the AER in a feedback loop (Agarwal et al., 2003). We observed normal expression of *Fgf8*, indicating that Tbx5 function is not likely to be affected, however we cannot exclude the possibility that ectopic Tbx6 affects binding of Tbx5 to other downstream targets or Tbx5 transcriptional ability at a level that cannot be observed by *in situ* hybridization.

Competition between T-box transcription factors for binding to endogenous targets within overlapping expression domains is not a novel concept. During mouse heart development, Tbx2, a transcriptional repressor, competes with Tbx5, a transcriptional activator for binding to the *ANF* promoter within the atrio-ventricular canal and outflow tract (Habets et al., 2002). We propose a similar competitive mechanism in our 3-component embryos, where ectopic expression of Tbx6 within the limb buds and somites interferes with the transcriptional

repressors Tbx15 and Tbx18 through a competition for binding sites in the enhancers of downstream targets. Competition between Tbx6 and Tbx18 for binding to the T-box binding sites within the *Dll1* enhancer has been previously proposed (Farin et al., 2007). Herein we provide *in vivo* support of such competition.

Our luciferase assays suggest that Tbx6 and Tbx18 directly compete at the *Dll1-msd* promoter. This could indeed reflect the *in vivo* situation observed in our 3-component embryos. Tbx6 is required in the PSM to activate *Dll1* expression (White and Chapman, 2005), and Tbx18 directly represses *Dll1* within the rostral portion of each somite, restricting its expression to the caudal half (Bussen et al., 2004). Similar competition is observed at the *Dll1-msd* enhancer between Tbx6 and Tbx15, suggesting a common mechanism to compete for binding to enhancer regions. We proposed that this competition occurs via direct binding of Tbx6 to the enhancer regions of Tbx15/Tbx18 downstream targets, thus precluding Tbx15/Tbx18 binding. This competition is not exclusive to the *Dll1* enhancer, as we also noted that competition could occur at the *ANF* enhancer in luciferase assays. Farin et al. hypothesized that Tbx18 may compete with Tbx5 at the *ANF* promoter in the heart to inhibit expression within the sinus horn mesenchyme (Farin et al., 2007). Tbx6 only weakly activates transcription from the *ANF* promoter, however *ANF* is not an endogenous Tbx6 target. Nevertheless, Tbx6-mediated activation decreased in a linear fashion with the addition of Tbx18. Addition of Tbx15 did not decrease Tbx6 activation, which could reflect inefficient binding of Tbx15 to the *ANF* promoter, or may be because considerably less myc-Tbx15 is produced compared to myc-Tbx6 when equivalent amounts of plasmid were transfected.

It is interesting to note that although the closely related Tbx18 and Tbx15 both interact with the Groucho co-repressor via a conserved eh1 domain within the N-terminus of the protein

to serve as transcriptional repressors (Farin et al., 2007), Tbx18 competes more efficiently than Tbx15 in our luciferase assays at both tested enhancers. Again, this result may be due to the amount of protein produced: Tbx18 is expressed at a 2-fold greater level than Tbx6, while Tbx15 is expressed at a 1.7-fold lower level than Tbx6. These differences in expression level could contribute to their ability to compete with Tbx6 in our luciferase transcriptional assays.

PCR-based binding site selection assays demonstrated that both Tbx15 and Tbx18 bind to the canonical 5'-AGGTGT-3' sequences and prefer a GA dinucleotide 3' of the half-site. Differences in binding site selection between Tbx18 and Tbx15 may lie in preferences for orientation of multiple half-sites. Tbx18 exclusively preferred a palindromic binding sequence, while Tbx15 equally selected palindromic sequences and direct repeats of the core sequence (Farin et al., 2007). This may result in differential ability to compete with Tbx6 at each of the enhancers tested. Another possibility that is not mutually exclusive is that Tbx18 could simply serve as a stronger transcriptional repressor than Tbx15. This is consistent with previous observations whereby Tbx18 fused to the Gal4 DNA binding domain repressed transcription from a Gal4-UAS synthetic enhancer more effectively than Tbx15 (Farin et al., 2007).

In the absence of additional targets for these transcription factors and antibodies suitable for ChIP, we can only speculate at the mechanism underlying the phenotypes generated by ectopic Tbx6 expression. Based on these phenotypes and our luciferase assay results, we suggest that competition for binding to endogenous enhancers occurs between the ectopically expressed Tbx6 and endogenously expressed Tbx15 and Tbx18 within somitic tissue and the limb bud mesenchyme. The DNA binding domain of Tbx18 and Tbx15 are also able to compete with Tbx6 to effectively lower its activity, further suggesting that the observed competition occurs at the level of DNA binding. Alternatively, ectopic expression of Tbx6 may compete for a

common co-factor of Tbx15 and Tbx18 that binds within the DNA binding domain. Previous studies have suggested that Tbx15 and Tbx18 may interact with the paired box transcription factor, Pax3 via the DNA binding domain (Farin et al., 2008). It is not known whether Tbx6 can interact with Pax3.

In tissues where multiple T-box factors are normally co-expressed, competition likely represents a mechanism that controls proper transcription of target genes. This competition would be held in check when all factors are expressed at their proper levels; however in situations whereby the relative levels of T-box transcription factors are altered would result in inappropriate target gene expression. Haploinsufficiency for human *TBX1, TBX3, TBX4, TBX5,* and *TBX22* result in the syndromes DiGeorge, ulna-mammary, Small Patella, Holt-Oram, and X-linked cleft palate with ankyloglossia, respectively (Baldini, 2003; Bamshad et al., 1997; Basson et al., 1997; Bongers et al., 2004; Braybrook et al., 2001). Developmental processes are therefore exquisitely sensitive to the levels of T-box transcription factors. Accordingly, reducing the effective amount of any particular T-box factor could allow for competition between resident T-box factors and this competition could contribute to the observed phenotypes. This highlights the importance of maintaining not only proper expression domains, but also the appropriate levels of T-box transcription factors during development.

5.0 CONCLUSIONS AND FUTURE DIRECTIONS

T-box transcription factors are critical regulators of cell type specification, differentiation, and proliferation during embryonic development. My thesis work has centered around two particular family members; T and Tbx6 and how they control cell fate decisions in their endogenous domains. In addition, we investigated how Tbx6 interacts with other T-box transcription factors when it is ectopically expressed. While this work focuses on one family of transcription factors, it implies that transcription factors that share a conserved DBD may also compete for downstream targets *in vivo* when they are co-expressed and that this competition contributes to the overall developmental dynamics of the organism.

5.1 CONCLUSIONS

5.1.1 T and Tbx6 compete for binding at the *Dll1-msd* enhancer

T and Tbx6 are expressed in an overlapping domain in the PS and tailbud, and bind to the same core 5'-AGGTGT-3' sequence, leading to the hypothesis that T and Tbx6 compete for a common subset of shared targets *in vivo*. Additional genetic evidence for competition between these two factors *in vivo* is discussed in detail in section 2.1.1. We chose to use the *Dll1-msd* enhancer as a model to further characterize the potential competition between these two factors because *Dll1* is

a known downstream target of Tbx6 and is co-expressed in the PS and tailbud with both T and Tbx6 (White and Chapman, 2005). The work described in this thesis has demonstrated that *Dll1* is not only a target of Tbx6, but also of T, the first known downstream target of T in the mouse embryo. Furthermore, we demonstrated that Tbx6 and T can differentially activate the *Dll1-msd* enhancer in luciferase transcriptional assays, and that addition of increasing amounts of T to a constant amount of Tbx6 lowers the overall RLUs observed in a dose-dependent manner. We propose that the observed differential activity is the result of distinctive binding preferences of T and Tbx6 for the four putative T-box binding sites within the *Dll1-msd* enhancer. In addition, as a non-exclusive alternative, T and Tbx6 may compete for a common co-factor that could result in the observed competition in our luciferase assays. We believe that in the case of *Dll1-msd*, competition for physically binding at the enhancer plays a role in the transcriptional activity of T and Tbx6 when co-expressed in our luciferase assays based on ChIP and EMSA data, although competition for a common co-factor may also contribute.

Tbx6 binds to BS1 and BS2 with similar affinity, while T only binds significantly to BS2, with a ten-fold lower affinity than Tbx6. This difference in binding affinity may account for the differences in transcriptional activity levels noted in our luciferase assays. We therefore hypothesize that the observed competition between T and Tbx6 in our luciferase assays is due to the fact that addition of increasing amounts of T simply increases the stochastic chance that T will be bound at BS2 rather than Tbx6. These assays do not take into account the possibility that either T or Tbx6 may bind any combination of the four binding sites in *Dll1-msd* cooperatively. Cooperative binding of Tbx6 to the *Dll1-msd* enhancer may explain why such a significant drop in RLUs is observed upon the addition of relatively small amounts of T in our luciferase assays. Future experiments to address potential cooperativity are outlined in section 5.2.2.

It has been previously noted that T-box transcription factors prefer different flanking sequences outside of the core 5'-AGGTGT-3' sequence and also orientation of the half-sites. Contrary to this observation, inspection of the crystal structures of the DBD of T and Tbx3 reveals no obvious structural reasons why T-box transcription factors should prefer different flanking nucleotides, as all amino acids that make direct contact with the DNA are highly conserved (Coll et al., 2002; Muller and Herrmann, 1997). Despite this, PCR based binding site selection experiments have demonstrated that T prefers a GA dinucleotide in the +1 and +2 positions 3' to the core sequence, while Tbx6 prefers N(A/G), (where N is any nucleotide) (Kispert and Hermann, 1993; White and Chapman, 2005). Accordingly, both T and Tbx6 are able to bind BS2, which contains a GA dinucleotide 3' to the core sequence, while T does not significantly bind BS1, which contains a TG dinucleotide. Interestingly, neither T nor Tbx6 bound BS3, which contains a consensus core sequence followed by a 3' CC dinucleotide that is not preferred by either. This demonstrates that nucleotides 3' to the core sequence are indeed important for T-box factor specificity, although perhaps this is an artifact of the *in vitro* assay system. It is possible that yet unknown, distinct factors may interact with T and Tbx6 to influence the binding preferences for flanking sequences.

