Molecular studies to understand the role of Tbx6 in somitogenesis

by

Phillip H. White

BS, Animal Bioscience, Pennsylvania State University, 2001

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

2006

UNIVERSITY OF PITTSBURGH

COLLEGE OF ARTS AND SCIENCES

This thesis was presented

by

Phillip H White

It was defended on

October 25, 2006

and approved by

Dr. Jeffrey L. Brodsky, Biological Sciences

Dr. Jeffrey D. Hildebrand, Biological Sciences

Dr. Lewis A. Jacobson, Biological Sciences

Dr. Cynthia Lance-Jones, Neurobiology

Thesis Director: Dr. Deborah L. Chapman, Biological Sciences

Molecular studies to understand the role of Tbx6 in somitogenesis

Phillip H. White, PhD

University of Pittsburgh, 2006

Tbx6 was shown to be a T-box transcription factor expressed in the primitive streak and presomitic mesoderm of the developing mouse embryo; null mutations in Tbx6 resulted in ectopic neural tube formation in place of posterior somites and embryonic lethality. However nothing was known about expression, transcriptional activity, or downstream targets of Tbx6 protein. Using antibodies generated against Tbx6, Tbx6 was confirmed to be a 58 kDa protein *in vivo*, and Tbx6 protein and mRNA have similar spatial and temporal expression patterns. T-box proteins were defined by the presence of the T-domain, the DNA binding region. Given the conservation of the T-domain, it was not surprising that many of the characterized T-box binding sites were also somewhat conserved. Therefore, I first demonstrated that Tbx6 recognized the T palindromic consensus sequence and half-site sequence *in vitro* and then determined that the Tbx6 consensus binding site was 5'- AGGTGTBRNNN -3' using a PCR-based binding site selection assay. Using luciferase reporter assays in cell culture, Tbx6 was determined to be a weak transcriptional activator. Lastly, initial studies with T and Tbx6 suggested that these T-box proteins do not synergistically activate a luciferase reporter.

In other work, the spontaneous mutation *rib-vertebrae* was determined to be a *Tbx6* hypomorphic mutation due to an insertion of 185 bp into a *Tbx6* enhancer. Finally, I also sought to understand how the Notch signaling pathway and *Tbx6* interacted. The Notch transcription factor, RBP-Jκ, was demonstrated to control the maintenance of Tbx6 in the PSM. Furthermore, Tbx6 bound to several putative binding sites *in vitro* within a previously known mesodermal *Dll1* enhancer. These results suggested that Notch signaling functions upstream and downstream of Tbx6. The work described in this thesis used a variety of *in vivo* and *in vitro* approaches, molecular and biochemical, to gain further understanding into how Tbx6, a member of a large family of transcription factors, functions in the context of a developing embryo in the presence of other family members to regulate the expression of target genes.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

I would like to thank my thesis advisor, Dr. Deborah Chapman, for taking me on as a graduate student; I am grateful for having the opportunity to work with such a dedicated scientist, who has given me the freedom to succeed on my own and took many risks in doing so. I thank her for giving me a challenging and rewarding thesis project. Secondly, I thank the past and present members of the Chapman Lab for help throughout my time here. I specifically would like to thank Deborah Farkas for her role in the microinjection and dissection/staining of the Su[H] site directed mutant when she was the mouse technician of the lab. Also, I would like to thank Amy Wehn for suggestions and advice with tissue culture procedures.

I thank my thesis committee members, Dr. Jeffrey Brodsky, Dr. Jeffrey Hildebrand, Dr. Lew Jacobson and Dr. Cynthia Lance-Jones for all their help and guidance. In addition, I would like to thank various members of the Department of Biological Sciences for all their help and support. Many thanks to the Schwacha, Gilbert, and Grabowski labs for help with the use of data collection equipment and to Dr. Vern Twombly, for allowing me to use/borrow a microplate reader. I would like to thank Dr. Jeffrey Brodsky for the use of his lab/luminometer and Dr. Jeffrey Hildebrand for the use of reagents. I would also like to thank Dr. Lisa Engler and Dr. Paul Sapienza for helpful discussions with experimental design and Dr. Gerard Campbell for reagents and supplies. In the initial stages of my cell culture work, I received a great amount of suggestions and advice from Dr. Ping An, Mick Yoder, and Megan Dietz, and I am grateful for all of it. Thank you to my teaching mentors Ms. Carole Lafave and Dr. Paula Grabowski for offering to act in this manner. Lastly, I would like to thank Ms. Cathy Barr and Ms. Pat Dean for all of their help throughout my years in the department.

I would also like to thank Dr. Lucy Xu and Dr. Richard Carthew (Northwestern University) for the GST-Su[H] plasmid and Dr. Frank Conlon for the Xbra protein expression construct.

Most of all, I would like to thank my friends and family for always being there to support and believe in me. To my mother, you have always been the kind and self-sacrificing person that I respect, admire, and love. Finally, to my fiancé Melissa Moser, fate allowed us to meet here, and I could not be done now if it were not for your support and advice. Above all else, I thank you for bringing happiness and love into my life, happiness and love we will share forever.

1.0 INTRODUCTION

Development of various organ systems that function in a coordinated manner requires vast levels of control to ensure that all steps from a single fertilized egg to the birth of a multicellular organism occur properly. Many genes are so crucial that loss of their function results in failure of fetal development, leading to spontaneous abortion of the fetus (miscarriage in humans). At the same time, mutations in other genes permit the embryo to develop but lead to congenital defects; sometimes these defects require surgery or other medical interventions, while others have no known medical treatments. To understand these errors in development, one must first understand the steps that lead to proper development of the organism. It is only after the correct pathways are known that we can find means of intervention when they fail. Our lab is interested in the development of one type of tissue, the mesoderm. We use the mouse as a model system to understand how the mesoderm ultimately becomes subdivided into different populations and how these events are coordinated during development.

1.1 DEVELOPMENT OF EUKARYOTES IS REGULATED BY GENE EXPRESSION

Development of a multicellular organism, like that of a mouse or human, begins with a fertilized egg. As development proceeds, the egg undergoes cell divisions and ultimately cells in this developing embryo become different from each other so that they can make the multitude of cell types that compose the body such as blood, muscle, bone, nerves, etc. How do cells acquire different identities and do so at the proper time and place in the embryo? Cell differentiation is driven by cell-cell interactions, which ultimately leads to the expression of a specific set of genes in one cell type that is different than those expressed in another. Tissue specific transcription factors ultimately regulate cell differentiation.

1.2 THE FORMATION OF MESODERM DURING MOUSE DEVELOPMENT

After fertilization, the embryo undergoes cleavage divisions to form a ball of cells called the blastocyst. In the mouse, this occurs at approximately 3.5 days post-coitus (e3.5) (0.5 day post-coitus (e0.5) is designated as noon on the day the vaginal plug is found). A portion of the cells in the blastocyst, the inner cell mass, gives rise to the embryo proper while the remaining cells, the trophectoderm, gives rise to the extra-embryonic mesoderm (Kaufman, 1992; Wolpert, 1998). Shortly thereafter, at approximately e4.5, the embryo implants into the uterine wall, after which the different embryonic and extra-embryonic cells become more apparent as the embryo begins to take on a cylindrical shape. At e7.0, the primitive streak (PS) appears and marks the beginning of gastrulation, the generation of the three germ layers: the ectoderm, mesoderm, and endoderm. The ectoderm gives rise to the neural tissues and epidermis, the mesoderm gives rise to the skeletal and muscular systems, connective tissue, and other internal organs, and the endoderm gives rise to the gut and its derivatives. The PS is the source of mesoderm, and the mesoderm becomes subdivided into axial, paraxial, intermediate, lateral plate, and extraembryonic mesoderm. These subdivisions are determined by the position and timing of ingression of cells through the PS. Cells destined to form the paraxial mesoderm (PAM) are derived from more distal regions of the PS (Kinder et al., 1999). The PAM will give rise to the somites, or paired blocks of mesoderm, which will undergo regional differentiation to form the skeletal muscle, vertebrae, ribs, and dermis. Somite development is the focus of this work.

After emerging from the PS, PAM migrates to a position between the axial and intermediate mesoderm, thus forming the presomitic mesoderm (PSM) or segmental plate. The PSM continues to form from cells ingressing through the PS until e10.5 (with approximately 35-39 pairs of somites), thus extending the body axis and tail by the addition of precursors to somites. After closure of the posterior neuropore at e10.5, the PS is replaced by the tailbud as the source of PAM. The tailbud continues to replace progenitor cells to produce somites until e13.5 when all 60 somites have formed (Tam and Tan, 1992). Below, I will describe the progression of the somite precursors through somitogenesis into their later stages of mesoderm development.

1.2.1 Differentiation of the somites

PAM cells migrate on both sides of the notochord forming a bilateral mesenchymal tissue between the axial (notochord) and intermediate mesoderm; thus PAM cells acquire the PSM fate (reviewed in Saga and Takeda, 2001). This state is maintained until groups of cells condense at the anterior end of the bilateral PSM forming somitomeres immediately before initial segmentation begins during somite formation. This initial segmentation will then generate the paired somites at regular intervals (90 – 120 minutes in the mouse) via a mesenchymal-to-epithelial transformation. These epithelial somites are spherical structures consisting of epithelial cells that are formed after mesodermal cells begin to become segmented (Saga and Takeda, 2001). Because formation and further differentiation of somites is an ongoing process, at any given time, somites at different stages of development are present along the axis with more highly differentiated somites found in more anterior regions and more immature somites found in more posterior regions (see Figure 1-1).

The somites will undergo regional differentiation to generate the skeletal muscle, dermis, vertebrae and ribs of the adult (reviewed in Saga and Takeda, 2001). Briefly, as somites are formed, they immediately begin to become influenced by nearby cells; ventral somitic cells undergo an epithelial-to-mesenchymal transition to form the sclerotome, while more dorsal cells initially remain epithelial and form the dermamyotome (see Figure 1-1). The sclerotome compartments of the five anterior-most somites contribute to the occipital bone while the sclerotomes of all remaining somites produce the vertebral bodies, vertebral arches, and ribs (Christ and Ordahl, 1995; Couly et al., 1992; Theiler, 1988). Cells from the dermamyotome give rise to the myotome, which underlies the dermatome and gives rise to the epaxial and hypaxial of the trunk and skeletal muscle of the limbs, as well as the dermatome, which contributes to the dermis of the dorsal skin. Cells from all regions of the somite contribute to the vascular epithelium. At all of these stages, cells are becoming further specified, thus locking in specific identities. These changes occur in response to changes in gene expression and are controlled by tissue-specific transcription factors.

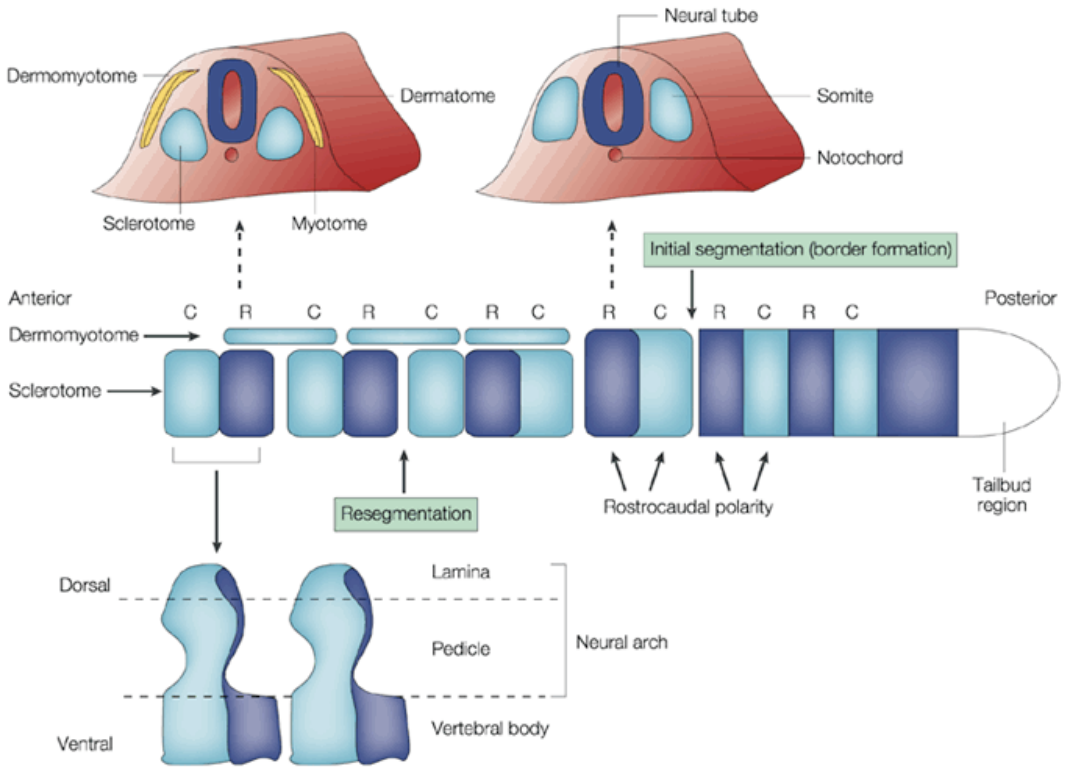


Figure 1-1: Segmentation and resegmentation of somites in the mouse

The above cartoon depicts the many steps in somitogenesis. As described in the text, a group of PSM cells undergoes segmentation and thus form borders that separate the individual somites (initial segmentation). This process continues as long as PSM cells exist. Each somite can be subdivided into anterior (rostral, R) and posterior (caudal, C) halves. Once formed, somitic cells begin to differentiate in response to signals from surrounding tissues. The sclerotome region then undergoes a process called resegmentation whereby the caudal half of the more anterior somite regroups with the rostral half of the next posterior somite to form the vertebral unit. Figure from Saga and Takeda (2001).