Overall, this work has contributed to our understanding of how co-expressed T-box transcription factors may interact when co-expressed in a tissue *in vivo.* Prior to this work, it was unknown how T and Tbx6 interacted in the PS where they are co-expressed. Several studies have documented competition for binding to enhancer regions and common co-factors between co-expressed T-box transcription factors in other tissues (Farin et al., 2007; Goering et al., 2003; Habets et al., 2002). Specifically, this work implies that there is a common subset of genes shared between T and Tbx6, and that a competition may exist *in vivo* for binding at the enhancers

of these genes. Additional competition between T and Tbx6 may occur via competing for a common co-factor, although this is difficult to investigate further as there is no known interacting partners of murine T or Tbx6 known to date. Experimental strategies to elucidate both common and unique downstream targets of both Tbx6 and T are outlined in section 5.2.3.

5.1.2 *In vivo* relevance of T and Tbx6 competition

We hypothesize that T is required to maintain cells in a PS-like state, while Tbx6 is required to push cells toward a PSM fate. In this scheme, maintaining the correct relative levels of Tbx6 and T is essential for the proper cell fate determination. The phenotypes of embryos with altered levels of Tbx6 and T relative to each other also support this hypothesis. For example, increased amounts of Tbx6 relative to T in the *Tg46* homozygous embryos resulted in axis truncations similar to a *T* heterozygous phenotype, perhaps indicating that cells are prematurely exiting the PS and adopting a PSM fate.

Furthermore, we hypothesize that there is a common subset of transcriptional targets shared between Tbx6 and T in the tailbud, where they may have inherently different transcriptional activities. T may function at these targets to attenuate a Tbx6-based transcriptional program, thereby allowing cells to maintain a PS-like fate and continue to extend the body axis. Cells that accumulate more Tbx6 relative to T would then exit the PS and adopt a PSM fate. In this manner, transcriptional interplay between Tbx6 and T may regulate the transition between PS and PSM fate *in vivo*.

5.1.3 Ectopic expression of Tbx6 results in competition with other T-box transcription factors

In addition to the competition between Tbx6 and T in the PS and tailbud of the embryo, we have also demonstrated that Tbx6 can compete with other T-box transcription factors when expressed outside of its endogenous expression domain. We developed a 3-component transgenic system that allowed us to drive ectopic expression of Tbx6 in the PSM and in all downstream derivatives in particular the somites, in addition to the limb buds. Although expression was highly mosaic, common skeletal defects were noted in 3-component embryos treated with DOX at e8.5 and dissected at e13.5. These included shortened and malformed fore- and hindlimbs, expansion of the vertebral pedicles, fusion and malformations of the atlas and axis, missing or underdeveloped digits and distal portions of the ribs, and occasionally an ectopic foramen within the scapular blade. These phenotypes are strikingly similar to those observed in *Tbx18* and *Tbx15* null embryos (Bussen et al., 2004; Singh et al., 2005), suggesting that Tbx6 has the ability to compete with other T-box transcription factors when expressed outside of its endogenous domain. Although it had been previously proposed that Tbx6 and Tbx18 compete for binding at the *Dll1* enhancer (Farin et al., 2007), our work supplies the first *in vivo* evidence of competition between these factors. Although we hypothesize that this competition occurs at the level of DNA binding, it is certainly possible that ectopic expression of Tbx6 competes for a common co-factor of Tbx15 and Tbx18. Tbx15 and Tbx18 interact with the paired box protein, Pax3 via their DNA binding domains (Farin et al., 2008). It is not known whether Tbx6 has the ability to interact with Pax3, although it is certainly possible that Tbx6 can also interact with Pax3, thereby reducing the ability of Tbx15 or Tbx18 to exert their transcriptional effects *in vivo*.

In our luciferase assays, myc-Tbx18 was a much better competitor than Tbx15 at both tested enhancers (*Dll1-msd,* and *ANF*). PCR-based binding site selection assays revealed that both Tbx15 and Tbx18 bind the same sequences, however Tbx18 exclusively preferred a palindromic binding sequence, while Tbx15 equally selected palindromic sequences and direct repeats of the core sequence (Farin et al., 2007). This indicates that like T and Tbx6, differences in binding preferences may exist between Tbx18 and Tbx15. These differences may ultimately result in the differential ability to compete with Tbx6 at each of the tested enhancers. Additionally, it has been previously shown that Tbx18 is a better transcriptional repressor at a synthetic enhancer compared to Tbx15 (Farin et al., 2007). We observed differential expression of Tbx18 and Tbx15 relative to Tbx6 that may play a role in their lowered ability to repress a Tbx6-mediated transcriptional program. The studies by Farin et al. did not examine the expression levels of these proteins relative to one another. Overall, this data suggests that ectopically expressed Tbx6 is able to compete at the level of DNA binding with other T-box transcription factors resulting in phenotypes resembling *Tbx15* and *Tbx18* null embryos. There are no known downstream targets of Tbx15, and *Dll1* and ANF are the only known downstream targets of Tbx18. Once additional targets of Tbx15 and Tbx18 are identified, it will be interesting to examine how their expression level changes when Tbx6 is misexpressed. We predict that ectopic expression of Tbx6 may result in an upregulation of Tbx15 and Tbx18 target genes in this ectopic expression domain, as both Tbx15 and Tbx18 serve as transcriptional repressors.

5.1.4 Competition between other transcription factors with conserved DNA binding domains

Proper temporal and spatial gene expression requires a high degree of transcription factor specificity for recognition sequences. This represents a daunting task if one considers that a random 6-mer sequence will be represented within the human genome greater than 700,000 times just by chance (Georges et al., 2010).

Although this work focuses on the T-box family of transcription factors, there are other families of transcription factors that share a conserved DBD, and therefore bind similar DNA sequences. This implies that this work can be extrapolated to other paralogous transcription factors co-expressed either during development or in adult tissues. Within these paralogous transcription factor families limited differences in specificity for binding sites or interactions with different co-factors that modify behavior partially encompass mechanisms of specific modulation of target genes. Two examples of other well-characterized families of transcription factors with similar properties to the T-box family of transcription factors are the Forkhead box and Homeobox transcription factors.

5.1.4.1 Forkhead box (Fox) transcription factor family specificity

Fox transcription factors contain a 110 amino acid DBD that is conserved from yeast to humans. Humans have over forty Fox proteins, grouped into subfamilies that are varied and share little sequence similarity or global domain architecture outside of the DBD. Similar to the T-box family, the conserved nature of the Forkhead DBD allows for binding to a consensus 5'-(G/A)(T/C)(A/C)AA(C/T)A-3' sequence (Friedman and Kaestner, 2006). Specificity for particular nucleotides within this consensus sequence are subfamily dependent and can partially,

but not entirely, explain Fox transcription factor specificity (Georges et al., 2010). As with T-box transcription factors, direct sequence-specific contact between the Fox DBD and DNA is primarily achieved through invariantly conserved residues (Littler et al., 2010). Additional levels of specificity of the Fox transcription factor family has been partially attributed to unique affinities for tandem copies (either palindromic repeats or direct repeats) of the consensus binding site, although these are rare and poorly conserved throughout the genome (Korver et al., 1997; Littler et al., 2010), suggesting that additional levels of specificity exist for Fox transcription factors. As with T-box factors, concurrent binding of other transcription factors is essential for activation of downstream targets. For example, cis-binding of both Foxa1 and the glucocorticoid nuclear hormone receptor is required during the fasting response to regulate expression of downstream targets in the adult mouse, such as *proglucagon* (Kaestner et al., 1999; Zhang et al., 2005). Additionally, Fox transcription factors also interact directly with other transcription factors to regulate their downstream targets. For instance, Engrailed, a homeodomain transcription factor, directly interacts with Foxa2 to regulate the downstream target *MAP1B,* which encodes a neuronal associated microtubule binding protein that is important for the development of the nervous system (Foucher et al., 2003). This data in combination with what is known about the regulation of T-box targets indicates a common theme amongst families of paralogous transcription factors for proper regulation of downstream targets *in vivo.*

5.1.4.2 Homeobox (Hox) transcription factor family specificity

Hox genes encode transcription factors found in all animal species that act during development and contain a highly conserved sixty amino acid motif (the homeodomain) that binds DNA in a sequence-specific manner (Levine and Hoey, 1988). In mammals, there are 39 Hox genes

organized in four clusters (A-D) on different chromosomes, and are thought to have arisen through cis-amplification and trans-duplication (Scott, 1992). Similar to T-box transcription factors, certain Hox transcription factors are more effective at activating transcription from some enhancer regions than other Hox transcription factors. Thus, the overall transcriptional output is the result of a competition between co-expressed Hox genes (Goff and Tabin, 1997). In support of this hypothesis is the observation that misexpression of Hoxd-13 in the developing chick limb appears to act in a dominant-negative manner to suppress Hoxd-11 mediated transcription (Goff and Tabin, 1997). This highlights the importance of the relative levels of paralogous transcription factors.

Crystal structures of several homeodomain proteins bound to the consensus 5'-TAAT-3' motif revealed conservation of amino acids that directly contact the DNA (Kissinger et al., 1990; Piper et al., 1999; Qian et al., 1989). As with T-box transcription factors, the flanking nucleotides 3' of the conserved core binding sequence directly influence DNA binding affinity (Pellerin et al., 1994), despite the fact that the residue that makes contact with the nucleotides 3' of the core sequence is completely conserved across all family members (Ekker et al., 1992). The perplexing issue of altered affinity for flanking sequences when all amino acids making direct DNA contacts are highly conserved is clearly a common theme in transcriptional regulation by paralogous transcription factors.

Hox protein specificity is also partially determined by interactions with co-factors. Many identified Hox binding sites are paired with Pbx/Meis binding sites in target genes. Several Hox proteins directly interact with Pbx proteins to form a higher order complex where each component binds DNA to regulate transcription (Svingen and Tonissen, 2006). In the case of Hoxb8, which specifies thoracic vertebrae, mutation of the Pbx interaction domain results in an

anterior homeotic transformation, similar to a *Hoxb8* null phenotype (Medina-Martinez and Ramirez-Solis, 2003).

Overall, the similarities between target regulation in the Hox family and T-box family indicates a common theme for paralogous families of transcription factors in target regulation.

5.1.5 Summary

Integration of multiple strategies likely helps differentiate transcription factors with related DBDs to ensure specific downstream transcriptional responses. Arrangement and spacing of binding sites, flanking nucleotides, synergistic activation of transcription in concert with other transcription factors, and direct interaction with a wide array of other protein modifiers can act in a complementary fashion to ensure specificity of a particular transcription factor. Each of these effects alone may not entirely explain transcription factor specificity, but in combination may allow for exquisitely sensitive control of this process. We hypothesize that results from studies within this thesis can therefore be applied not only to the T-box family of transcription factors, but can be extrapolated to other families such as the Fox and Hox families with conserved DBDs.

5.2 FUTURE DIRECTIONS

5.2.1 Potential uses for 3-component fibroblasts

We have demonstrated that approximately 50% of fibroblasts derived from 3-component embryos express myc-Tbx6. These myc-Tbx6 positive fibroblasts were derived from mesodermal tissue cell types. As a true 'embryonic mesoderm' cell line does not exist, these fibroblast cell lines may also be useful for studies to identify interacting proteins and protein modifications in a more relevant embryonic mesoderm cell type. As these fibroblasts are not immortalized, they can only be passaged approximately four times before they senesce. This means that access to a source of the embryos themselves must be maintained to perform multiple rounds of these types of experiments. To this end, one could utilize the population of cells for immunoprecipitation followed by mass spectrometry to identify potential Tbx6 interacting proteins. These potential interacting proteins may not be expressed in other commercially available cell lines commonly used for investigation of T-box transcription factors, such as HEK293T, HeLa, or Cos-7. To date, there are no known interacting partners of murine Tbx6.

5.2.2 Investigation of potential cooperativity at the *Dll1-msd* enhancer

While our luciferase transcriptional assays demonstrate Tbx6 and T activate transcription from the *Dll1-msd* enhancer at different levels, these assays did not take into account the possibility of cooperative binding of any combination of the four binding sites. Cooperative binding of Tbx6 to the *Dll1-msd* enhancer may explain why a significant drop in RLUs is observed upon the addition of relatively small amounts of T in our luciferase assays. Although T does not bind

significantly to BS1, in this scenario, binding of T to BS2 would block the ability of Tbx6 to bind not only to BS2, but also BS1, thereby greatly reducing the overall RLUs measured when a small amount of T is added.

To investigate the potential role of cooperative binding, we propose to make site-directed mutants of each of the four BS individually, and in combination and test how these mutations affect T and Tbx6. Based on our EMSA data, we hypothesize that mutation of BS3 and BS4 will have no effect on Tbx6 or T transcriptional activity. We expect that mutation of BS2 would effectively mitigate T's transcriptional activation, as T only binds significantly to BS2. In contrast, if there is no cooperativity of binding, mutation of BS2 should not completely diminish Tbx6 activation, as it would still be able to bind BS1. Mutation of BS1 may only affect the transcriptional activity of Tbx6, as T did not significantly bind BS1. If mutation of BS1 affects T's transcriptional activity, it would suggest that BS1 is a physiologically relevant site for T.

We can also make combinatorial mutations, such as a mutation in BS2 and BS1, which we predict would abrogate all transcriptional activity from the *Dll1-msd* enhancer for both T and Tbx6. If transcriptional activity were not completely extinguished for both T and Tbx6, it would indicate that BS3 and BS4 could potentially play a role in activation from the enhancer at a level that cannot be detected in our EMSA experiments. Cooperativity in binding can enhance specificity of transcription factors and generally implies dimerization between the two transcription factors. This dimerization leads to a sigmoidal response to transcription factor concentrations, allowing for a exquisitely sensitive molecular switch (Georges et al., 2010). This is interesting in that sigmoidal binding behavior was observed in our quantitative EMSA experiments using BS1 and BS2 of the *Dll1-msd* and the DBD of Tbx6 and T.

Overall, dissecting of the role of each binding site should lend insight into potential differences and/or similarities in T and Tbx6 transcriptional regulation from the *Dll1-msd* enhancer. Additionally, these assays should allow us to investigate the potential role of cooperativity for binding each of the four binding sites.

5.2.3 ChIP-seq to identify common and unique targets of Tbx6 and T

In this work, we demonstrated that Tbx6 and T share the downstream target *Dll1*, and that they have inherently different transcriptional activities at the *Dll1-msd* enhancer in our luciferase assays. We hypothesize that there are additional shared targets of Tbx6 and T, in addition to unique targets for each. To identify these targets, we plan to use ChIP followed by next-generation sequencing (ChIP-seq). To this end, we have currently collected over 1,500 tailbud and somitic tissue fragments for use in ChIP-seq experiments. We estimate 700 tissue fragments are necessary for each antibody used.

Briefly, the samples will be subjected to identical immunoprecipitation conditions as for the ChIP experiments discussed in section 2.4.3, and subsequently sent to the University of Pittsburgh Center for Genomics and Proteomics, who will further prepare samples to run on the Applied Biosystems' SolID3 massively parallel sequencer. Sequences generated from immunoprecipitated fragments will be compared to sequences generated from input DNA to account for biases in sequencing and shearing efficiency. Enrichment of the immunoprecipitated regions as compared to input DNA will delineate where each factor is bound within the genome. Furthermore, The Center for Genomics and Proteomics will develop algorithms to search within enriched sequences for the canonical 5'-AGGTGT-3' sequences. We expect that true targets of Tbx6 and T will harbor this sequence within the immunoprecipitated region. A similar approach

has been used to identify common and unique downstream targets of two related transcription factors, FoxA1 and FoxA3, in the development and specification of the hepatic lineage (Motallebipour et al., 2009).

Comparison of the selected genomic regions from Tbx6 and T will allow us to begin to uncover shared and unique targets of each factor. Sequence data can be compiled to begin to identify differences in orientation of half-sites and flanking nucleotides preferences of Tbx6 and T *in vivo*, and compared to *in vitro* PCR-based binding site selection data. Once these targets are uncovered, we can use our genetic toolbox to alter the expression levels of Tbx6 and T within the tailbud, followed by directed ChIP and quantitative PCR to determine whether occupancy at these targets is altered when relative expression levels of Tbx6 or T is changed.

Lastly, we plan to search the enhancer regions of T and Tbx6 targets for potential binding sites of other transcription factors. Expression of the four known downstream targets of Tbx6 is dependent on not only Tbx6, but also other signaling pathways (see section 1.4.4), ensuring tight regulation of expression of these targets. For example, expression of *Dll1* within the PSM depends on the presence of both Tbx6 and LEF/TCF transcription factors (Hofmann et al., 2004). We predict that other downstream targets of Tbx6 and T will also require input from other signaling pathways to properly regulate expression of these targets.

The enhancer regions of target genes can then be further dissected using cell-based assays and biochemical techniques as is described in this work to characterize how T and Tbx6 act independently and together to regulate expression of downstream targets *in vivo.* The broad concepts derived from this work can be applied to other tissues in development with co-expression of more than one T-box transcription factor, as well as other families of transcription factors such as the Fox family and the Hox family that share a conserved DBD.

6.0 MATERIALS AND METHODS

6.1 GENERAL CELL CULTURE AND TRANSFECTIONS

Cell lines were maintained in a 37^{o}C incubator with 5% CO_2 and passaged regularly (every 2-3 days). HEK293T, NIH3T3, and isolated embryonic fibroblast cells were grown in DMEM + 10% fetal calf serum, 1x Penicillin/Streptomyocin, and 1x L-glutamine with regular media changes. All transfections were performed with cells in suspension using 5μL of Lipofectamine (Invitrogen) 2000 per 35mm dish and DNA diluted in Optimem media (Gibco).

6.2 WESTERN BLOTTING

Tissue dissected from embryos or tissue culture cells were treated with RIPA buffer with added protease inhibitor cocktail (Sigma); tissue was also dounce homogenized to facilitate lysis. When indicated, Bradford dye assay was performed to determine total protein concentration to ensure equal loading. Lysates were loaded onto a SDS-PAGE gel, and transferred to nitrocellulose (Immobilon). The membrane was blocked in TBTT + 5% dry milk (blocking buffer), and then incubated in the appropriate primary antibody (1:500 rabbit-anti-Tbx6, 1:500 mouse-anti-myc (Sigma), 1:500 goat-anti-T (Santa Cruz), 1:500 mouse-anti-tubulin (Sigma) diluted in blocking buffer, and subsequently incubated in horseradish peroxidase-conjugated

secondary antibody (1:2500 dilution in blocking buffer). Some blots were subsequently stripped and re-probed with anti-tubulin (1:500 dilution, Sigma) in blocking buffer to ensure equal loading.

6.3 IMMUNOFLUORESCENCE

Cells were transiently transfected in 6-well plate with a sterilized coverslip. Twenty-four hours later, cells were fixed in 4% paraformaldehyde in PBS, and permeablized to allow antigen retrieval in PBS + 0.1% Triton X-100. 5% normal goat serum (NGS) or normal donkey serum (NDS, for goat-anti-T primary antibody) in PBS was used as a blocking solution. Primary antibodies were diluted in blocking serum: 1:250 mouse-anti-myc (Sigma), 1:250 rabbit-anti-Tbx6, 1:250 goat-anti-T (Santa Cruz) for one hour at room temperature. Coverslips were washed three times in PBS. Appropriate secondary antibodies (Alexa fluor, Invitrogen) at 1:500 concentration with 1:2000 TO-PRO3 were subsequently diluted in blocking serum and added for one hour. Coverslips were washed three times in PBS and mounted in Vectashield mounting medium.

All images were captured on a BioRad scanning laser confocal microscope, and assembled using ObjectImage2.10 and Photoshop.

6.4 PREPARATION OF NUCLEAR LYSATES

Nuclear lysates were prepared according to a protocol adapted from Wadman et al. (Wadman et al., 1997). Briefly, $5x10^7$ cells (approximately one confluent well of a 6-well plate) were washed twice with ice cold PBS and lysed with 500μL of Buffer A (10mM HEPES, 1.5 mM $MgCl_2$, 10mM KCl, 0.5 mM DTT, with added protease inhibitors, pH 7.9) and incubated on ice for 15 minutes. NP-40 was added to a final concentration of 0.5%, and cells were vortexed for 10 seconds. Nuclei were pelleted by centrifugation at 6500g for 30 seconds. Supernatant from this spin was considered the "cytoplasmic fraction". The nuclear pellet was resuspended in 150μL Buffer C (20mM HEPES, 1.5mM $MgCl_2$, 420mM NaCl, 0.2mM EDTA, 25% v/v glycerol, with protease inhibitors, pH 7.9) and rotated at 4^oC for 30 minutes. Lysed nuclei were centrifuged at full-speed for 10 minutes, and the supernatant was collected as the "nuclear fraction".

6.5 EMSA

Sources of proteins included nuclear lysates prepared from transfected HEK293T cells or purified proteins, as indicated. In all cases, a Bradford assay was performed to determine total protein concentration, and equivalent amounts of total protein were used in EMSA reactions. 3μg of total protein from nuclear extracts were used. Amounts of purified proteins used are as indicated.