1.3 THE T-BOX FAMILY OF TRANSCRIPTION FACTORS PERFORM VITAL ROLES DURING THE DEVELOPMENT OF MULTICELLULAR ORGANISMS

Transcription factors are proteins expressed within cells that perform tasks in many ways, most commonly by activating or repressing gene expression. During development, many transcription factors function to adjust gene expression levels, thus providing the specificity necessary to enable cells to assume different fates (Wolpert, 1998). It is currently thought that transcription factors and a signaling pathway together lead to regulated responses within the cell; these responses change the transcriptional landscape, thereby enabling the cell to function. There are many different families of transcription factors known to function during developmental processes. My thesis focuses on transcription factors belonging to the T-box family, which are critical for assigning cell fates in a variety of organisms (reviewed in Naiche et al., 2005).

The T-box family of transcription factors are essential for normal development in organisms as widely diverse as nematodes and humans. The T-box domain in all family members encodes the DNA binding domain and is the only region of homology shared among all family members. Family members are expressed in a variety of tissues, and functional studies in mice, zebrafish, *Xenopus* and *C. elegans* have demonstrated the important roles played by these genes during development, including specification of germ layers and cell types (Papaioannou, 2001; Showell et al., 2004). Expression of T-box family members have been observed in all developmental stages from the oocyte to adult (reviewed in Naiche et al., 2005). Mutations in T-box genes led to a variety of observed defects during development; many of the genes showed haploinsufficiency, or some form of mutant phenotype when present as a heterozygous mutation, indicating that dosage of these transcription factors was important for the function of T, Tbx1, Tbx3, and Tbx5 (reviewed in Naiche et al., 2005). The T-box genes were named after their founding member, *T* or *Brachyury*, which was originally identified as a mouse mutant that displayed a short tail, thus *Brachyury*, Greek for short tail (Dobrovolskaia-Zavadskaia, 1927). In the following sections, I will discuss the function of both *T* and *Tbx6* (including their orthologs) during development. Both of these T-box genes are crucial for posterior mesoderm formation and are the crux of experiments described in this thesis.

1.4 *BRACHYURY*, THE FOUNDING T-BOX FAMILY MEMBER, REGULATES MESODERM FORMATION AND MIGRATION IN THE EMBRYO

1.4.1 T is expressed in the PS and notochord during development

As previously described, gastrulation is the process that generates the three germ layers; during gastrulation, these three germ layers are organized in the embryo with the mesoderm sandwiched between the ectoderm and endoderm. It is at the PS that cells lose their epithelial characteristic, becoming mesenchymal and adopting a mesodermal fate as they ingress through the PS. Prior to the initiation of gastrulation, *T* is expressed as a ring of epiblast cells at the boundary between the embryonic and extraembryonic portions of the embryo (Beddington et al., 1992). By e7.0, T transcripts are localized to the PS where expression continues throughout gastrulation; *T* is also expressed in the axial mesoderm, the notochord (Wilkinson et al., 1990). Once this newly formed mesoderm moves to the paraxial position, *T* is down-regulated (Herrmann, 1991; Wilkinson et al., 1990). Therefore, T has two distinct expression domains, one in the notochord (axial mesoderm) and one in the PS (nascent mesoderm) (Wilkinson et al., 1990).

1.4.2 Mutant Analysis of *T*

The original *T* mutation was a spontaneous mutation deleting 160 kb region of mouse chromosome 17 (Herrmann et al., 1990). *T* homozygous null embryos failed to form posterior mesoderm and died at approximately e10.5 because the allantois did not make a connection to the chorion, thus failing to form the chorioallantoic placenta (Chesley, 1935; Dobrovolskaia-Zavadskaia, 1927; Gluecksohn-Schoenheimer, 1944). *T* appeared to be a dosage dependent gene because heterozygotes had short tails, while homozygotes had more drastic posterior truncations resulting in embryonic lethality (Chesley, 1935; Dobrovolskaia-Zavadskaia, 1927; Johnson, 1974). By e8.5, homozygous mutant embryos showed that the notochord failed to form, the neural tube was wavy, the axis was truncated, and while 7-9 anterior somites formed, the somites posterior to the forelimb bud were not formed (Chesley, 1935; Fujimoto and Yanagisawa, 1983). The *T* mutation was detectable in histological sections by low mesoderm:endoderm ratios starting at e7.0 and an underdeveloped allantois beginning at e7.5 (Yanagisawa et al., 1981).

In addition to the 160 kb deletion mutant, other mutations in T have been described that result from gene deletion, representing null mutations or altered reading frames (reviewed in Herrmann and Kispert, 1994; reviewed in Lyon et al., 1995). The T^{Wis}, T^{c}, and $T^{c\text{-}2H}$ mutations have been identified as mutations that alter the T protein reading frame, resulting in truncated T proteins (Cattanach and Rasberry, 1987; Herrmann et al., 1990; Searle, 1966; Shedlovsky et al., 1988). The T^{Wis} allele resulted from an insertion of a transposable element into the splice donor site of exon 7 within the *T* gene; the mutation interfered with RNA splicing that resulted in at least eight different RNA transcripts and encoded various truncated T proteins (Goldin and Papaioannou, 2003; Herrmann et al., 1990). Both the T^{C} and $T^{C\text{-}2H}$ alleles occurred during radiation-induced mutagenesis, resulting in either a 19 base-pair (bp) deletion (T^{C}) or a 'CG' to 'A' mutation ($T^{C\text{-}2H}$) in exon 8 that were predicted to frame-shift the resultant proteins, creating C-terminally truncated protein products (Herrmann and Kispert, 1994). All of the mutations described above led to a more severe heterozygous phenotype than null alleles depending on genetic background, and demonstrated by the loss of the tail and skeletal abnormalities (Conlon et al., 1995; Kispert and Herrmann, 1994). In addition, the homozygous T^{Wis} mutant embryos were more severe than the homozygous *T* null mutant embryos, with no discernable anterior somites (Herrmann, 1991). Therefore, it was thought that by producing a truncated protein, the T^{Wis} mutation was a dominant negative mutation with the truncated protein interfering with a T-related protein (Herrmann, 1991).

T appeared to have a role in cell adhesion and the epithelial-to-mesechymal transition in the PS (Wilson and Beddington, 1997; Wilson et al., 1995; Wilson et al., 1993b; Yanagisawa and Fujimoto, 1977; Yanagisawa et al., 1981). Since the proliferation rate of *T*/*T* cells was unchanged, it was thought that cells which express mutant forms of *T* were unable to migrate away from the PS *in vivo*, thereby accumulating in the PS. Support for cell migration defects came from cell culture experiments in which mesodermal cells isolated from *T* mutant embryos were unable to migrate on an ECM and from chimera studies in which *T* null cells accumulated in the PS (Hashimoto et al., 1987; Rashbass et al., 1991; Wilson and Beddington, 1997; Wilson et al., 1995; Wilson et al., 1993b; Yanagisawa et al., 1981). In addition, chimera phenotypic analyses have shown that T acted cell autonomously, thus affecting only *T* expressing cells (Wilson et al., 1993b).

1.4.3 The expression and function of T is evolutionarily conserved

T orthologs have been identified in many organisms allowing T function to be more closely examined (Holland et al., 1995; Kispert et al., 1995b; Schulte-Merker et al., 1992; Smith et al., 1991). Many studies have been conducted using the *Xenopus* ortholog of *T*, *Xbra*, which exhibited a high degree of similarity in sequence similarity, gene expression, and mutant phenotype (Smith et al., 1991). *Xbra* was shown to be both necessary and sufficient for mesoderm formation since misexpression of *T* could divert prospective ectoderm to take on a mesodermal fate (Cunliffe and Smith, 1992). This fate change was found to be dependent upon the concentration of T, as different doses of Xbra were found to be capable of inducing different mesoderm cell types (O'Reilly et al., 1995). Studies in zebrafish have also shown that the role of T in morphogenesis and cell fate was evolutionarily conserved. The zebrafish mutant *no tail* (*ntl*) was found to have homozygous phenotypes that mirrored that of the mouse (Schulte-Merker et al., 1994). Interestingly, however, *ntl* heterozygotes appeared physically normal (Halpern et al., 1993; Schulte-Merker et al., 1994). Cell mobility experiments in zebrafish further supported the notion that T was involved in cell adhesion and cell movements. While early cell movements during gastrulation occurred normally in *ntl* mutants, later cell movements were defective during convergence, when cells exhort traction on neighboring cells thereby narrowing a tissue field (Glickman et al., 2003). The authors concluded that loss of *ntl* demonstrated that convergence could be controlled separately from extension of a body axis during convergent extension because the field of cells expanded the axis without narrowing the field showing the requirement for the cell adhesion property of *ntl* during normal function (Glickman et al., 2003).

Altogether, these results demonstrate that T is involved in cell movement and induction of a mesodermal cell fate. Cells ingress through the PS, expressing T and obtaining a mesodermal identity. Upon emerging from the PS, these cells acquire different mesodermal fates. Since the studies described in this thesis focused on formation of one of these mesodermal tissues, the PAM, my focus will now shift toward another T-box gene, *Tbx6*, which is required for the proper development of the PAM.

1.5 THE ROLE OF TBX6 IN MESODERM FORMATION

1.5.1 *Tbx6* expression in the PS and PSM

In the mouse, *Tbx6* expression commences at e7.0 in the PS, and by e7.5, it is also expressed in the lateral wings of mesoderm that are fated to form PAM. By e10.5, Tbx6 is expressed in the tailbud, which takes over from the PS as the source of somitic mesoderm (Figure 1-3). *Tbx6* expression ceases after e12.5 when the tail bud no longer produces embryonic mesoderm (Chapman et al., 1996a). *Tbx6* expression in the PSM is rapidly down regulated as the PAM undergoes segmentation to form the somites. Therefore, *T* and *Tbx6* are both expressed in the PS and the tailbud (Chapman et al., 1996a). Along with this overlapping expression, *T* and *Tbx6* also have unique domains of expression with *T* expressed in the notochord and *Tbx6* expressed in the PSM (Chapman et al., 1996a; Wilkinson et al., 1990).

1.5.2 Null mutations in *Tbx6* result in loss of PAM

Tbx6 is essential for the specification of the PAM. This was shown by targeted disruption of the *Tbx6* gene ($Tbx6^{tm1Pa}$), in which the first two exons and a portion of the third exon were replaced by a selectable marker (Chapman and Papaioannou, 1998). As detailed in Chapman and Papaioannou (1998), $Tbx6^{tm1Pa}$/ + animals were viable, fertile, and had no visible abnormalities, however homozygous $Tbx6^{tm1Pa}$ mutations led to embryonic lethality. By e9.5, *Tbx6* homozygous mutants could be recognized by a lack of trunk somites and an enlarged tail bud. By e11.5, approximately 20% of the homozygous $Tbx6^{tm1Pa}$ embryos were dead as determined by lack of a heartbeat; all died by e12.5 due to vasculature problems. The homozygous $Tbx6^{tm1Pa}$ mutant embryos formed ectopic neural tubes, one on either side of the primary neural tube in place of the somites; using several neural markers, it was further determined that these ectopic neural tubes were properly patterned along their dorsal-ventral axis (Chapman and Papaioannou, 1998).