Oligonucleotides were end-labeled with γ-^{32}P-ATP using T4 polynucleotide kinase. Oligonuclotides were then annealed by placing in a water bath at 95^oC for 5 minutes, then allowing to cool to room temperature (power turned off on water bath). After annealing,

purification was performed using Micro Bio-Spin P-30 Tris purification columns (BioRad). The percentage of double-stranded versus single-stranded probe was determined by running probes on an 18% non-denaturing PAGE, and quantitating the fraction of double stranded probe using the Fuji BAS-2500 Phosphoimager and ImageGauge software. The percentage of double stranded probe recovered was greater than 85% in all cases. Subsequently, the percentage of double-stranded probe for each experiment was standardized so equivalent amounts were used. All EMSA binding reactions were prepared in a final reaction volume of 10μL in BBT buffer (25mM HEPES pH 7.4, 75mM NaCl, 1mM DTT. 0.25mM EDTA, 0.1% NP-40, 1mM $MgCl_2$, 10% glycerol, 10μg/mL BSA). 0.1mg/mL Poly dI-dC was added as a non-specific competitor.

Binding reactions were incubated at room temperature for 20 minutes before addition of 1μg of antibody (anti-Tbx6 or anti-myc). After addition of antibody, reactions were incubated at room temperature for 15 minutes before loading on 4-6% non-denaturing PAGE (37.5:1) run in 1x TAE. Gels were dried unfixed, exposed to a phosphoimager screen and imaged on a Fuji BAS-2500 Phosphoimager.

Oligonucleotides used: (bold letters indicate core binding sequence)

T^{bind}: 5'-CTAGTC**ACACCTAGGTGT**GAAATT-3'

T^{half}: 5'-ATCGAATTC**AGGTGT**GAAATTGGATCCACT-3

T^{mut}: 5'-CTAGTC**ACACCT**ATTTTTGAAATT-3'

Dll1BS1: 5'-TCACTGT**AGGTGT**TGCTGTCCTGT-3'

Dll1BS2: 5'-TCCCG**AGGTGT**GATTCTTGGA-3'

Dll1BS3: 5'-GTGGATCC**AGGTGT**CCTCACTGGGCTGC-3'

Dll1BS4: 5'-TGGATCCT**AGG**G**TGT**ACCTGACGGCTGC-3'

6.6 PRODUCTION OF RECOMBINANT TBX6-DBD AND T-DBD

The region of T and Tbx6 corresponding to the T-box DNA binding domain (Tbx6: AA 94-195; T: AA 41-223) was PCR amplified and cloned into the pTOPO151 vector, producing a Histidine-tagged fusion protein. Transformed bacterial cultures were auto-induced, lysed and His-tagged fusion proteins were purified over a nickel-NTA column. Subsequently a sizing column was utilized to exclude large aggregates. Limited trypsin proteolysis revealed that >90% of isolated, purified proteins were correctly folded.

6.7 QUANTITATIVE EMSA

EMSA reactions were prepared as described in section 6.6, except increasing amounts of Tbx6-DBD (range: $2.1 \times 10^{-8} – 2.1 \times 10^{-5}$M) or T-DBD (range: $4.0 \times 10^{-6} – 2.4 \times 10^{-5}$M) were added to a constant, limiting amount of labeled BS1-4 oligonucleotides (10pM) and incubated one hour at room temperature to ensure reactions were at equilibrium. Reactions were run on a 6% non-denaturing PAGE. Quantitation was performed as described in (Harada et al., 1994). Briefly, the amount of free and bound DNA was quantitated using a Fuji BAS-2500 phosphoimager and analysis with ImageGauge software. Percentage of bound DNA was determined by the following formula: (Shifted DNA)/(Shifted DNA+Free DNA). The concentration of Tbx6-DBD or T-DBD was plotted versus the percentage of DNA bound. The data was fit to a 3-parameter Hill equation using SigmaPlot software (equation: $y= ax^b/(c^b + x^b)$, where a= the maximum value of y (percent bound), b= the Hill co-efficient, and c= K_d).

6.8 LUCIFERASE ASSAYS

$1x10^5$ HEK293T cells were plated per well in a tissue culture treated 96-well dish, and transfected in suspension. 10-25ng of the indicated luciferase reporter vector was transfected per well along with 1 ng of *pRenilla Luciferase-CMV*, which served as an internal control. Amounts of protein expression construct are as indicated in experiments and empty pCS3 vector was transfected where needed to keep amounts of transfected DNA constant. Twenty-four hours after transfection, cells were processed with Dual Glo luciferase reagent (Promega) according to manufacturer's directions, and intensity measured on a Berthold XS^3 LB960 luminometer. Luciferase readings were normalized to the Renilla luciferase, and ratios were normalized to levels of luciferase reporter when transfected with a protein empty vector control as follows: (Luciferase reading/Renilla reading) = Ratio for each condition. Ratio for each condition was then divided by the ratio obtained with the empty vector control to account for background activation of the reporter. Mean values for results performed in triplicate, and repeated at least twice were graphed and standard error was calculated (Standard Error = standard deviation/square root of number of replicates).

6.9 CHROMATIN IMMUNOPRECIPITATIONS (CHIP)

Confluent 100mm dishes of cells or approximately 50 dissected tailbuds or equivalent portions of somitic tissue were processed using the ChIP kit from Millipore, according to the manufacture's directions. Briefly, formaldehyde at a final concentration of 1% in PBS was added to cells or tissue and incubated at 37^oC for ten minutes. After fixation, cells were lysed

with SDS lysis buffer (provided in kit) with added protease inhibitor cocktail (Sigma) for ten minutes on ice. 200μL was added for every $1x10^6$ cells or fifty dissected tissue pieces. Dissected tissue was also dounce homogenized at this time to break up large tissue pieces. The released DNA with bound proteins was sonicated three times for three seconds each to shear the DNA to 200-1000 base-pair fragments. Shearing efficiency was initially tested by collecting 20μL of sheared lysate and running it on a 1.5% agarose gel. The lysate was pre-cleared with protein sepharose A or G, depending on the antibody used, for thirty minutes at 4^oC. 2.5% of the total volume was removed as "input". Lysates were split into two pools: one which had primary antibody added (3μg goat anti-T, Santa Cruz Biotechnology or rabbit anti-Tbx6) and one pool which had no antibody added, overnight at 4^oC. Protein sepharose A (rabbit anti-Tbx6) or protein sepharose G (goat anti-T) were subsequently added to collect all antibody bound fragments. Washing and collection of DNA fragments was performed according to ChIP kit manufacturer's directions except Qiagen PCR purification columns were used to purify immunoprecipitated DNA rather than Phenol:Sevaq extraction. Two separate radioactive PCR reactions were performed using 2μL of elutate with a standard PCR protocol. An extension time of 30 seconds with an annealing temperature of 55^oC for *Dll1* and 65^oC for *Wnt7a* for a total of 25 cycles was utilized.

PCR primers are as follows:

Dll1-Fwd: 5' – CATGCGAAGGTTTTCCTC – 3'

Dll1-Rev: 5' – CCATCTTTAGAGGGGCC – 3'

Wnt7a-Fwd: 5' - CTCTTCGGTGGTAGCTCTGG – 3'

Wnt7a-Rev: 5' – CCTTCCCGAAGACAGTACGC – 3'

PCR products were run on a 6% polyacrylamide gel (29:1) in TAE, exposed to a phosphoimager plate, and imaged with a Fuji-BAS-2500 phosphoimager.

6.10 GENERATION OF *DLL1-MSD CRE* RECOMBINASE TRANSGENIC MICE

The *Dll1-msd* enhancer element was PCR amplified from C57Bl6/J wild type genomic DNA using the following primers: 5'-GAAGAAGAAGAGCAGGAAGGAG-3' and 5'-AGGTTTTTACACATCCATCAG-3'. The resulting 1.5 kb product was digested with NotI and XbaI and the *Dll1-msd* enhancer element was cloned into the NotI/SpeI sites of pAKH-Cre9, upstream of the β-globin minimal promoter, a nuclear localized Cre recombinase and β-actin polyadenylation signal (pAKH gift of K. Hadjantonakis, Sloan Kettering). The transgene insert was gel purified and injected into the pronuclei of fertilized eggs from FVB/N mice (Transgenic Core Facility, University of Pittsburgh Medical Center). Two founder lines were established, both with similar Cre recombinase activity domains. One line, *Dll1-msd Cre Tg33*, was chosen for further characterization. Mice were genotyped for the presence of the *Dll1-msd Cre* transgene by PCR using primers specific to the Cre recombinase (forward 5'-GGACATGTTCAGGGATCGCCAGGC-3'; reverse 5'- CGACGATGAAGCATGTTTAGCTG 3'), which generates a 219 bp product.

6.11 CHARACTERIZATION OF *DLL1-MSD CRE* TRANSGENIC MICE

To examine the spatial and temporal activity of the *Dll1-msd* Cre recombinase, *Dll1-msd Cre* transgenic mice were mated to *ROSA26-lacZ* reporter mice (*Gt(ROSA)26Sor*^{tm1Sor}*/J*) (Soriano, 1999). Embryos were dissected between e7.0 and e14.5, (embryonic day 0.5 corresponding to the day the vaginal plug was found). Reporter gene activity was assayed by β-galactosidase staining according to standard protocols (Nagy et al., 2003). Embryos were stained for two hours either at room temperature (RT) or 37^{o}C or overnight at RT. Following β-galactosidase staining, some embryos were fixed in Bouin's fixative, paraffin embedded and sectioned. Sections were counterstained with eosin. e13.5 and e14.5 embryos were fixed in 4% paraformaldehyde and frozen in OCT for cryosectioning and subsequent β-galactosidase staining according to standard protocols (Nagy et al., 2003).

6.12 GENERATION *OF DLL1-MSD CRE, ROSA26-RTTA* EMBRYONIC FIBROBLASTS

Fibroblasts were derived from embryos hemizygous for the *Dll1-msd Cre* transgene and heterozygous for *ROSA26-rtTA* (*Gt(ROSA)26Sor*$^{tm1(rtTA,EGFP)Nagy}$*/J*, Jackson Lab*)* (Belteki *et al.*, 2005) according to standard protocols (Abbondanzo *et al.*, 1993). Briefly, e13.5 embryos were dissected free of extraembryonic tissue and the head and liver were removed. In a second set of experiments, neural tissue, heart and lungs were also removed. Isolated tissue was macerated by passing it through an 18-gauge needle and then cultured in DMEM supplemented with 10% fetal

calf serum. Fibroblasts were expanded and at the fourth passage plated onto coverslips for immunocytochemistry. Cells on coverslips were fixed in 4% paraformaldehyde, permeablized in 0.1% Triton X-100 in PBS, and blocked in 5% goat serum in PBS. Coverslips were incubated with rabbit anti-GFP antibody (Torey Pines Biolabs, 1:1000 dilution in blocking buffer), washed and then incubated with goat anti-rabbit 568-secondary antibody (Molecular Probes, 1:400 dilution in blocking buffer) and TO-PRO3 nuclear stain (Molecular Probes, 1:1000 dilution in blocking buffer). Cells were mounted on slides using Vectashield mounting medium and optical sections were visualized on a BioRad scanning laser confocal microscope. Photoshop was used to merge the different channel images. The number of EGFP positive nuclei or Tbx6 positive nuclei was divided by the total number of nuclei in each frame to determine the percentage of EGFP positive cells. Two images derived from two separate coverslips were counted to obtain the average percentage of fibroblasts that were EGFP positive.