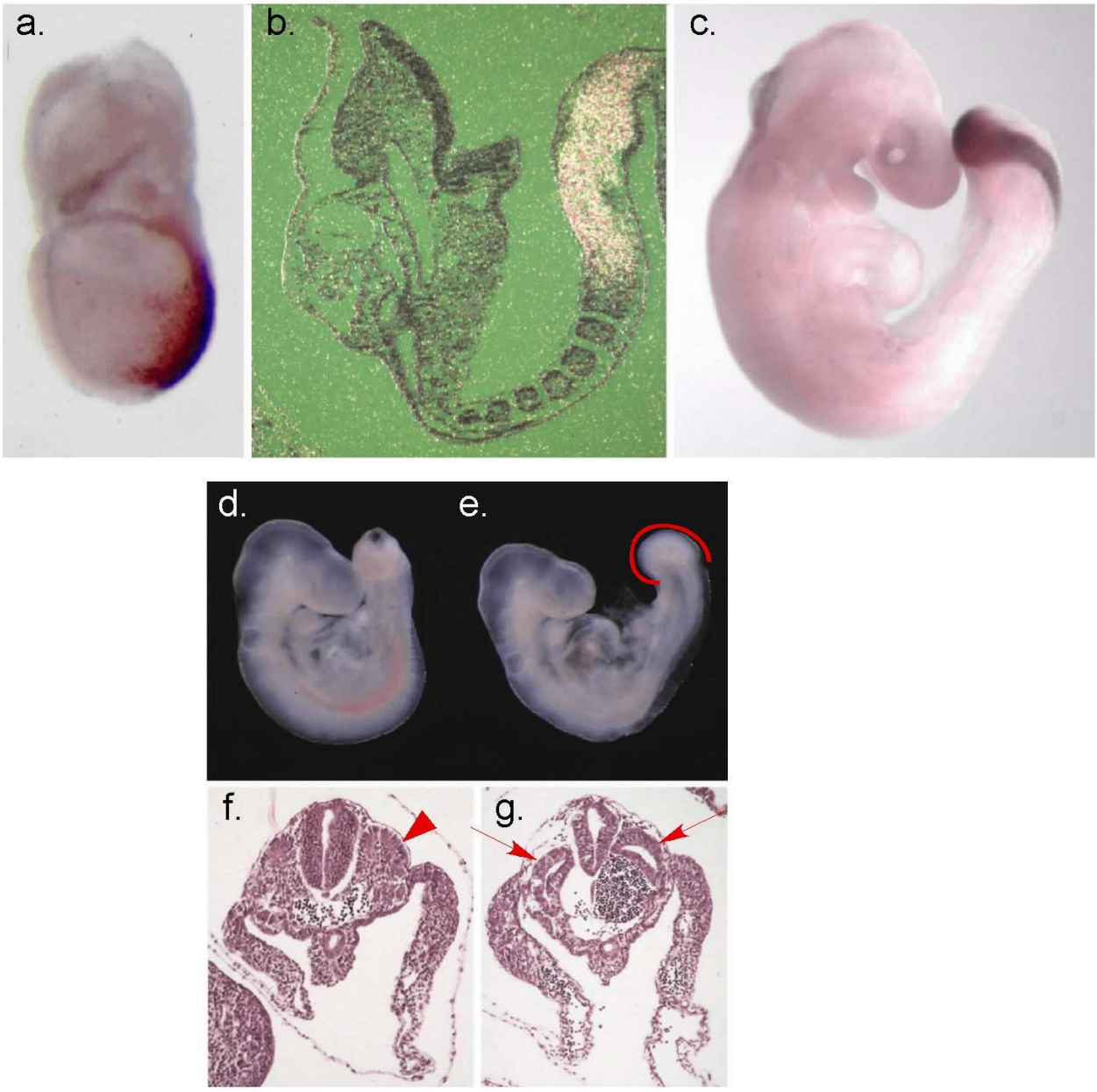


Figure 1-2: Loss of *Tbx6* expression leads to ectopic neural tube formation

Panels (a-c) adopted from Chapman et al. (1996a) and (d-g) modified from Chapman and Papaioannou (1998). Normal (a-d, f) and mutant (e-g.) *Tbx6* embryos at e7.5 (a), e8.5 (b), and e9.5 (c-g). The wild type expression of *Tbx6* during development using whole mount immunocytochemistry (a, c) and section *in situ* hybridization (b). Comparison of wild type and mutant embryos during development (d-g) showed that Tbx6 null embryos had no visible trunk somites and an enlarged tail bud (red line, e). Histological transverse sections through the trunk showed the formation of bilateral somites flanking the neural tube in wild type embryos (arrowhead) compared to the formation of ectopic neural tubes in Tbx6 mutant embryos (arrows) in place of somites.

T was expressed throughout the expanded tail bud, and *sonic hedgehog* (*shh*) was expressed in the notochord (confirming no axial mesoderm defects). *wnt3a*, normally expressed in the PS and tailbud, was expressed in the expanded tailbud, but more central than T (Chapman and Papaioannou, 1998; Takada et al., 1994). Three mesodermal markers, *Dll1*, *paraxis*, and *Meox1* (*Mox1*), normally expressed in the PSM, were all lost in the homozygous *Tbx6*^{tm1Pa} mutants (Chapman and Papaioannou, 1998). Therefore, PS markers such as *T* and *wnt3a* were found to be expressed in the *Tbx6* null mutant, while PSM markers such as *Dll1* and *paraxis* were not, suggesting that differentiation of PS into PSM was blocked (Chapman and Papaioannou, 1998).

1.5.3 Tracing the evolutionary lineage of *Tbx6*

Tbx6 orthologs have also been identified in many organisms, however the identification of orthologs has been more difficult than that of T. Human *TBX6* has been cloned and initial characterization completed (Papapetrou et al., 1999). *TBX6* has been localized to chromosome 16, and RT-PCR showed *TBX6* expression in adult testis, kidney, lung, muscle, and thymus tissues in addition to its expression during development (Papapetrou et al., 1999).

Aside from human TBX6, the *Xenopus laevis Tbx6* has been cloned (Uchiyama et al., 2001). *XTbx6* expression was similar to that of the mouse, with expression around the blastopore (the *Xenopus* equivalent of a PS) excluding the dorsal midline, throughout the PSM and lateral mesoderm, and in the tailbud but not in the somites (Uchiyama et al., 2001). Overexpression of *XTbx6* by mRNA microinjection into a 4-cell stage embryo resulted in gastrulation defects in 50% of the injected embryos. These defects included no head region, enlarged somites, notochord differentiation errors, and a neural tube containing cells with pyknotic nuclei, or condensed chromatin within cells typically linked to programmed cell death (Uchiyama et al., 2001). In addition, *XTbx6* had been overexpressed in animal cap cells to determine whether *XTbx6* affected cell fate decisions. *XTbx6* and *XTbx6* with a VP16 activator developed into mesenchyme and mesothelium, while *XTbx6* with the repression domain of *Drosophila engrailed* gene developed into epidermis (Uchiyama et al., 2001). Uchiyama et al. (2001) concluded that *XTbx6* was a moderate activator because the *XTbx6* with a VP16 activator was roughly 100-fold more potent than *XTbx6* at inducing mesoderm in animal cap assays. Interestingly, co-injection

of *XTbx6* with *Noggin* mRNA, an antagonist of BMP signaling, resulted in differentiation of the animal cap cells into well-developed muscle, while *Noggin* injection alone had no effect on cell fate; these results suggested that *Tbx6* acted together with a dorsalizing factor, such as *noggin*, in *Xenopus* somite formation (Uchiyama et al., 2001). Uchiyama et al. (2001) also reported that sequence analysis suggested that XTbx6 was most closely related to murine Tbx6 by predicted T-box amino acids sequence similarity. All of the above data suggested that *XTbx6* was a functional ortholog of the mouse *Tbx6*.

Genes related to *Tbx6* have been observed in *Drosophila* (Hamaguchi et al., 2004). The *Dorsocross* genes (*Doc1-3*) were determined to be *Drosophila* T-box genes that were most closely related to murine *Tbx6* by sequence identity. The expression patterns of the three *doc* genes overlapped and closely resembled the expression pattern of *dpp,* a member of the *Drosophila* BMP family of genes (Hamaguchi et al., 2004; Reim et al., 2003). These three genes appeared to be functionally redundant in wing development, and the expression of *doc1* required the presence of two BMP homologs, *dpp* and *screw* (*scw*) (Hamaguchi et al., 2004). Hamaguchi et al. (2004) tested whether *Doc1* was an ortholog of *XTbx6* by injecting *Doc1* mRNA into *Xenopus* animal caps, similar to studies with *XTbx6* (Uchiyama et al., 2001). *Doc1* induced ventral mesodermal tissues such as mesenchyme and mesothelium, although the activity of *Doc1* was noted to be slightly lower; both *Doc1* and *XTbx6* induced *Xvent-1* in animal caps, although again the activity of *Doc1* was noted to be slightly lower (Hamaguchi et al., 2004; Uchiyama et al., 2001).

Interestingly, while zebrafish, chicken, and Fugu have *Tbx6* genes, these are unlikely to be the true orthologs of mouse/human *Tbx6*. It is known that genome duplications have occurred during evolution, and it is currently thought that the true ortholog of *Tbx6* was lost in these species (Agulnik et al., 1996). However, I will discuss the T-box genes found in the PS and PSM in zebrafish because studies have examined the interactions between these T-box proteins.

Zebrafish have two T-box genes that are expressed in the PSM, namely *spadetail* (*spt*) and zf*tbx6*. Although named *tbx6*, zf*tbx6* is now known to be paralogous to *Tbx16*; neither zf*tbx6* nor *spt* appeared to be functional homologs of mouse or *Xenopus Tbx6* (Ruvinsky et al., 1998). When ectopically expressed in zebrafish notochord precursors, *spt* promoted a switch to muscle cell fate like *XTbx6* (Halpern et al., 1995). However, trunk PAM continued to develop in

spt mutants, unlike in mouse and *Xenopus Tbx6* mutants (Griffin et al., 1998; Ho and Kane, 1990). zf*tbx6* was determined to provide a redundant function for *spt* (Griffin et al., 1998).

Even though they were not orthologs of mouse *Tbx6*, zf*tbx6* and *spt* genetically interacted and shared some functional similarities to mouse Tbx6 during somitogenesis. In zebrafish, *spt*, zf*tbx6*, and *ntl* were expressed in partially overlapping domains (Griffin et al., 1998; Hug et al., 1997; Schulte-Merker et al., 1992; Schulte-Merker et al., 1994). Experiments by Goering et al. (2003) and others had shown that *ntl*, *spt*, and zf*tbx6* interacted in various ways: combinatorial interactions, additive contributions to express target genes, and competitive antagonism of target gene activity. A combinatorial interaction was observed between *ntl* and *spt* because both were necessary for the expression of *mesogenin* and *wnt8* (Goering et al., 2003). *ntl*/*spt* double homozygous mutants caused a more severe phenotype than either homozygous mutant on its own, supporting the idea that there were additive contributions to development (Amacher et al., 2002). Support for additive contributions of the T-box genes *ntl* and *spt* came from genetic experiments: mutations in either *ntl* or *spt* resulted in a decrease in zf*tbx6* expression, however loss of even one copy of *ntl* in a *spt* mutant embryo resulted in the absence of zf*tbx6* expression (Goering et al., 2003). Lastly, it appeared that zf*tbx6* could act as an antagonist of *ntl*. It is known that ectopic expression of *ntl* in animal cap cells induced *MyoD* expression in a dose-dependent manner (Cunliffe and Smith, 1992). Studies by Goering et al. (2003) showed that co-injection of *ntl* along with increasing concentrations of either zf*tbx6* or *ntl-En*^{R} construct resulted in a failure to induce *MyoD*; the dosage responses using either zf*tbx6* or *ntl-En*^{R} appeared similar, such that inhibition may have been due to competitive binding of zfTbx6 and Ntl. These studies showed that co-expressed T-box genes interacted both *in vivo* and *in vitro*.

Outside of the mouse, frog and zebrafish, *Tbx6* orthologs had been characterized by showing similar expression patterns to mouse *Tbx6* with links to mesoderm formation. There have been reports of *Tbx6*-related genes identified in the chicken and several invertebrates, including three *Tbx6* variants in the *Ciona intestinalis* genome and the ascidians *Molgula tectiformis* and *Halocynthia roretzi* (Mitani et al., 1999; Takada et al., 2002). These proteins were classified as *Tbx6* orthologs using initial sequence alignments based upon predicted protein T-box sequences when only a few T-box genes were known. As described above for zf*tbx6* in zebrafish, there are two chicken variants of *Tbx6*, *Ch-Tbx6T* and *Ch-Tbx6L*, and neither chick gene is thought to be the ortholog of murine *Tbx6* (Knezevic et al., 1997).