6.13 *IN SITU* HYBRIDIZATIONS

Whole-mount in situ hybridization was performed as previously described (Wilkinson, 1992) using antisense riboprobes for *Dll1, Fgf8, mesp2, myogenin, Ripply2, Tbx6, Tbx18, Tbx15* and *uncx4.1*. Hybridizations and washes were performed at 63°C.

6.14 WHOLE-MOUNT IMMUNOCYTOCHEMISTRY

Immunocytochemistry was performed as described in Nagy *et al.* (2003). The Tbx6 N-terminal affinity purified antibody was used at a 1:500 dilution. Goat anti-rabbit:HRP-conjugated secondary antibody (Jackson ImmunoResearch) was used at a 1:500 dilution and staining was performed in the presence of DAB, hydrogen peroxide and nickel chloride.

BIBLIOGRAPHY

Abbondanzo, S. J., Gadi, I. and Stewart, C. L. (1993). Derivation of embryonic stem cell lines. *Methods Enzymol* **225**, 803-23.

Agarwal, P., Wylie, J. N., Galceran, J., Arkhitko, O., Li, C., Deng, C., Grosschedl, R. and Bruneau, B. G. (2003). Tbx5 is essential for forelimb bud initiation following patterning of the limb field in the mouse embryo. *Development* **130**, 623-33.

Agulnik, S. I., Papaioannou, V. E. and Silver, L. M. (1998). Cloning, mapping, and expression analysis of TBX15, a new member of the T-Box gene family. *Genomics* **51**, 68-75.

Andreou, A. M., Pauws, E., Jones, M. C., Singh, M. K., Bussen, M., Doudney, K., Moore, G. E., Kispert, A., Brosens, J. J. and Stanier, P. (2007). TBX22 missense mutations found in patients with X-linked cleft palate affect DNA binding, sumoylation, and transcriptional repression. *Am J Hum Genet* **81**, 700-12.

Artavanis-Tsakonas, S., Rand, M. D. and Lake, R. J. (1999). Notch signaling: cell fate control and signal integration in development. *Science* **284**, 770-6.

Asteria, C. (2002). T-box and isolated ACTH deficiency. *Eur J Endocrinol* **146**, 463-5.

Aulehla, A. and Johnson, R. L. (1999). Dynamic expression of lunatic fringe suggests a link between notch signaling and an autonomous cellular oscillator driving somite segmentation. *Dev Biol* **207**, 49-61.

Aulehla, A., Wehrle, C., Brand-Saberi, B., Kemler, R., Gossler, A., Kanzler, B. and Herrmann, B. G. (2003). Wnt3a plays a major role in the segmentation clock controlling somitogenesis. *Dev Cell* **4**, 395-406.

Baldini, A. (2003). DiGeorge's syndrome: a gene at last. *Lancet* **362**, 1342-3.

Bamshad, M., Lin, R. C., Law, D. J., Watkins, W. C., Krakowiak, P. A., Moore, M. E., Franceschini, P., Lala, R., Holmes, L. B., Gebuhr, T. C. et al. (1997). Mutations in human TBX3 alter limb, apocrine and genital development in ulnar-mammary syndrome. *Nat Genet* **16**, 311-5.

Barlund, M., Monni, O., Kononen, J., Cornelison, R., Torhorst, J., Sauter, G., Kallioniemi, O.-P. and Kallioniemi, A. (2000). Multiple genes at 17q23 undergo amplification and overexpression in breast cancer. *Cancer Res* **60**, 5340-4.

Basson, C. T., Bachinsky, D. R., Lin, R. C., Levi, T., Elkins, J. A., Soults, J., Grayzel, D., Kroumpouzou, E., Traill, T. A., Leblanc-Straceski, J. et al. (1997). Mutations in human TBX5 cause limb and cardiac malformation in Holt-Oram syndrome. *Nat Genet* **15**, 30-5.

Beckers, J., Caron, A., Hrabe de Angelis, M., Hans, S., Campos-Ortega, J. A. and Gossler, A. (2000a). Distinct regulatory elements direct Delta1 expression in the nervous system and paraxial mesoderm of transgenic mice. *Mech Dev* **95**, 23-34.

Beckers, J., Schlautmann, N. and Gossler, A. (2000b). The mouse rib-vertebrae mutation disrupts anterior-posterior somite patterning and genetically interacts with a delta1 null allele. *Mech Dev* **95**, 35-46.

Beddington, R. S., Rashbass, P. and Wilson, V. (1992). Brachyury--a gene affecting mouse gastrulation and early organogenesis. *Dev Suppl*, 157-65.

Belteki, G., Haigh, J., Kabacs, N., Haigh, K., Sison, K., Costantini, F., Whitsett, J., Quaggin, S. E. and Nagy, A. (2005). Conditional and inducible transgene expression in mice through the combinatorial use of Cre-mediated recombination and tetracycline induction. *Nucleic Acids Res* **33**, e51.

Bettenhausen, B., Hrabe de Angelis, M., Simon, D., Guenet, J. L. and Gossler, A. (1995). Transient and restricted expression during mouse embryogenesis of Dll1, a murine gene closely related to Drosophila Delta. *Development* **121**, 2407-18.

Biris, K. K., Dunty, W. C., Jr. and Yamaguchi, T. P. (2007). Mouse Ripply2 is downstream of Wnt3a and is dynamically expressed during somitogenesis. *Dev Dyn* **236**, 3167-72.

Bollag, R. J., Siegfried, Z., Cebra-Thomas, J. A., Garvey, N., Davison, E. M. and Silver, L. M. (1994). An ancient family of embryonically expressed mouse genes sharing a conserved protein motif with the T locus. *Nat Genet* **7**, 383-9.

Bongers, E. M., Duijf, P. H., van Beersum, S. E., Schoots, J., Van Kampen, A., Burckhardt, A., Hamel, B. C., Losan, F., Hoefsloot, L. H., Yntema, H. G. et al. (2004). Mutations in the human TBX4 gene cause small patella syndrome. *Am J Hum Genet* **74**, 1239-48.

Braybrook, C., Doudney, K., Marcano, A. C., Arnason, A., Bjornsson, A., Patton, M. A., Goodfellow, P. J., Moore, G. E. and Stanier, P. (2001). The T-box transcription factor gene TBX22 is mutated in X-linked cleft palate and ankyloglossia. *Nat Genet* **29**, 179-83.

Bronner-Fraser, M. (2000). Rostrocaudal differences within the somites confer segmental pattern to trunk neural crest migration. *Curr Top Dev Biol* **47**, 279-96.

Brown, D. D., Martz, S. N., Binder, O., Goetz, S. C., Price, B. M., Smith, J. C. and Conlon, F. L. (2005). Tbx5 and Tbx20 act synergistically to control vertebrate heart morphogenesis. *Development* **132**, 553-63.

Bulman, M. P., Kusumi, K., Frayling, T. M., McKeown, C., Garrett, C., Lander, E. S., Krumlauf, R., Hattersley, A. T., Ellard, S. and Turnpenny, P. D. (2000). Mutations in the human delta homologue, DLL3, cause axial skeletal defects in spondylocostal dysostosis. *Nat Genet* **24**, 438-41.

Bush, J. O., Lan, Y., Maltby, K. M. and Jiang, R. (2002). Isolation and developmental expression analysis of Tbx22, the mouse homolog of the human X-linked cleft palate gene. *Dev Dyn* **225**, 322-6.

Bussen, M., Petry, M., Schuster-Gossler, K., Leitges, M., Gossler, A. and Kispert, A. (2004). The T-box transcription factor Tbx18 maintains the separation of anterior and posterior somite compartments. *Genes Dev* **18**, 1209-21.

Cai, C. L., Zhou, W., Yang, L., Bu, L., Qyang, Y., Zhang, X., Li, X., Rosenfeld, M. G., Chen, J. and Evans, S. (2005). T-box genes coordinate regional rates of proliferation and regional specification during cardiogenesis. *Development* **132**, 2475-87.

Carlson, H., Ota, S., Campbell, C. E. and Hurlin, P. J. (2001). A dominant repression domain in Tbx3 mediates transcriptional repression and cell immortalization: relevance to mutations in Tbx3 that cause ulnar-mammary syndrome. *Hum Mol Genet* **10**, 2403-13.

Carreira, S., Dexter, T. J., Yavuzer, U., Easty, D. J. and Goding, C. R. (1998). Brachyury-related transcription factor Tbx2 and repression of the melanocyte-specific TRP-1 promoter. *Mol Cell Biol* **18**, 5099-108.

Casey, E. S., O'Reilly, M. A., Conlon, F. L. and Smith, J. C. (1998). The T-box transcription factor Brachyury regulates expression of eFGF through binding to a non-palindromic response element. *Development* **125**, 3887-94.

Chan, T., Kondow, A., Hosoya, A., Hitachi, K., Yukita, A., Okabayashi, K., Nakamura, H., Ozawa, H., Kiyonari, H., Michiue, T. et al. (2007). Ripply2 is essential for precise somite formation during mouse early development. *FEBS Lett* **581**, 2691-6.

Chapman, D. L., Agulnik, I., Hancock, S., Silver, L. M. and Papaioannou, V. E. (1996). Tbx6, a mouse T-Box gene implicated in paraxial mesoderm formation at gastrulation. *Dev Biol* **180**, 534-42.

Chapman, D. L., Cooper-Morgan, A., Harrelson, Z. and Papaioannou, V. E. (2003). Critical role for Tbx6 in mesoderm specification in the mouse embryo. *Mech Dev* **120**, 837-47.

Chapman, D. L. and Papaioannou, V. E. (1998). Three neural tubes in mouse embryos with mutations in the T-box gene Tbx6. *Nature* **391**, 695-7.

Chen, Y. L., Liu, B., Zhou, Z. N., Hu, R. Y., Fei, C., Xie, Z. H. and Ding, X. (2009). Smad6 inhibits the transcriptional activity of Tbx6 by mediating its degradation. *J Biol Chem* **284**, 23481-90.

Chesley, P. (1935). Development of the short-tailed mutant in the house mouse. *J. Exp. Zool.* **70**, 429-435.

Christ, B. and Ordahl, C. P. (1995). Early stages of chick somite development. *Anat Embryol (Berl)* **191**, 381-96.

Ciruna, B. and Rossant, J. (2001). FGF signaling regulates mesoderm cell fate specification and morphogenetic movement at the primitive streak. *Dev Cell* **1**, 37-49.

Ciruna, B. G., Schwartz, L., Harpal, K., Yamaguchi, T. P. and Rossant, J. (1997). Chimeric analysis of fibroblast growth factor receptor-1 (Fgfr1) function: a role for FGFR1 in morphogenetic movement through the primitive streak. *Development* **124**, 2829-41.

Coll, M., Seidman, J. G. and Muller, C. W. (2002). Structure of the DNA-bound T-box domain of human TBX3, a transcription factor responsible for ulnar-mammary syndrome. *Structure* **10**, 343-56.

Conlon, F. L., Fairclough, L., Price, B. M., Casey, E. S. and Smith, J. C. (2001). Determinants of T box protein specificity. *Development* **128**, 3749-58.

Conlon, R. A., Reaume, A. G. and Rossant, J. (1995). Notch1 is required for the coordinate segmentation of somites. *Development* **121**, 1533-45.

Crossley, P. H. and Martin, G. R. (1995). The mouse Fgf8 gene encodes a family of polypeptides and is expressed in regions that direct outgrowth and patterning in the developing embryo. *Development* **121**, 439-51.

Dale, J. K., Maroto, M., Dequeant, M. L., Malapert, P., McGrew, M. and Pourquie, O. (2003). Periodic notch inhibition by lunatic fringe underlies the chick segmentation clock. *Nature* **421**, 275-8.

Davenport, T. G., Jerome-Majewska, L. A. and Papaioannou, V. E. (2003). Mammary gland, limb and yolk sac defects in mice lacking Tbx3, the gene mutated in human ulnar mammary syndrome. *Development* **130**, 2263-73.

Davis, A. C., Wims, M., Spotts, G. D., Hann, S. R. and Bradley, A. (1993). A null c-myc mutation causes lethality before 10.5 days of gestation in homozygotes and reduced fertility in heterozygous female mice. *Genes Dev* **7**, 671-82.

Deng, C. X., Wynshaw-Boris, A., Shen, M. M., Daugherty, C., Ornitz, D. M. and Leder, P. (1994). Murine FGFR-1 is required for early postimplantation growth and axial organization. *Genes Dev* **8**, 3045-57.

Dobrovolskaia-Zavadskaia, N. (1927). Sur la moritifcation spontanee de la queue chez la souris nouveau-nee et sur l'existence d'un caractere hereditaire 'non viable'. *C.R. Hebd. Seanc. Soc. Biol.* **97**, 114-116.

Duband, J. L., Dufour, S., Hatta, K., Takeichi, M., Edelman, G. M. and Thiery, J. P. (1987). Adhesion molecules during somitogenesis in the avian embryo. *J Cell Biol* **104**, 1361-74.

Dubrulle, J., McGrew, M. J. and Pourquie, O. (2001). FGF signaling controls somite boundary position and regulates segmentation clock control of spatiotemporal Hox gene activation. *Cell* **106**, 219-32.

Dunty, W. C., Jr., Biris, K. K., Chalamalasetty, R. B., Taketo, M. M., Lewandoski, M. and Yamaguchi, T. P. (2008). Wnt3a/beta-catenin signaling controls posterior body development by coordinating mesoderm formation and segmentation. *Development* **135**, 85-94.

Ekker, S. C., von Kessler, D. P. and Beachy, P. A. (1992). Differential DNA sequence recognition is a determinant of specificity in homeotic gene action. *Embo J* **11**, 4059-72.

Evrard, Y. A., Lun, Y., Aulehla, A., Gan, L. and Johnson, R. L. (1998). lunatic fringe is an essential mediator of somite segmentation and patterning. *Nature* **394**, 377-81.

Farin, H. F., Bussen, M., Schmidt, M. K., Singh, M. K., Schuster-Gossler, K. and Kispert, A. (2007). Transcriptional repression by the T-box proteins Tbx18 and Tbx15 depends on Groucho corepressors. *J Biol Chem*.

Farin, H. F., Mansouri, A., Petry, M. and Kispert, A. (2008). T-box protein Tbx18 interacts with the paired box protein Pax3 in the development of the paraxial mesoderm. *J Biol Chem* **283**, 25372-80.

Fernando, R. I., Litzinger, M., Trono, P., Hamilton, D. H., Schlom, J. and Palena, C. (2010). The T-box transcription factor Brachyury promotes epithelial-mesenchymal transition in human tumor cells. *J Clin Invest*.

Foucher, I., Montesinos, M. L., Volovitch, M., Prochiantz, A. and Trembleau, A. (2003). Joint regulation of the MAP1B promoter by HNF3beta/Foxa2 and Engrailed is the result of a highly conserved mechanism for direct interaction of homeoproteins and Fox transcription factors. *Development* **130**, 1867-76.

Friedman, J. R. and Kaestner, K. H. (2006). The Foxa family of transcription factors in development and metabolism. *Cell Mol Life Sci* **63**, 2317-28.

Georges, A. B., Benayoun, B. A., Caburet, S. and Veitia, R. A. (2010). Generic binding sites, generic DNA-binding domains: where does specific promoter recognition come from? *Faseb J* **24**, 346-56.

Ghebranious, N., Blank, R. D., Raggio, C. L., Staubli, J., McPherson, E., Ivacic, L., Rasmussen, K., Jacobsen, F. S., Faciszewski, T., Burmester, J. K. et al. (2008). A missense T (Brachyury) mutation contributes to vertebral malformations. *J Bone Miner Res* **23**, 1576-83.

Ghosh, T. K., Packham, E. A., Bonser, A. J., Robinson, T. E., Cross, S. J. and Brook, J. D. (2001). Characterization of the TBX5 binding site and analysis of mutations that cause Holt-Oram syndrome. *Hum Mol Genet* **10**, 1983-94.

Glickman, N. S., Kimmel, C. B., Jones, M. A. and Adams, R. J. (2003). Shaping the zebrafish notochord. *Development* **130**, 873-87.

Goering, L. M., Hoshijima, K., Hug, B., Bisgrove, B., Kispert, A. and Grunwald, D. J. (2003). An interacting network of T-box genes directs gene expression and fate in the zebrafish mesoderm. *Proc Natl Acad Sci U S A* **100**, 9410-5.

Goff, D. J. and Tabin, C. J. (1997). Analysis of Hoxd-13 and Hoxd-11 misexpression in chick limb buds reveals that Hox genes affect both bone condensation and growth. *Development* **124**, 627-36.

Goh, K. L., Yang, J. T. and Hynes, R. O. (1997). Mesodermal defects and cranial neural crest apoptosis in alpha5 integrin-null embryos. *Development* **124**, 4309-19.

Goldin, S. N. and Papaioannou, V. E. (2003). Unusual misregulation of RNA splicing caused by insertion of a transposable element into the T (Brachyury) locus. *BMC Genomics* **4**, 14.

Habets, P. E., Moorman, A. F., Clout, D. E., van Roon, M. A., Lingbeek, M., van Lohuizen, M., Campione, M. and Christoffels, V. M. (2002). Cooperative action of Tbx2 and Nkx2.5 inhibits ANF expression in the atrioventricular canal: implications for cardiac chamber formation. *Genes Dev* **16**, 1234-46.

Harada, R., Dufort, D., Denis-Larose, C. and Nepveu, A. (1994). Conserved cut repeats in the human cut homeodomain protein function as DNA binding domains. *J Biol Chem* **269**, 2062-7.

Harrelson, Z. and Papaioannou, V. E. (2006). Segmental expression of the T-box transcription factor, Tbx2, during early somitogenesis. *Dev Dyn* **235**, 3080-4.

Hashimoto, K., Fujimoto, H. and Nakatsuji, N. (1987). An ECM substratum allows mouse mesodermal cells isolated from the primitive streak to exhibit motility similar to that inside the embryo and reveals a deficiency in the T/T mutant cells. *Development* **100**, 587-98.

Hasson, P., Del Buono, J. and Logan, M. P. (2007). Tbx5 is dispensable for forelimb outgrowth. *Development* **134**, 85-92.

Hatada, Y. and Stern, C. D. (1994). A fate map of the epiblast of the early chick embryo. *Development* **120**, 2879-89.

Herrmann, B. G., Labeit, S., Poustka, A., King, T. R. and Lehrach, H. (1990). Cloning of the T gene required in mesoderm formation in the mouse. *Nature* **343**, 617-22.

Hitachi, K., Danno, H., Kondow, A., Ohnuma, K., Uchiyama, H., Ishiura, S., Kurisaki, A. and Asashima, M. (2008). Physical interaction between Tbx6 and mespb is indispensable for

the activation of bowline expression during Xenopus somitogenesis. *Biochem Biophys Res Commun* **372**, 607-12.

Hofmann, M., Schuster-Gossler, K., Watabe-Rudolph, M., Aulehla, A., Herrmann, B. G. and Gossler, A. (2004). WNT signaling, in synergy with T/TBX6, controls Notch signaling by regulating Dll1 expression in the presomitic mesoderm of mouse embryos. *Genes Dev* **18**, 2712-7.

Hrabe de Angelis, M., McIntyre, J., 2nd and Gossler, A. (1997). Maintenance of somite borders in mice requires the Delta homologue DII1. *Nature* **386**, 717-21.

Hsueh, Y. P., Wang, T. F., Yang, F. C. and Sheng, M. (2000). Nuclear translocation and transcription regulation by the membrane-associated guanylate kinase CASK/LIN-2. *Nature* **404**, 298-302.

Huang, R., Zhi, Q., Brand-Saberi, B. and Christ, B. (2000a). New experimental evidence for somite resegmentation. *Anat Embryol (Berl)* **202**, 195-200.

Huang, R., Zhi, Q., Patel, K., Wilting, J. and Christ, B. (2000b). Dual origin and segmental organisation of the avian scapula. *Development* **127**, 3789-94.