1.5.4 Genetic interactions between *T* and *Tbx6* in the mouse

Even though *T* is expressed earlier in development than *Tbx6*, *T* does not directly regulate the onset of *Tbx6* expression because *Tbx6* is initially expressed in *T/T* mutant embryos. The morphological abnormalities of *T/T* mutants were recognizable suggesting that maintenance of *Tbx6* expression required *T* either directly or indirectly (Chapman et al., 1996a). Double homozygous mutants of *Tbx6* and T^{Wis} were generated to understand how *Tbx6* and *T* mutations genetically interact. These $Tbx6^{tm1Pa}/Tbx6^{tm1Pa}$ T^{Wis}/T^{Wis} double homozygous mutants were indistinguishable from T^{Wis}/T^{Wis} mutant embryos, suggesting that *T* was epistatic to *Tbx6* (Chapman et al., 2003). However, $Tbx6^{tm1Pa}/Tbx6^{tm1Pa}$ $T^{Wis}/+$ embryos showed a combined effect of two mutations. The tailbud was typically enlarged in the *Tbx6* null homozygous mutants, however in $Tbx6^{tm1Pa}/Tbx6^{tm1Pa}$ $T^{Wis}/+$ embryos, the tailbud was approximately the size of a normal embryo. This was presumably due to the combined effects of both mutations: excess tail tissue due to the *Tbx6* mutant and reduced mesoderm production due to the T^{Wis} mutation (Chapman et al., 2003).

1.5.5 Separate enhancer regions regulate *Tbx6* expression

Studies by other members of the Chapman lab have begun to address the genetic pathway leading to somite formation. *Tbx6* is essential for posterior PAM formation (Chapman and Papaioannou, 1998). To understand the genetic pathway responsible for posterior PAM, a transgenic approach was used to identify cis-acting regulatory regions capable of driving the expression of a reporter gene in a *Tbx6*-specific pattern during development.

In the mouse, the *Tbx6* gene is approximately 4 kb. To recapitulate the expression pattern of *Tbx6*, an 8.8 kb *Tbx6* genomic reporter construct containing the entire *Tbx6* gene, along with approximately 2.5 kb upstream and approximately 2.3 kb downstream of the gene was generated. This 8.8 kb genomic fragment was shown to direct expression of a *lacZ* reporter in a similar spatial and temporal pattern as endogenous *Tbx6* (White et al., 2003). The only observable difference between endogenous *Tbx6* and the staining of β-galactosidase was residual staining in the somites that was due to perdurance of β-galactosidase and not continued reporter gene expression (White et al., 2003).

Work by others in the Chapman lab generated a series of reporter constructs to narrow down this 8.8 kb enhancer and the expression patterns described below are summarized in Figure 1-3 (modified from White et al., 2005). Expression of a reporter construct containing the first 2.3 kb of the previous 8.8 kb sequence (2.5 kb upstream to approximately 200 bp upstream of the start of *Tbx6* transcription) directed expression of the β-galactosidase reporter in the PSM, like that of endogenous *Tbx6*, but also resulted in ectopic expression of β-galactosidase in the caudal halves of somites. The timing of this PSM reporter, known as $Tbx6^{Tg122/lacZ}$, was slightly delayed compared to endogenous *Tbx6*, with expression commencing at approximately e8.5, rather than at e7.5, in the PSM. This reporter was not active in the PS. These results suggested that sequences responsible for restricting *Tbx6* expression to the PSM were contained within this 2.3 kb PSM enhancer region (Figure 1-3). Expression of a reporter construct containing approximately 200 bp upstream of *Tbx6* and the entire *Tbx6* gene directed expression of the *lacZ* reporter in the PS at e7.5, like that of endogenous *Tbx6*, but also resulted in ectopic expression of *Tbx6* in various forms of mesoderm at various times in development (Figure 1-3). Therefore, this enhancer (which contains the entire *Tbx6* genomic region and a small portion of upstream sequences) contained sequences responsible for driving *Tbx6* expression in the PS, but also lacked sequences that restrict *Tbx6* expression, as evident in ectopic expression of β-galactosidase throughout various forms of mesoderm.

1.6 T-BOX TRANSCRIPTION FACTORS AND SIGNALING PATHWAYS

Transcription factors are the downstream effectors of signaling pathways that function to control the expression of downstream target genes. There are a handful of signaling pathways and these are recycled throughout development with some specificity provided by the downstream tissue-specific transcription factors. It has been well-documented that the Notch signaling pathway plays critical roles in the formation of somites (reviewed in Saga and Takeda, 2001). Interestingly, expression of *Delta-like ligand 1* (*Dll1*), a Notch signaling receptor, is lost in *Tbx6* null embryos suggesting that Notch signaling is downstream of *Tbx6* (Chapman and

Papaioannou, 1998). Below I will describe the generalized Notch signaling pathway and the current knowledge of how Notch signaling functions in somitogenesis.

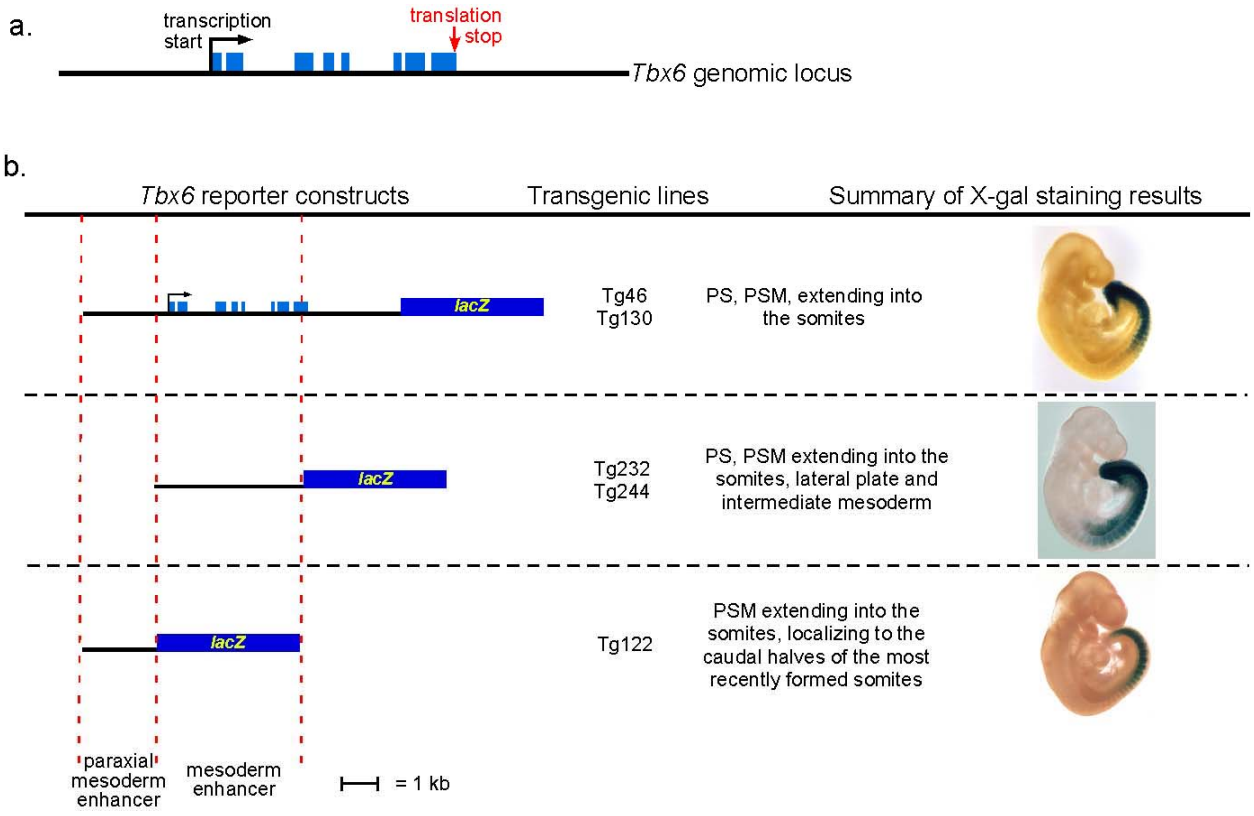


Figure 1-3: Analysis of the cis-acting regulatory elements for Tbx6 expression

Modified from Figure 1 in White et al. (2005). (a.) Schematic of the Tbx6 genomic locus with the positions of the start of transcription (black arrow), exons (blue boxes), and stop codon (red arrow) shown. (b.) Schematic and names of Tbx6 transgenic constructs and a summary of β-galactosidase staining results at e9.5.

1.7 THE NOTCH SIGNALING PATHWAY

1.7.1 Notch signaling pathway overview

Notch signaling is known to play many vital roles throughout development. The signaling pathway is evolutionarily conserved, functioning to exchange signals between neighboring cells through the Notch receptor and its ligands. The Notch signaling pathway was so named by the phenotype of heterozygous mutations in the Notch receptor within the fly wing that resulted in a 'Notched' wing (Artavanis-Tsakonas et al., 1999; Mohr, 1919). As reviewed in Schweisguth (2004) and depicted in Figure 1-4, the Notch receptor is a cell-surface type 1 transmembrane protein, exhibiting both intracellular and extracellular domains. The extracellular portion of Notch binds the ligands of the signaling pathway, Delta and Serrate in *Drosophila* (see below). Following ligand binding, the Notch receptor undergoes successive proteolytic cleavages; the protein Furin cleaves the extracellular domain, releasing a small portion of the Notch extracellular domain, while the intracellular domain is processed by the gamma-secretase complex, including Presenilin and leading to the release of the intracellular Notch domain (ICN). The ICN contains a nuclear localization signal (NLS) and is the active form of the receptor. The ICN translocates to the nucleus, where it forms a complex with the signaling pathway's primary, if not only, transcription factor Suppressor of Hairless (Su[H], or RBP-Jκ in mammals). In the absence of Notch signaling, Su[H] recruits repressors to the cis-acting regulatory regions of target genes, thus acting as a repressor (Bray and Furriols, 2001; Kao et al., 1998). Activation of the Notch signaling pathway results when ICN binds to Su[H] and switches Su[H] to a transcriptional activator (Bray and Furriols, 2001; reviewed in Schweisguth, 2004). This type of signaling, using proteolytic cleavage of receptors, has several features of interest, including irreversible signaling (each receptor can only be used once) and no additional messengers needed to exist in the pathway (reviewed in Schweisguth, 2004).

As mentioned previously, Notch binds the ligands of the signaling pathway, Delta and Serrate in *Drosophila*. In mammals, the pathway is more complicated as there are four Notch receptors (Notch1-Notch4), three versions of Delta (Delta1-3), and two Jagged ligands (the Serrate homolog, Jagged1, and Jagged2) (reviewed in Weinmaster, 1997). Most of these genes and their products are expressed ubiquitously throughout the developing organism. Just as the

Notch receptor is processed by proteolytic cleavage, so are its ligands; the roles that these cleaved ligands play in signaling are only now being revealed.

Notch signaling functions in gonadal cell fate decisions in *C. elegans*, in the development of the eye in *Drosophila*, in all three germ layers during development, and in many processes in mammals, including development of neural and mesodermal tissues (reviewed in Greenwald, 1998; Hartenstein et al., 1992; reviewed in Louvi and Artavanis-Tsakonas, 2006). Mutations in the components of the Notch signaling pathway not surprisingly resulted in many developmental abnormalities and human pathologies (Louvi et al., 2006). Notch signaling occurs in stem cell populations and is now being considered as a cancer therapeutic target through inhibition of the gamma-secretase complex as well as general inhibitors of Notch signaling (reviewed in Chiba, 2006; reviewed in Miele et al., 2006). Among the many systems studied involving Notch signaling, the Notch signaling pathway has been observed to be crucial in somitogenesis (reviewed in Saga and Takeda, 2001).

1.7.1.1 *Dll1* contains enhancer elements responsible for driving expression in the PS and PSM during somitogenesis

The Notch ligand *Dll1* is expressed in the PS and PSM, later becoming localized to the caudal halves of the somites in the mouse (Bettenhausen et al., 1995). *Dll1* is expressed in a number of other tissues, including the neural tube, but for this work, we are specifically concerned with *Dll1* expression in the PS and PSM. *Dll1* enhancer elements capable of driving *lacZ* reporter expression in *Dll1*-specific expression domains have previously been identified (Beckers et al., 2000a). The $Dll1^{tg4.3/lacZ}$ enhancer consisted of 4.3 kb upstream of the *Dll1* translation start site and included two mesoderm enhancer elements, $Dll1^{tg'msd'/lacZ}$ and $Dll1^{tg1.6/lacZ}$. The 1.5 kb $Dll1^{tg'msd'/lacZ}$ enhancer, located between two conserved neural enhancer elements, functioned as a PSM enhancer. Interestingly, this enhancer was not continuously active in the PAM, as demonstrated by the absence of reporter expression in the nascent somites and its later activity in more mature somites and their derivatives (Beckers et al., 2000a). A second mesoderm-specific enhancer, $Dll1^{tg1.6/lacZ}$, consisted of a 1.6 kb genomic region upstream of the *Dll1* start of translation and was active in the PAM as well as in the ventral neural tube at e10.5.