Jacobs, J. J., Keblusek, P., Robanus-Maandag, E., Kristel, P., Lingbeek, M., Nederlof, P. M., van Welsem, T., van de Vijver, M. J., Koh, E. Y., Daley, G. Q. et al. (2000). Senescence bypass screen identifies TBX2, which represses Cdkn2a (p19(ARF)) and is amplified in a subset of human breast cancers. *Nat Genet* **26**, 291-9.

Jouve, C., Palmeirim, I., Henrique, D., Beckers, J., Gossler, A., Ish-Horowicz, D. and Pourquie, O. (2000). Notch signalling is required for cyclic expression of the hairy-like gene HES1 in the presomitic mesoderm. *Development* **127**, 1421-9.

Jukkola, T., Trokovic, R., Maj, P., Lamberg, A., Mankoo, B., Pachnis, V., Savilahti, H. and Partanen, J. (2005). Meox1Cre: a mouse line expressing Cre recombinase in somitic mesoderm. *Genesis* **43**, 148-53.

Kaestner, K. H., Katz, J., Liu, Y., Drucker, D. J. and Schutz, G. (1999). Inactivation of the winged helix transcription factor HNF3alpha affects glucose homeostasis and islet glucagon gene expression in vivo. *Genes Dev* **13**, 495-504.

Kawamura, A., Koshida, S., Hijikata, H., Ohbayashi, A., Kondoh, H. and Takada, S. (2005). Groucho-associated transcriptional repressor ripply1 is required for proper transition from the presomitic mesoderm to somites. *Dev Cell* **9**, 735-44.

Kawamura, A., Koshida, S. and Takada, S. (2008). Activator-to-repressor conversion of T-box transcription factors by the Ripply family of Groucho/TLE-associated mediators. *Mol Cell Biol* **28**, 3236-44.

Kispert, A. and Hermann, B. G. (1993). The Brachyury gene encodes a novel DNA binding protein. *Embo J* **12**, 4898-9.

Kispert, A., Koschorz, B. and Herrmann, B. G. (1995). The T protein encoded by Brachyury is a tissue-specific transcription factor. *Embo J* **14**, 4763-72.

Kissinger, C. R., Liu, B. S., Martin-Blanco, E., Kornberg, T. B. and Pabo, C. O. (1990). Crystal structure of an engrailed homeodomain-DNA complex at 2.8 A resolution: a framework for understanding homeodomain-DNA interactions. *Cell* **63**, 579-90.

Kondow, A., Hitachi, K., Ikegame, T. and Asashima, M. (2006). Bowline, a novel protein localized to the presomitic mesoderm, interacts with Groucho/TLE in Xenopus. *Int J Dev Biol* **50**, 473-9.

Kondow, A., Hitachi, K., Okabayashi, K., Hayashi, N. and Asashima, M. (2007). Bowline mediates association of the transcriptional corepressor XGrg-4 with Tbx6 during somitogenesis in Xenopus. *Biochem Biophys Res Commun* **359**, 959-64.

Korver, W., Roose, J., Wilson, A. and Clevers, H. (1997). The winged-helix transcription factor Trident is expressed in actively dividing lymphocytes. *Immunobiology* **198**, 157-61.

Kraus, F., Haenig, B. and Kispert, A. (2001). Cloning and expression analysis of the mouse T-box gene Tbx18. *Mech Dev* **100**, 83-6.

Kusumi, K., Sun, E. S., Kerrebrock, A. W., Bronson, R. T., Chi, D. C., Bulotsky, M. S., Spencer, J. B., Birren, B. W., Frankel, W. N. and Lander, E. S. (1998). The mouse pudgy mutation disrupts Delta homologue Dll3 and initiation of early somite boundaries. *Nat Genet* **19**, 274-8.

Lausch, E., Hermanns, P., Farin, H. F., Alanay, Y., Unger, S., Nikkel, S., Steinwender, C., Scherer, G., Spranger, J., Zabel, B. et al. (2008). TBX15 mutations cause craniofacial dysmorphism, hypoplasia of scapula and pelvis, and short stature in Cousin syndrome. *Am J Hum Genet* **83**, 649-55.

Leimeister, C., Dale, K., Fischer, A., Klamt, B., Hrabe de Angelis, M., Radtke, F., McGrew, M. J., Pourquie, O. and Gessler, M. (2000). Oscillating expression of c-Hey2 in the presomitic mesoderm suggests that the segmentation clock may use combinatorial signaling through multiple interacting bHLH factors. *Dev Biol* **227**, 91-103.

Levine, M. and Hoey, T. (1988). Homeobox proteins as sequence-specific transcription factors. *Cell* **55**, 537-40.

Littler, D. R., Alvarez-Fernandez, M., Stein, A., Hibbert, R. G., Heidebrecht, T., Aloy, P., Medema, R. H. and Perrakis, A. (2010). Structure of the FoxM1 DNA-recognition domain bound to a promoter sequence. *Nucleic Acids Res.*

Liu, C., Shen, A., Li, X., Jiao, W., Zhang, X. and Li, Z. (2008). T-box transcription factor TBX20 mutations in Chinese patients with congenital heart disease. *Eur J Med Genet* **51**, 580-7.

Macindoe, I., Glockner, L., Vukasin, P., Stennard, F. A., Costa, M. W., Harvey, R. P., Mackay, J. P. and Sunde, M. (2009). Conformational stability and DNA binding specificity of the cardiac T-box transcription factor Tbx20. *J Mol Biol* **389**, 606-18.

Medina-Martinez, O. and Ramirez-Solis, R. (2003). In vivo mutagenesis of the Hoxb8 hexapeptide domain leads to dominant homeotic transformations that mimic the loss-of-function mutations in genes of the Hoxb cluster. *Dev Biol* **264**, 77-90.

Messenger, N. J., Kabitschke, C., Andrews, R., Grimmer, D., Miguel, R. N., Blundell, T. L., Smith, J. C. and Wardle, F. C. (2005). Functional specificity of the Xenopus T-domain protein brachyury is conferred by its ability to interact with smad1. *Dev Cell* **8**, 599-610.

Moreno, T. A. and Kintner, C. (2004). Regulation of segmental patterning by retinoic acid signaling during Xenopus somitogenesis. *Dev Cell* **6**, 205-18.

Morimoto, M., Sasaki, N., Oginuma, M., Kiso, M., Igarashi, K., Aizaki, K., Kanno, J. and Saga, Y. (2007). The negative regulation of Mesp2 by mouse Ripply2 is required to establish the rostro-caudal patterning within a somite. *Development* **134**, 1561-9.

Morimoto, M., Takahashi, Y., Endo, M. and Saga, Y. (2005). The Mesp2 transcription factor establishes segmental borders by suppressing Notch activity. *Nature* **435**, 354-9.

Morley, R. H., Lachani, K., Keefe, D., Gilchrist, M. J., Flicek, P., Smith, J. C. and Wardle, F. C. (2009). A gene regulatory network directed by zebrafish No tail accounts for its roles in mesoderm formation. *Proc Natl Acad Sci U S A* **106**, 3829-34.

Motallebipour, M., Ameur, A., Reddy Bysani, M. S., Patra, K., Wallerman, O., Mangion, J., Barker, M. A., McKernan, K. J., Komorowski, J. and Wadelius, C. (2009). Differential binding and co-binding pattern of FOXA1 and FOXA3 and their relation to H3K4me3 in HepG2 cells revealed by ChIP-seq. *Genome Biol* **10**, R129.

Muller, C. W. and Herrmann, B. G. (1997). Crystallographic structure of the T domain-DNA complex of the Brachyury transcription factor. *Nature* **389**, 884-8.

Nagy, A., Gertsenstein, M., Vintersten, K. and Behringer, R. R. (2003). Manipulating the Mouse Embryo. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

Naiche, L. A., Harrelson, Z., Kelly, R. G. and Papaioannou, V. E. (2005a). T-Box Genes in Vertebrate Development. *Annu Rev Genet* **39**, 219-239.

Naiche, L. A., Harrelson, Z., Kelly, R. G. and Papaioannou, V. E. (2005b). T-box genes in vertebrate development. *Annu Rev Genet* **39**, 219-39.

Naiche, L. A. and Papaioannou, V. E. (2003). Loss of Tbx4 blocks hindlimb development and affects vascularization and fusion of the allantois. *Development* **130**, 2681-93.

Oginuma, M., Niwa, Y., Chapman, D. L. and Saga, Y. (2008). Mesp2 and Tbx6 cooperatively create periodic patterns coupled with the clock machinery during mouse somitogenesis. *Development* **135**, 2555-62.

Oka, C., Nakano, T., Wakeham, A., de la Pompa, J. L., Mori, C., Sakai, T., Okazaki, S., Kawaichi, M., Shiota, K., Mak, T. W. et al. (1995). Disruption of the mouse RBP-J kappa gene results in early embryonic death. *Development* **121**, 3291-301.

Packard, D. S., Jr. (1978). Chick somite determination: the role of factors in young somites and the segmental plate. *J Exp Zool* **203**, 295-306.

Palmeirim, I., Henrique, D., Ish-Horowicz, D. and Pourquie, O. (1997). Avian hairy gene expression identifies a molecular clock linked to vertebrate segmentation and somitogenesis. *Cell* **91**, 639-48.

Pellerin, I., Schnabel, C., Catron, K. M. and Abate, C. (1994). Hox proteins have different affinities for a consensus DNA site that correlate with the positions of their genes on the hox cluster. *Mol Cell Biol* **14**, 4532-45.

Perantoni, A. O., Timofeeva, O., Naillat, F., Richman, C., Pajni-Underwood, S., Wilson, C., Vainio, S., Dove, L. F. and Lewandoski, M. (2005). Inactivation of FGF8 in early mesoderm reveals an essential role in kidney development. *Development* **132**, 3859-71.

Piper, D. E., Batchelor, A. H., Chang, C. P., Cleary, M. L. and Wolberger, C. (1999). Structure of a HoxB1-Pbx1 heterodimer bound to DNA: role of the hexapeptide and a fourth homeodomain helix in complex formation. *Cell* **96**, 587-97.

Pourquie, O. (2001). Vertebrate somitogenesis. *Annu Rev Cell Dev Biol* **17**, 311-50.

Qian, Y. Q., Billeter, M., Otting, G., Muller, M., Gehring, W. J. and Wuthrich, K. (1989). The structure of the Antennapedia homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors. *Cell* **59**, 573-80.

Rallis, C., Bruneau, B. G., Del Buono, J., Seidman, C. E., Seidman, J. G., Nissim, S., Tabin, C. J. and Logan, M. P. (2003). Tbx5 is required for forelimb bud formation and continued outgrowth. *Development* **130**, 2741-51.

Russ, A. P., Wattler, S., Colledge, W. H., Aparicio, S. A., Carlton, M. B., Pearce, J. J., Barton, S. C., Surani, M. A., Ryan, K., Nehls, M. C. et al. (2000). Eomesodermin is required for mouse trophoblast development and mesoderm formation. *Nature* **404**, 95-9.

Saga, Y., Hata, N., Koseki, H. and Taketo, M. M. (1997). Mesp2: a novel mouse gene expressed in the presegmented mesoderm and essential for segmentation initiation. *Genes Dev* **11**, 1827-39.

Scott, M. P. (1992). Vertebrate homeobox gene nomenclature. *Cell* **71**, 551-3.

Selkoe, D. and Kopan, R. (2003). Notch and Presenilin: regulated intramembrane proteolysis links development and degeneration. *Annu Rev Neurosci* **26**, 565-97.

Sewell, W. and Kusumi, K. (2007). Genetic analysis of molecular oscillators in mammalian somitogenesis: clues for studies of human vertebral disorders. *Birth Defects Res C Embryo Today* **81**, 111-20.

Shedlovsky, A., King, T. R. and Dove, W. F. (1988). Saturation germ line mutagenesis of the murine t region including a lethal allele at the quaking locus. *Proc Natl Acad Sci U S A* **85**, 180-4.

Shimojima, K., Inoue, T., Fujii, Y., Ohno, K. and Yamamoto, T. (2009). A familial 593-kb microdeletion of 16p11.2 associated with mental retardation and hemivertebrae. *Eur J Med Genet* **52**, 433-5.

Singh, M. K., Petry, M., Haenig, B., Lescher, B., Leitges, M. and Kispert, A. (2005). The T-box transcription factor Tbx15 is required for skeletal development. *Mech Dev* **122**, 131-44.

Smith, L. J. (1964). The Effects of Transection and Extirpation on Axis Formation and Elongation in the Young Mouse Embryo. *J Embryol Exp Morphol* **12**, 787-803.

Soriano, P. (1999). Generalized lacZ expression with the ROSA26 Cre reporter strain. *Nat Genet* **21**, 70-1.

Sparrow, D. B., Chapman, G., Wouters, M. A., Whittock, N. V., Ellard, S., Fatkin, D., Turnpenny, P. D., Kusumi, K., Sillence, D. and Dunwoodie, S. L. (2006). Mutation of the LUNATIC FRINGE gene in humans causes spondylocostal dysostosis with a severe vertebral phenotype. *Am J Hum Genet* **78**, 28-37.

Stott, D., Kispert, A. and Herrmann, B. G. (1993). Rescue of the tail defect of Brachyury mice. *Genes Dev* **7**, 197-203.

Sun, X., Meyers, E. N., Lewandoski, M. and Martin, G. R. (1999). Targeted disruption of Fgf8 causes failure of cell migration in the gastrulating mouse embryo. *Genes Dev* **13**, 1834-46.

Suzuki, T., Takeuchi, J., Koshiba-Takeuchi, K. and Ogura, T. (2004). Tbx Genes Specify Posterior Digit Identity through Shh and BMP Signaling. *Dev Cell* **6**, 43-53.

Svingen, T. and Tonissen, K. F. (2006). Hox transcription factors and their elusive mammalian gene targets. *Heredity* **97**, 88-96.

Tada, M., Casey, E. S., Fairclough, L. and Smith, J. C. (1998). Bix1, a direct target of Xenopus T-box genes, causes formation of ventral mesoderm and endoderm. *Development* **125**, 3997-4006.

Tada, M. and Smith, J. C. (2000). Xwnt11 is a target of Xenopus Brachyury: regulation of gastrulation movements via Dishevelled, but not through the canonical Wnt pathway. *Development* **127**, 2227-38.

Tada, M. and Smith, J. C. (2001). T-targets: clues to understanding the functions of T-box proteins. *Dev Growth Differ* **43**, 1-11.

Takada, S., Stark, K. L., Shea, M. J., Vassileva, G., McMahon, J. A. and McMahon, A. P. (1994). Wnt-3a regulates somite and tailbud formation in the mouse embryo. *Genes Dev* **8**, 174-89.

Takeuchi, J. K., Koshiba-Takeuchi, K., Suzuki, T., Kamimura, M., Ogura, K. and Ogura, T. (2003). Tbx5 and Tbx4 trigger limb initiation through activation of the Wnt/Fgf signaling cascade. *Development* **130**, 2729-39.

Tam, P. P. (1984). The histogenetic capacity of tissues in the caudal end of the embryonic axis of the mouse. *J Embryol Exp Morphol* **82**, 253-66.

Tam, P. P. and Beddington, R. S. (1987). The formation of mesodermal tissues in the mouse embryo during gastrulation and early organogenesis. *Development* **99**, 109-26.

Tam, P. P. and Loebel, D. A. (2007). Gene function in mouse embryogenesis: get set for gastrulation. *Nat Rev Genet* **8**, 368-81.

Tam, P. P., Loebel, D. A. and Tanaka, S. S. (2006). Building the mouse gastrula: signals, asymmetry and lineages. *Curr Opin Genet Dev* **16**, 419-25.

Tam, P. P. and Tan, S. S. (1992). The somitogenetic potential of cells in the primitive streak and the tail bud of the organogenesis-stage mouse embryo. *Development* **115**, 703-15.

Theiler, K. and Varnum, D. S. (1985). Development of rib-vertebrae: a new mutation in the house mouse with accessory caudal duplications. *Anat Embryol (Berl)* **173**, 111-6.

Thomas, P. and Beddington, R. (1996). Anterior primitive endoderm may be responsible for patterning the anterior neural plate in the mouse embryo. *Curr Biol* **6**, 1487-96.

Trindade, M., Tada, M. and Smith, J. C. (1999). DNA-binding specificity and embryological function of Xom (Xvent-2). *Dev Biol* **216**, 442-56.

Vermot, J. and Pourquie, O. (2005). Retinoic acid coordinates somitogenesis and left-right patterning in vertebrate embryos. *Nature* **435**, 215-20.

Vujovic, S., Henderson, S., Presneau, N., Odell, E., Jacques, T. S., Tirabosco, R., Boshoff, C. and Flanagan, A. M. (2006). Brachyury, a crucial regulator of notochordal development, is a novel biomarker for chordomas. *J Pathol* **209**, 157-65.

Wadman, I. A., Osada, H., Grutz, G. G., Agulnick, A. D., Westphal, H., Forster, A. and Rabbitts, T. H. (1997). The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. *Embo J* **16**, 3145-57.

Wehn, A. K., Gallo, P. H. and Chapman, D. L. (2009). Generation of transgenic mice expressing Cre recombinase under the control of the Dll1 mesoderm enhancer element. *Genesis*.

White, P. H. and Chapman, D. L. (2005). Dll1 is a downstream target of Tbx6 in the paraxial mesoderm. *Genesis* **42**, 193-202.

White, P. H., Farkas, D. R. and Chapman, D. L. (2005). Regulation of Tbx6 expression by Notch signaling. *Genesis* **42**, 61-70.

White, P. H., Farkas, D. R., McFadden, E. E. and Chapman, D. L. (2003). Defective somite patterning in mouse embryos with reduced levels of Tbx6. *Development* **130**, 1681-90.

Wilkinson, D. G. (1992). Whole mount in situ hybridization of vertebrate embryos. In *In situ hybridization: A practical approach*, (ed. D. G. Wilkinson), pp. 75-83. Oxford: IRL Press.

Wilkinson, D. G., Bhatt, S. and Herrmann, B. G. (1990). Expression pattern of the mouse T gene and its role in mesoderm formation. *Nature* **343**, 657-9.

Wilson, V., Manson, L., Skarnes, W. C. and Beddington, R. S. (1995). The T gene is necessary for normal mesodermal morphogenetic cell movements during gastrulation. *Development* **121**, 877-86.

Wittler, L., Shin, E. H., Grote, P., Kispert, A., Beckers, A., Gossler, A., Werber, M. and Herrmann, B. G. (2007). Expression of Msgn1 in the presomitic mesoderm is controlled by synergism of WNT signalling and Tbx6. *EMBO Rep* **8**, 784-9.

Yamaguchi, T. P., Harpal, K., Henkemeyer, M. and Rossant, J. (1994). fgfr-1 is required for embryonic growth and mesodermal patterning during mouse gastrulation. *Genes Dev* **8**, 3032-44.

Yamaguchi, T. P., Takada, S., Yoshikawa, Y., Wu, N. and McMahon, A. P. (1999). T (Brachyury) is a direct target of Wnt3a during paraxial mesoderm specification. *Genes Dev* **13**, 3185-90.

Yang, X. R., Ng, D., Alcorta, D. A., Liebsch, N. J., Sheridan, E., Li, S., Goldstein, A. M., Parry, D. M. and Kelley, M. J. (2009). T (brachyury) gene duplication confers major susceptibility to familial chordoma. *Nat Genet* **41**, 1176-8.

Yasuhiko, Y., Haraguchi, S., Kitajima, S., Takahashi, Y., Kanno, J. and Saga, Y. (2006). Tbx6-mediated Notch signaling controls somite-specific Mesp2 expression. *Proc Natl Acad Sci U S A* **103**, 3651-6.

Yasuhiko, Y., Kitajima, S., Takahashi, Y., Oginuma, M., Kagiwada, H., Kanno, J. and Saga, Y. (2008). Functional importance of evolutionally conserved Tbx6 binding sites in the presomitic mesoderm-specific enhancer of Mesp2. *Development* **135**, 3511-9.

Yoon, J. K. and Wold, B. (2000). The bHLH regulator pMesogenin1 is required for maturation and segmentation of paraxial mesoderm. *Genes Dev* **14**, 3204-14.

Zhang, L., Rubins, N. E., Ahima, R. S., Greenbaum, L. E. and Kaestner, K. H. (2005). Foxa2 integrates the transcriptional response of the hepatocyte to fasting. *Cell Metab* **2**, 141-8.

Zhang, N. and Gridley, T. (1998). Defects in somite formation in lunatic fringe-deficient mice. *Nature* **394**, 374-7.

Zohn, I. E., Li, Y., Skolnik, E. Y., Anderson, K. V., Han, J. and Niswander, L. (2006). p38 and a p38-interacting protein are critical for downregulation of E-cadherin during mouse gastrulation. *Cell* **125**, 957-69.