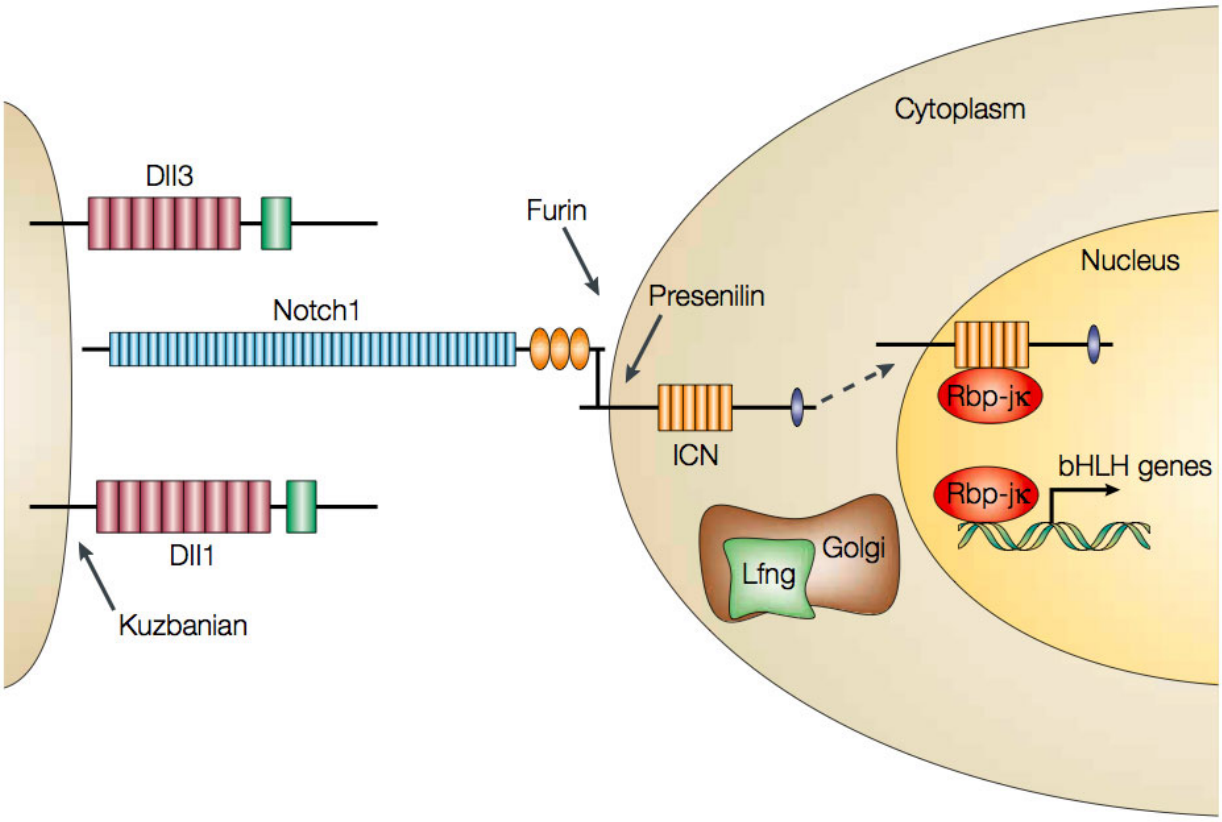


Figure 1-4: Components of the Notch signaling pathway thought to function in somitogenesis

Extracellular regions of the Notch1 receptor interact with the Delta ligands (Dll1, Dll3) to activate Notch signaling. Upon activation, the extracellular domain of Notch is cleaved by Furin, and the membrane-bound intracellular domain of Notch (ICN) is further processed by the gamma-secretase complex, which includes Presenilin (Psen1), resulting in release of the ICN; the ICN interacts with RBP-Jκ and translocates to the nucleus to activate downstream targets. Figure from Saga and Takeda, 2001.

1.7.1.2 A genetic interaction between *Dll1* and *Tbx6*

Dll1 and *Tbx6* were shown to be co-expressed in the PS and PSM (Bettenhausen et al., 1995; Chapman et al., 1996a). Like other genes that were normally expressed in the PSM, *Dll1* expression was lost in *Tbx6* null mutant embryos (Chapman and Papaioannou, 1998). There were two different explanations why *Dll1* may be lost; either *Dll1* was a target of Tbx6 or the

loss of *Dll1* was an indirect effect due to the loss of PSM tissue in the *Tbx6* mutant. Mice heterozygous for either *Tbx6* or *Dll1* were normal, however animals heterozygous for both *Tbx6* and *Dll1* null mutations had fusions of ribs and vertebrae; this demonstrated a genetic interaction between *Dll1* and *Tbx6* (White et al., 2003). Altogether, these results supported the hypothesis that *Dll1* was a direct target of Tbx6.

1.8 BIOCHEMICAL ANALYSIS OF T-BOX PROTEIN FAMILY

Genetic and mutant analyses have shed light on the function of various T-box genes. Because *T* was the first T-box mutant characterized, it had been the focus of the majority of the *in vitro* studies involving structure and function of T-box genes. Prior to the work described in this thesis, characterization of the Tbx6 protein was limited. However several other T-box proteins had been characterized. After detailing the major finding with T, I will discuss below the comparative analysis of the T-box family and then look more closely at a select few members that aid in our understanding of Tbx6 structure and function.

1.8.1 Biochemical analysis of T reveals a new, evolutionarily conserved mode for a DNA binding protein to recognize DNA

1.8.1.1 T is an evolutionarily conserved DNA binding protein

T was demonstrated to be a DNA binding protein (Kispert and Herrmann, 1993). Using a PCR-based binding site selection procedure in which the T protein selected its preferred binding sequence from a pool of random double-stranded oligonucleotides, a consensus sequence representing the binding site for T was identified (Kispert and Herrmann, 1993). These initial binding site studies showed that T bound to an almost perfect 24 bp palindromic sequence; specifically it preferred the sequence 5'- AGGTGTGAAATT -3' as one half-site and the second half-site showed less of a preference (Kispert and Herrmann, 1993). This binding site was evolutionarily conserved because it has been shown that the *Xenopus* and zebrafish orthologs

bound to the palindromic sequences as well (Kispert and Herrmann, 1993). The next step was to determine exactly how T recognized this consensus sequence.

Various interference analyses were utilized to determine the contribution of all of the individual bases within the consensus sequence (Kispert and Herrmann, 1993). DNase I footprinting showed that the T protein protected the entire palindromic sequence and some flanking bases (Kispert and Herrmann, 1993). Methylation interference analysis examined steric interference of protein-DNA interactions; methylation of the guanosine residue at position +5 in the palindromic sequence of T exhibited strong interference, lending support to T activity as a major groove contacting protein, however other bases were contacted suggesting that T also bound within the minor groove (Kispert and Herrmann, 1993). The authors used depurination interference experiments to demonstrate that removal of the guanine residue at position +5 of the palindromic sequence showed strongest interference while removal of purines at positions +1, +2, +3, and +7 interfered less (Kispert and Herrmann, 1993). Lastly, removal of any pyrimidines with hydrazine treatment, except flanking thymines at positions +11 and +12, interfered with DNA binding suggesting that T primarily contacted pyrimidines (Kispert and Herrmann, 1993). These contacts characterized the sequence specificity of T for its DNA target sequence described above. The authors sought to determine how much of the palindromic sequence was necessary for T binding; deletion analysis showed that less than 20 bp of the sequence resulted in the loss of T binding while 22 bp or more sequence resulted in strong binding of T as determined in EMSAs (Kispert and Herrmann, 1993).

1.8.1.2 The T-box family of proteins recognize target DNA sequences in a novel manner

The specificity of T to its sequence had been detailed in terms of the sequence itself, however the next major step was to determine the portion of the T protein necessary for DNA binding. This was ultimately narrowed down to an approximately 190 amino acid (AA) T domain (Kispert and Herrmann, 1993). Deletion analysis using various expression constructs showed that the N-terminal portion of the protein was necessary for DNA binding; deletion of the first 18 AA of the protein resulted in very weak binding as observed in EMSA experiments. Therefore, the DNA binding function required a minimum of AA 18 – 229, including the T-box domain (Kispert and Herrmann, 1993).

The X-ray crystal structure for Xbra, the *Xenopus* ortholog of T, has been solved. As previously stated, murine T and its orthologs in *Xenopus* and zebrafish bound the palindromic sequence *in vitro* (Kispert and Herrmann, 1993). Therefore, the structure was generated using a bacterially expressed Xbra protein consisting of AA 1 – 226 that was co-crystallized with a 24 bp DNA duplex matching the T consensus sequence (Muller and Herrmann, 1997). The structure was solved to 2.5 Å resolution. The Xbra crystal structure showed that T binds as a dimer to the imperfect palindrome sequence with the dimer interface lying above the major groove (Muller and Herrmann, 1997). The solved structure contained AA 39 – 222 of one monomer, AA 39 – 221 of the second monomer, and the entire palindromic sequence (Muller and Herrmann, 1997).

The Xbra crystal structure showed that the dimer formed a very large arc that spanned the DNA and allowed Xbra to contact its 20 bp of recognition sequence, forming a large 10 Å diameter hole between the dimer interface and the contacts made with DNA (Muller and Herrmann, 1997). The region required for dimerization was localized within the T-domain; the amino acids involved with dimerization contacts formed a salt bridge and hydrophobic patch. Since the amino acids required for dimerization have been identified, it should be possible to use this information to predict whether other T-box proteins form dimers. These amino acids were predicted to be lost in many homologs, including Tbx3, Tbx5, and Tbx6, suggesting that loss of these AAs may result in failure of these homologs to dimerize upon binding DNA (Muller and Herrmann, 1997). This dimerization interface was relatively small (250 $Å^2$), which supported observations that T was a monomer in solution, dimerizing only when in contact with DNA (Muller and Herrmann, 1997). As stated by Coll et al. (2002), a "typical" protein/protein interface generally required much greater amounts of contact between protein subunits to form stable, binary complexes in solution (Lo Conte et al., 1999). Therefore, the small interface in Xbra was consistent with sequential binding, allowing a more rapid complex formation while preventing the dimers from being kinetically trapped on low-affinity sites (Kohler and Schepartz, 2001). This small dimerization region also explained why the addition of polyclonal antibodies was observed to stabilize the binding of T to DNA (Kispert and Herrmann, 1993).

The Xbra protein contacted both the major and minor grooves of DNA, although most contacts were directed to the minor groove; these extensive contacts managed to occur without drastically bending the DNA, a rather rare outcome for DNA-binding proteins that extensively recognized the minor groove (Muller and Herrmann, 1997). Ultimately, this structure explained

why greater sequence conservation was required of one half-site while the second half-site was less stringent; one strong and one weak half-site were sufficient for DNA binding. However, for proteins that require dimerization like T, a single half site was not sufficient for binding DNA (Kispert and Herrmann, 1993). Therefore, *in vivo* target sites of T-domain proteins might not occur as palindromic sequences, but rather as imperfect indirect repeats (Muller and Herrmann, 1997). A subset of the few targets of T and other T-box proteins that have been identified are described below.

1.8.1.3 Characterization of the tissue-specific transcription factor, T, reveals activation and repression regions

T was shown to act as a transcriptional activator using an *in vitro* expression system. Briefly, the murine *T* was expressed in HeLa cells along with a chloramphenicol acetyltransferase CAT reporter under the control of a β-globin minimal promoter with one or two T palindromic sites; these experiments resulted in 17 or 79 fold activation, respectively (Kispert et al., 1995a). Interestingly, although T was unable to bind one half-site sequence by EMSA, T could bind multiple half-sites to activate CAT expression *in vivo* (Kispert and Herrmann, 1993; Kispert et al., 1995a). Activation was also observed when the murine *T* was expressed in HeLa cells along with a CAT reporter under the control of a β-globin minimal promoter and two T half-site sequences; the sequences containing two half-sites in the same orientation, separated by 6 bases, and in opposite orientations, separated by 24 bases, resulted in activation of CAT (Kispert et al., 1995a). EMSA and DNase I fingerprinting across the entire binding region (containing the two half-sites) confirmed that T bound these sequences *in vitro* (Kispert et al., 1995a). Deletion constructs were subsequently used in the same *in vitro* assay system to identify an activation domain; although an activation domain resided in the C-terminus of the protein, there were actually two activation domains, existing from AA 230 – 280 and AA 320 – 380, and two repressing domains, existing from AA 280 – 313 and AA 400 – 436 (Kispert et al., 1995a). As T functioned as a transcription factor, it must have a means to enter the nucleus; typically this was mediated by a nuclear localization sequence (NLS). Using deletion analyses as described above, the NLS was narrowed down to AA 137 – 320 (Kispert et al., 1995a).

1.8.1.4 Initial identification of *in vivo* T targets

Thus far, no palindromic sequences have been identified in the *in vivo* targets of T, however multiple half-sites have been found in the enhancer regions of several T-box target genes, including *eFGF*, *Bix1*, *Bix4* and *Xnr5* (Casey et al., 1998; Casey et al., 1999; Hilton et al., 2003; Tada et al., 1998). The importance of the individual half-sites was analyzed by mutating each site individually and then assaying the activity of the mutated enhancers either *in vitro* or *in vivo*. Results from these studies demonstrated that in some cases, one binding site was necessary and sufficient for activity of the enhancer, while in other cases, multiple half sites appeared to be required; either way, in all cases, at least one half-site consisted of the 5'- AGGTGTGAA -3' sequence. These studies supported the hypothesis derived from crystal structure analysis that one correct T half-site was required while other sites may be less similar to the consensus binding half-site sequence (Muller and Herrmann, 1997). In addition, studies by Conlon et al. (2001) showed that orientation and spacing of the T half-sites were critical for determining the binding specificity for Xbra, as well as for other T-box proteins (VegT and Eomesodermin). It is possible that future experiments will demonstrate that each T-box protein will show different requirements for the number, spacing and orientation of half-sites; these factors may contribute to the ability of all T-box protein(s) to modulate transcription of their target genes *in vivo*.

1.8.2 Comparative analysis of the T-box gene family through conservation within the T-box domain

Murine T is so named for the short tail phenotype of heterozygous mice. With the advent of genome sequencing projects and molecular biology, T orthologs (direct homologs) have been identified with a high degree of similarity in sequence, expression pattern and function, between a variety of vertebrates, including fish, frogs, mice and more recently dogs (Haworth et al., 2001; Papaioannou, 2001; Wilson and Conlon, 2002). Interestingly, mutations in canine T result in the short tails of Corgis and other breeds of dogs (Haworth et al., 2001). T is the founding member of the T-domain family of genes that consists of an ever-expanding list of proteins, both characterized and predicted proteins, across a range of invertebrates and vertebrates alike (Agulnik et al., 1995; Bollag et al., 1994; reviewed in Papaioannou, 2001). Based on sequence similarity within the T domain, the T-box family members are divided into the following

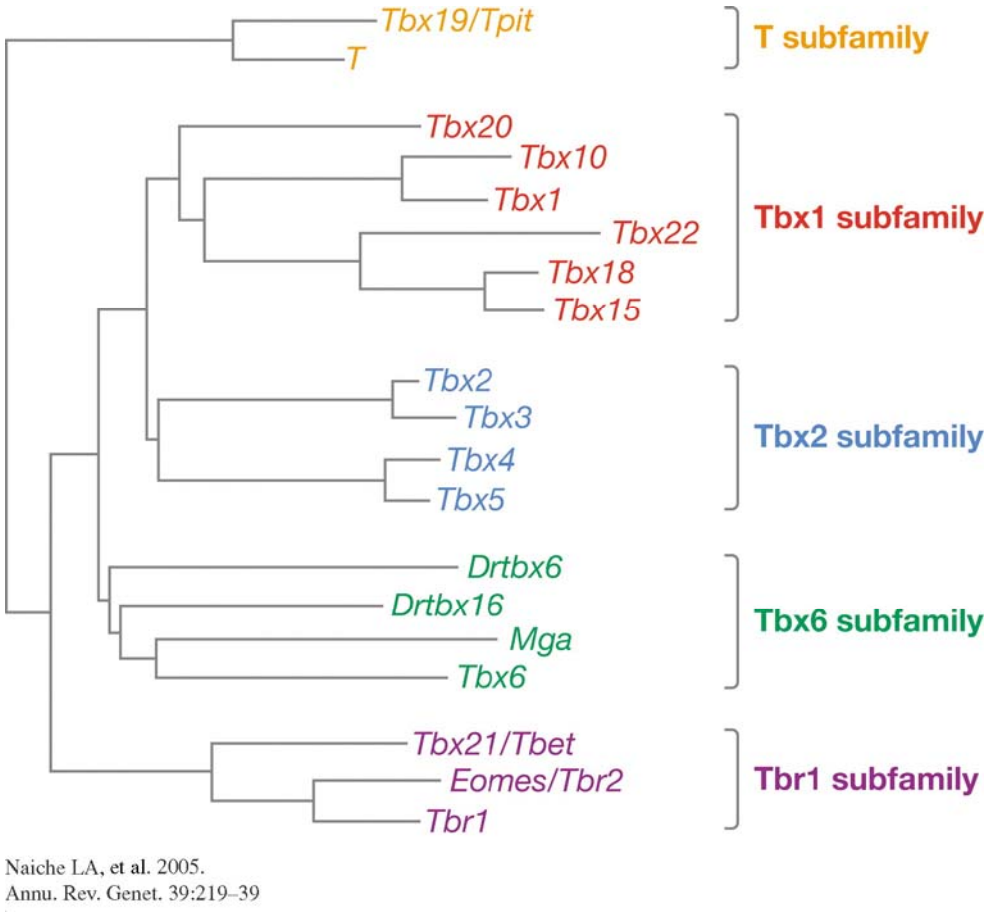


Figure 1-5: Family tree of T-box subfamilies based upon the conserved T-box domain

From Figure 1 of Naiche et al (2005). Schematic phylogenetic tree of the T-box gene family of vertebrates, prepared by comparing T-domain sequences, showing the relationship of T-box proteins in the five subfamilies indicated by the brackets on the right. All of these genes are present in mammals with the exception of the zebrafish genes Drtbx6 and Drtbx16, which do not have orthologs in mammals; therefore the mammalian Tbx6 subfamily consists of only Tbx6 and Mga.

subfamilies: the T, Tbx1, Tbx2, Tbx6, and Tbr1 subfamilies (see Figure 1-5 adapted from Naiche et al. (2005) based on Papaioannou and Goldin, 2003). The Tbx6 subfamily consists of Tbx6 itself, the zebrafish zftbx6 and spadetail (neither are true orthologs of Tbx6), and Mga, a T-box containing protein that also contains a basic helix-loop-helix-leucine zipper motif (Hurlin et al., 1999). The T subfamily, including T and its orthologs, are the furthest evolutionarily from all

other families (Naiche et al., 2005). Tbx6 is most closely related to the Tbx2 subfamily, which includes Tbx2, Tbx3, Tbx4, and Tbx5. This subfamily is among the most well studied of all T-box subfamilies. Below I will highlight some of the major findings for the Tbx2 subfamily as these findings have a major impact on my thesis. I will begin by discussing the binding specificity and crystal structure for Tbx3, then focus on Tbx5. The binding specificity of Tbx5 is known and more recent data showed cooperation between Tbx5 and Tbx20. This is the first demonstration of a synergistic relationship among T-box proteins with mammalian homologs.

1.8.2.1 Crystal structure analysis of TBX3 explains several mutations observed in Ulnar mammary syndrome

Mutations in human TBX3 cause Ulnar mammary syndrome (UMS), a pleiotropic disease affecting development of limbs, apocrine gland, teeth, hair, and genitals (Bamshad et al., 1997). UMS afflicted individuals have mutations either within the T-box DNA binding domain or truncation mutations, which remove various portions of the C-terminus of the protein (Bamshad et al., 1999; Bamshad et al., 1997). The X-ray crystal structure for human TBX3 had been solved in complex with a 24 bp palindromic DNA (Coll et al., 2002). The structure was generated using a bacterially expressed human TBX3 protein consisting of AA 99 – 291 (the T domain) and the 24 bp DNA duplex matching the T consensus sequence; the structure was solved to 1.7 Å resolution (Coll et al., 2002).

The crystal structure showed that TBX3 bound as a monomer to the T almost perfect palindromic site, unlike Xbra, which bound as a dimer. The solved structure contained AA 102 – 285 of one molecule, AA 100 – 285 of the second molecule, and the entire palindromic sequence, meaning that two Tbx3 molecules bound independently to the DNA sequence (Coll et al., 2002; Muller and Herrmann, 1997). Importantly, TBX3 did not contain the amino acids that were shown in Xbra to be involved with dimerization; rather, the amino acids in these positions were similar to the T domain of the other Tbx2 subfamily proteins (Coll et al., 2002; Muller and Herrmann, 1997). The two TBX3 proteins still contacted both the major and minor grooves of DNA like T and maintained all important protein/DNA contacts, except at the very end of the binding site where contacts with the last two thymine residues were lost (Coll et al., 2002; Muller and Herrmann, 1997). As a result, the proteins were rotated away from each other by

approximately 10 degrees in opposite directions; this further diminished the region for potential interactions between the two monomers (Coll et al., 2002).

Some patients with UMS have mutations in the T-box domain, and all of these mutations were predicted to impair DNA binding by rendering the T-box domain unstable, as with the TBX3 mutations L143P and Y149S (Coll et al., 2002). Not all afflicted patients have mutations in the T-box domain; some have mutations after the T-box domain that would result in a truncated protein much like the truncated T mutations in mice (Section 1.4.2). Other studies in Tbx3 had identified a key repression domain in the Tbx3 C-terminus that functioned as a portable repression domain; this domain was lost in the known TBX3 truncation mutants (Carlson et al., 2001).

1.8.2.2 Mutations in TBX5 lead to Holt-Oram syndrome in humans

TBX5 had also been extensively studied as mutations in this gene in humans resulted in Holt-Oram syndrome (HOS), a rare autosomal dominant human disease in which afflicted patients exhibited upper limb malformations and have a very high incidence (>85%) of congenital heart disease (Basson et al., 1997; Basson et al., 1999; Li et al., 1997). Most HOS mutations were caused by haploinsufficiency of *TBX5* (Basson et al., 1999; Li et al., 1997). Some patients with HOS have mutations in the T-box domain and all of these mutations were predicted to impair DNA binding by changing the confirmation of DNA-contacting residues, as with TBX5 mutations G80R, R237Q, and R237W (Coll et al., 2002). *Tbx5* was determined to be expressed during development and shown to function in cardiac morphogenesis and limb development (Bruneau et al., 1999; Chapman et al., 1996b). In the mouse, transgenic approaches have modeled HOS using a cre/lox deletion of *Tbx5*; heterozygous mice modeled HOS, while homozygous mice died early in embryogenesis, revealing an early and fundamental role for *Tbx5* in cardiac development (Bruneau et al., 2001).

Tbx5 bound equally well to the T palindromic consensus site and the T half-site as tested with EMSA experiments (Ghosh et al., 2001; Kispert and Herrmann, 1993). The binding site for Tbx5 had been determined using a PCR-based binding site selection procedure like that for T, however unlike T, the selected consensus sequence consisted of 5'- DRGGTGTBR -3', where D was any nucleotide except cytosine, R was adenine or guanine, and B was any nucleotide except adenine; the authors noted that some of the selected sequences contained only one half-site

(Ghosh et al., 2001; Kispert and Herrmann, 1993). The N-terminal half of the protein, containing the T-box DNA binding domain (AA 1 – 237), was the minimum region required for DNA binding; however the addition of four additional AAs (AA 1 – 241) abolished binding, and removal of the C-terminal half of Tbx5 (AA 238 – 518) significantly enhanced DNA binding affinity compared to full length Tbx5 (Ghosh et al., 2001). The later suggested that truncated proteins could bind to DNA sequences tightly, interfering with other DNA binding proteins that in turn affect the expression of downstream targets.

1.8.2.3 Tbx5 synergistically activates target genes during vertebrate heart development

Several targets have been identified for Tbx5, including *ANF*, *Connexin40* (*cx40*), and *Fgf10* (Bruneau et al., 2001; Koshiba-Takeuchi et al., 2006; Moskowitz et al., 2004). Tbx5 bound to the single Tbx5 half-site found within the 330 bp *ANF* enhancer/promoter to activate transcription of a luciferase reporter gene *in vitro* (Ghosh et al., 2001). Similarly, Tbx5 activated transcription from the *cx40* enhancer that contained 5 predicted Tbx5 binding sites within a 1 kb region (Bruneau et al., 2001). These studies used a variety of mammalian cell types to assay transcriptional activity, including COS-7 cells, a rat cardiomyocyte cell line, H9c2, primary cardiac myocytes, and cardiac fibroblasts (Bruneau et al., 2001; Ghosh et al., 2001). Interestingly, primary cardiac myocytes yielded 100 times greater activation than non-cardiac cell types (Bruneau et al., 2001), suggesting that co-factors present in these cardiac cells synergistically affected the ability of Tbx5 to activate transcription.

It was later found that Tbx5 activated expression of genes in combination with Nkx2.5, synergistically activating both *ANF* and *cx40* as determined by luciferase assays (Hiroi et al., 2001). *Tbx5* and *Tbx20* were determined to be expressed early in *Xenopus* heart development in partially overlapping domains (Brown et al., 2003; Bruneau et al., 1999; Horb and Thomsen, 1999). On its own, Tbx5 resulted in weak activation of a luciferase reporter gene while Tbx20 activated a luciferase reporter construct in a dosage dependent manner. The presence of Tbx20 appeared to alter the ability of Tbx5 to activate the luciferase reporter because at low and high levels of Tbx20, increased activation of the luciferase expression was observed, while at moderate doses, Tbx20 repressed Tbx5 activation of luciferase (Brown et al., 2005). In *Xenopus*, morpholino knockdown of *Tbx5* and *Tbx20* individually led to similar phenotypes, including pericardial edema, reduced cardiac mass, and loss of circulation; injection of morpholinos at half

the levels that previously led to complete knockdown of protein, as judged by penetrance of cardiac defect phenotypes, demonstrated that knockdown of both Tbx5 and Tbx20 together resulted in more penetrant heart defects (74% of embryos affected), compared to either Tbx5 morpholinos alone (4% of embryos affected) or Tbx20 morpholinos alone (13% of embryos affected) (Brown et al., 2005).

Tbx5 and Tbx20 were demonstrated to physically interact by co-immunoprecipitation; the N-terminus and T-box regions were required for co-immunoprecipitation of both proteins (Brown et al., 2005). All of the data above supports the notion that Tbx5 and Tbx20 forms heterodimers that enable these T-box proteins to modulate their transcriptional activities. This is especially interesting as Tbx5 and Tbx20 belong to different T-box subfamilies, thus opening up interactions between different T-box sub-families as possible mechanisms of T-box function.

1.9 THESIS AIMS

At the time I began my thesis research, the expression pattern and critical function of *Tbx6* during mouse development had been determined (Chapman et al., 1996a; Chapman and Papaioannou, 1998). However, nothing was known about the Tbx6 protein in terms of its molecular mass, transcriptional activity, or downstream targets. As a first step toward understanding Tbx6, we generated antibodies against Tbx6 to determine the molecular mass of Tbx6 *in vivo* and whether the protein and mRNA had similar spatial and temporal expression patterns. These Tbx6 polyclonal antibodies have allowed the lab to make progress in a number of areas; in particular they have been useful for the localization of the Tbx6 protein *in vivo* and in tissue culture, biochemical studies for DNA binding, immunoprecipitation studies, and for use in detecting Tbx6 expression from various luciferase reporter constructs.

The function of T-box proteins were mediated by the T-domain, the DNA binding region. Given the conservation of the T-domain, it was not surprising that many of the characterized T-box binding sites were also somewhat conserved. Results from several laboratories have shown that the T-box proteins bound to a core 5'-AGGTGT-3' with flanking sequences exhibiting more variability (Conlon et al., 2001; Kispert and Herrmann, 1993; Sinha et al., 2000; Tada et al., 1998). Many of the T-box proteins bound as monomers to the half-site, however other T-box

proteins, specifically T and TBX1, bound as dimers to the palindromic site (Conlon et al., 2001; Kispert et al., 1995a; Muller and Herrmann, 1997; Sinha et al., 2000; Tada et al., 1998). Therefore, I sought to first determine whether Tbx6 can recognize the T palindromic consensus sequence and half-site sequence *in vitro*, and ultimately to determine a preferred Tbx6 binding site using a PCR-based binding site selection assay. My studies also focus on how Tbx6 affected transcription using *in vitro* luciferase assays. As described earlier, T and Tbx6 expression overlapped in the PS and tailbud. Data thus far suggested that various T-box proteins recognized similar sequences *in vitro* and that physical interactions between specific T-box proteins occurred previously, as shown between Tbx5/Tbx20. I therefore sought to understand the relationship between T and Tbx6 and whether these two T-box proteins (1) compete for binding sites in target genes and/or (2) interact to vary activity of target genes when co-expressed.

In other work, I sought to determine whether the spontaneous mutation, *rib-vertebrae*, which mapped within 2cM of *Tbx6*, was actually a *Tbx6* mutation. Finally, I also sought to understand how the Notch signaling pathway and *Tbx6* interacted. These studies used both transgenic and biochemical techniques to unravel the complex interrelationship of these pathways. The work within this thesis used a variety of approaches *in vivo* and *in vitro*, molecular and biochemical, to gain further understanding of how Tbx6 functioned in the context of a developing embryo while in the presence of other T-box family members to ultimately regulate the expression of target genes.

2.0 CHARACTERIZATION OF TBX6 *IN VIVO*

2.1 INTRODUCTION

With the advent of genomic sequencing centers publicly releasing data, a large volume of sequencing data are now available for analysis. The genomic sequences are overlaid with annotated, published data from various sources; although the genomic data have many advantages, discrepancies can occur. *Tbx6* may be one of those discrepancies.

Tbx6 was predicted to encode a 540 amino acid protein with a molecular mass of 58.6 kDa (Chapman et al., 1996a). This was based on the Sanger dideoxy sequencing of a cloned *Tbx6* cDNA from an e8.5 mouse cDNA library; *Tbx6* was expressed in the PS, tailbud and PSM from e7.5 until e12.5 (Chapman et al., 1996a). However Tbx6 had not been characterized at the protein level and any differences between mRNA and protein expression of Tbx6 were unknown. Aside from the original *Tbx6* cDNA sequence (Accession # U57331, translated AAC53110) (Chapman et al., 1996a), another *Tbx6* cDNA sequence had also been deposited at Genbank that was predicted to generate a protein of approximately 47 kDa (Accession # NM_011538, translated AAT72924). Because of this, another sequencing depository (Ensembl, www.ensembl.org) also described a 1715 bp transcript for *Tbx6*, (ENSMUST00000094037, Ensembl) predicted to encode a 436 amino acid protein with a molecular mass of 47 kDa (Figure 2-1). However, it is important to note that there were only a few differences between these sequences, namely the addition of three one-base insertions in the cDNA at 995, 1155, and 1415 bp. All of these occurred in regions where a tract of 7 to 12 guanines and/or cytosines existed. This difference shifted the reading frame and truncated the predicted protein prematurely compared to the original sequence. We believe the differences between the *Tbx6* cDNAs could be due to sequencing misreads that resulted in frameshift mutations, and changed the predicted protein primary sequence.

2.1.1 Aims of these studies

To study Tbx6 protein function, we must know if these different variants exist *in vivo*. The timing and regions of *Tbx6* mRNA expression were found to be tightly regulated, commencing at approximately e7.0 and being undetectable by e13.5. Expression was limited to the PS, tailbud and PSM. However little was known about Tbx6 protein expression. I sought to determine not only how Tbx6 protein was localized *in vivo*, but also the molecular mass of the Tbx6 protein with the hopes of resolving the molecular mass of the endogenous Tbx6. To do so, I generated custom rabbit antibodies that detect Tbx6. These antibodies could also be useful for other studies, such as immunoprecipitation, gel shift studies, and identifying Tbx6 interacting proteins that could potentially uncover biochemical properties of Tbx6.

```
Multiple sequence alignment

2 Sequences Aligned           Alignment Score = 2090
Gaps Inserted = 0             Conserved Identities = 333

                10        20        30        40        50
AAC53110  MVHPRELVPSLGTGVRLGHPQPGADSTFPPALTEGVRYPDLDTSKLDCFLSGIEA
AAT72924  MVHPRELVPSLGTGVRLGHPQPGADSTFPPALTEGVRYPDLDTSKLDCFLSGIEA
          ************************************************************

             60        70        80        90       100       110
AAC53110  APHTLAAAAPLPPSPIALGPETAPPPPEALHSLPGVSLSLENQELWKEFSAVGTE
AAT72924  APHTLAAAAPLPLLPSALGPETAPPPPEALHSLPGVSLSLENQELWKEFSAVGTE
          ************  * ****************************************

            120       130       140       150       160
AAC53110  MIITKAGRRMFPACRVSVTGLDPEARYLFLLDVVPVDGARYRWQGPDWEPSGKAE
AAT72924  MIITKAGRRMFPACRVSVTGLDPEARYLFLLDVVPVDGARYRWQGQHWEPSGKAE
          ************************************************  *********

              170       180       190       200       210       220
AAC53110  PRLPDRVVIHPDSPATGAHWMRQPVSFHRVKLTNSTLDPHGHLILHSMHKVQPRI
AAT72924  PRLPDRVVIHPDSPATGAHWMRQPVSFHRVKLTNSTLDPHGHLILHSMHKVQPRI
          ************************************************************

            230       240       250       260       270
AAC53110  HLVRATQLCSQHWGGVASFRFPETTFISVTAYQNPRITQLKIAANPFAKGFRENG
AAT72924  HLVRATQLCSQHWGGVASFRFPETTFISVTAYQNPRITQLKIAANPFAKGFRENG
          ************************************************************

              280       290       300       310       320       330
AAC53110  RNCKRERDARVKRKLRGPEPVATEACGSGDTPGGPCDSTLGGDIRDSDPEQAPTP
AAT72924  RNCKRERDARVKRKLRGPEPVATEACGSGDTPGGPCDSTLGGDIRDSDPEQAPTP
          ************************************************************

            340       350       360       370       380
AAC53110  GKLLLPQLLHVGAPVLRPTFYTLPLFMAPPVTYQPGPPAWLRLQTLAPSPLLSCI
AAT72924  QEAASASAPPCGGPSAEAYLLHPAAFHGAPSHLPARTPSXAEAPDPGRPAPYSAA
                  * *           *   *       *                 *

              390       400       410       420       430       440
AAC53110  SGPTAWTRGLCLSGSSICTILCPTLHPRGSLPSTVPRTWRLSGHGIQANVLSLGV
AAT72924  FLDLQPGPGGSAYQAAPSVPSFAPHFIQGGPFPLPYPGPGGYLDMGSKPMY
                *                   *

            450       460       470       480       490
AAC53110  VTLCHLPLILQLPSPSLVASALKWYPSPPTKWLAWASLPPFSSHLDFTPPPLASK
AAT72924

              500       510       520       530       540
AAC53110  LRQGSSESSWGLLPLSSHLSANGGSLCLAKWGSQPSTPTSWGPCY
AAT72924
```

Figure 2-1: Alignment of two variants of Tbx6 from Genbank

Two variants of Tbx6 have been deposited in Genbank. Tbx6 variant AAC53110 reported a 540 amino acid protein, while the Tbx6 variant AAT72924 created a 436 amino acid protein. These two variants were virtually identical from amino acids 1 – 330, and then drastically differed in their C- terminus. (*) denotes identical matches between the sequences at the particular amino acid.

2.2 GENERATION OF TBX6-SPECIFIC ANTIBODIES

I generated Tbx6 antibodies to examine Tbx6 protein localization during development and for use in other biochemical studies. I created three GST-Tbx6 fusion proteins, consisting of the 26 kDa GST with one of three regions of Tbx6, such that all regions excluded the conserved T domain (Figure 2-2a). None of the Tbx6 regions used showed homology to any other protein found in the translated mouse genome (based on a BLAST search). GST fusion proteins were made for use as antigens from the following regions of the Tbx6 protein: the N-terminal Tbx6 fusion protein amino acids (AA) 2-78 [GST-NTbx6], the internal Tbx6 fusion protein AA 311-408 [GST-ITbx6], and the C-terminal Tbx6 fusion protein AA 467-537 [GST-CTbx6]. The constructs were cloned into expression vectors and sequenced, matching the original *Tbx6* sequence, Accession # U57331 (Chapman et al., 1996a).

The three GST-Tbx6 fusion proteins and GST control were expressed on a small scale and tested for solubility; soluble GST fusion proteins could be easily purified and used as antigens for antibody purification. Briefly, the fusion proteins and GST alone were expressed in bacteria using IPTG induction; pelleted bacteria were then resuspended in lysis buffer, sonicated, and centrifuged. Soluble proteins remained in the supernatant, while insoluble proteins pelleted. Soluble and insoluble fractions were separated on 14% SDS-polyacrylamide gels and stained. Figure 2-2b is a representative gel showing that two of the three proteins were isolated as soluble proteins; all three proteins ran at their predicted molecular masses.

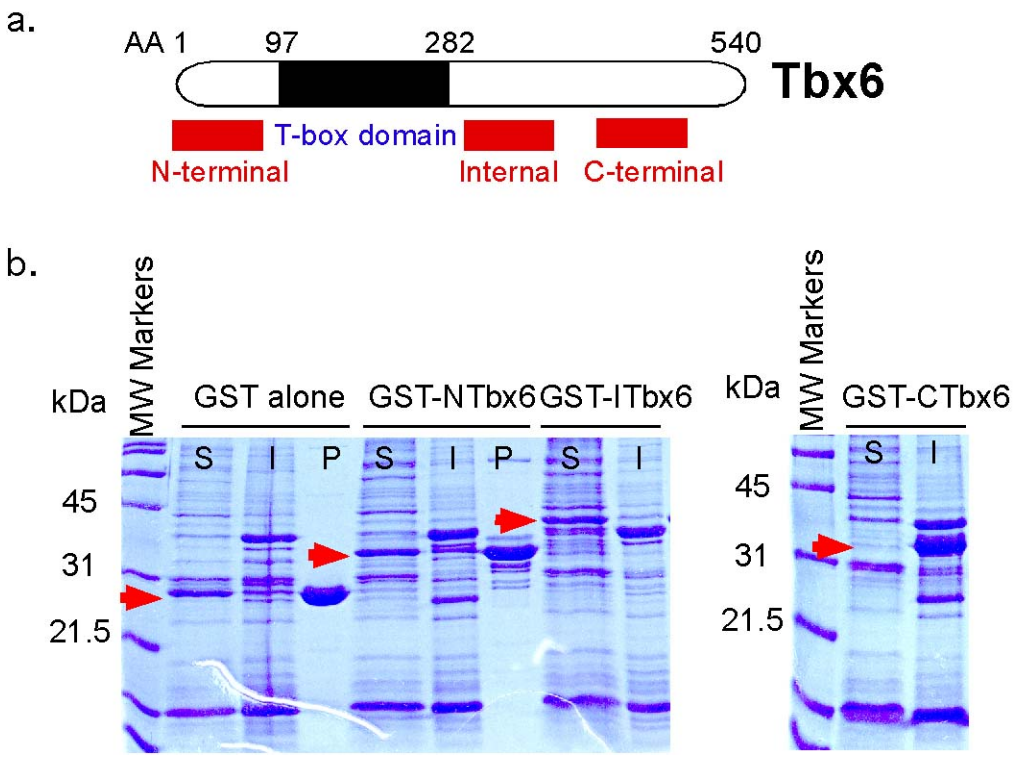


Figure 2-2: Schematic of Tbx6 antigens and initial antigen expression

(a.) Tbx6 antigen map showing the location of the three Tbx6 regions used for generating antibodies. The N-terminal and Internal antibodies were used for Western blots, immunoprecipitations, EMSA gel shifts, and immunocytochemistry experiments. (b.) Coomassie blue stained 14% SDS-polyacrylamide gel showing that GST, GST-NTbx6, and GST-ITbx6 were primarily soluble following small-scale induction of the fusion proteins in BL21 bacterial cells. However, the GST-CTbx6 fusion protein was primarily insoluble. Arrowheads denote bands representing fusion protein, (S) soluble fraction / supernatant, (I) insoluble fraction / pellet, (P) purified GST fusion protein bound to GST resin.

All GST-Tbx6 fusion proteins were expressed on a large scale, purified, and used as an immunogen to generate polyclonal antibodies in rabbits. Briefly, the fusion proteins and GST alone were expressed in bacteria using IPTG induction; pelleted bacteria were then resuspended in lysis buffer, sonicated, and centrifuged. For the soluble GST-NTbx6 and GST-ITbx6 fusion proteins, the supernatant containing soluble fusion protein was purified over a gravity-flow GST-bind resin column. After eluting purified GST-Tbx6 fusion proteins, the proteins were dialyzed against PBS and then used as antigens (Figure 2-3). For the insoluble GST-CTbx6 protein, the insoluble fraction was separated on a 14% SDS-polyacrylamide gel, and the protein band representing GST-CTbx6 was excised from the gel and used as an antigen.

All three GST-Tbx6 antigens were used for antibody production in rabbits by Cocalico Biologicals (Reamstown, PA). Each fusion protein was injected into two different rabbits to control for animal variability. The six rabbits were boosted and bled following the company's standard protocol, with two test bleeds, a larger production bleed procedure, and a terminal bleed. The antisera were given the designations anti-T6N143, N144, I145, I146, C147, and C148, where the N, I, and C were based on the antigen. These antisera were initially tested for Tbx6 specificity by visualization of supershifts in EMSA experiments (Figure 3-2) and immunoprecipitation of radiolabelled Tbx6. The N-terminal and Internal antibodies were useful for Western blots, IPs, EMSA gel shifts, immunofluorescence, and immunohistochemistry experiments. These antibodies were then affinity purified.

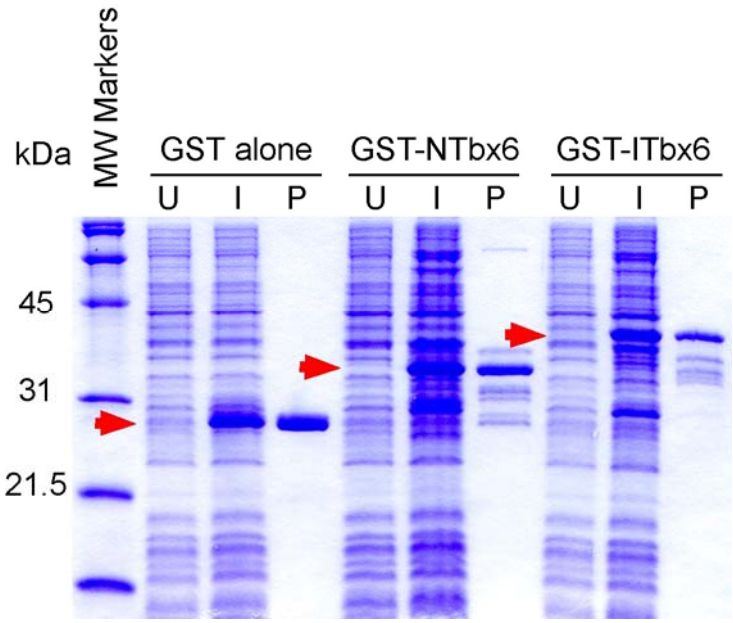


Figure 2-3: Large scale purification of soluble Tbx6 antigens

Representative Coomassie Brilliant blue stained 14% SDS-polyacrylamide gel showing that GST, GST-NTbx6, and GST-ITbx6 were soluble proteins after large-scale expression and purification of the fusion proteins in bacteria. Arrowheads denote locations of bands representing fusion protein, (U) uninduced total cell lysate, (I) induced total cell lysate, (P) purified GST or antigen.

2.3 *IN VITRO* EXPRESSION OF FULL-LENGTH TBX6 GENERATES A 58 KD PROTEIN

Before characterizing Tbx6 expression *in vivo*, we sought to express Tbx6 *in vitro* for use in biochemical assays. Previously, several groups had success using a rabbit reticulocyte lysate system to express T-box transcription factors (Bruneau et al., 2001; Conlon et al., 2001; Kispert and Herrmann, 1993; Papapetrou et al., 1999). Therefore, the full length *Tbx6* cDNA was cloned into a pET expression vector to produce the Tbx6 protein with a T7 epitope tag at its N-terminus, T7-Tbx6 (White and Chapman, 2005). To determine the molecular mass of the T7-Tbx6 protein produced from the expression vector, the construct was *in vitro* transcribed and translated in a

rabbit reticulocyte in the presence of ^{35}S-methionine. Luciferase and Xbra were similarly produced and served as controls in these experiments. Lysates programmed with the T7-Tbx6 plasmid primarily produced a protein of approximately 60 kDa, the expected molecular mass of the Tbx6 fusion protein (540 AA Tbx6 protein with a 7 AA T7 epitope tag). The molecular mass of T7-Tbx6 was determined by comparing the primary protein band against prestained markers, the known 54 kDa V5-His-tagged Xbra protein (gift of F. Conlon), and the 61 kDa luciferase positive control (Figure 2-4). Because our future studies required the immunoprecipitation of Tbx6, immunoprecipitations were performed using commercially available monoclonal antibodies against either the T7-tag or the V5-tag (C-terminus of Xbra fusion protein). Both commercial antibodies were able to immunoprecipitate the radiolabelled proteins (Figure 2-4a).

We next wanted to test our anti-Tbx6 antibodies for their ability to immunoprecipitate Tbx6; therefore, test bleed antisera from each of the six rabbits were used to specifically immunoprecipitate radiolabelled Tbx6. Antisera from rabbits injected with the Tbx6 N-terminal and internal antigens T6N143, N144, I145, and I146 immunoprecipitated a band at the expected molecular mass for Tbx6, while immunoprecipitation reactions using antisera from rabbits injected with the insoluble GST-CTbx6 antigen failed to immunoprecipitate Tbx6. As a negative control, pre-immune sera was used and resulted in no immunoprecipitation of Tbx6 as expected since the pre-immune sera was taken from the source animal before the animal was inoculated with the antigen (Figure 2-5). Altogether, these data indicated that rabbit antibodies generated against two different regions of Tbx6 were useful for immunoprecipitation of *in vitro* generated Tbx6. The N-terminal and Internal Tbx6 antisera were subsequently affinity purified and later re-tested for their ability to immunoprecipitate radiolabelled Tbx6. As shown in Figure 2-4b, affinity purified antibodies immunoprecipitated the ^{35}S-labeled T7-Tbx6 protein, while not precipitating luciferase, thus confirming antibody specificity. A single major band of approximately 60 kDa was present in all immunoprecipitation reactions of the T7-Tbx6 protein.

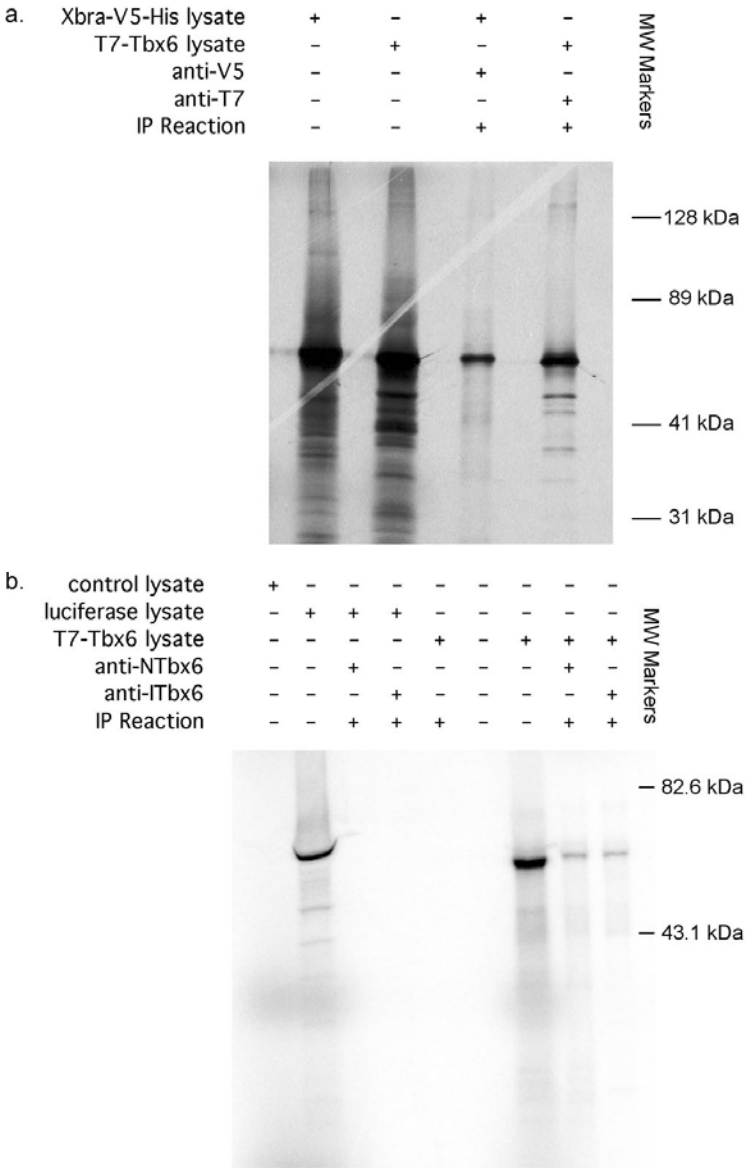


Figure 2-4: Translation of Tbx6 *in vitro* results in a specific 58 kDa protein

Figure published in White and Chapman (2005). (a.) *In vitro* transcribed and translated Xbra or T7-Tbx6 proteins were immunoprecipitated with the anti-V5 and anti-T7 antibodies, respectively. Both immunoprecipitation reactions resulted in the appearance of one major band of the predicted size for each protein. (b.) *In vitro* transcribed and translated T7-Tbx6 or luciferase proteins were immunoprecipitated with the Tbx6 N-terminal (anti-NTbx6) or Internal (anti-ITbx6) antibodies. The T7-Tbx6 fusion protein was immunoprecipitated by both antibodies, resulting in the appearance of one major band at approximately 60 kDa. Neither Tbx6 antibody immunoprecipitated the luciferase control. For each gel, 10 μL of lysate was loaded in lysate lanes, 20 μL of lysate was used in immunoprecipitation reactions, and one quarter of the final volume of the immunoprecipitation reaction was loaded in the immunoprecipitation lanes.

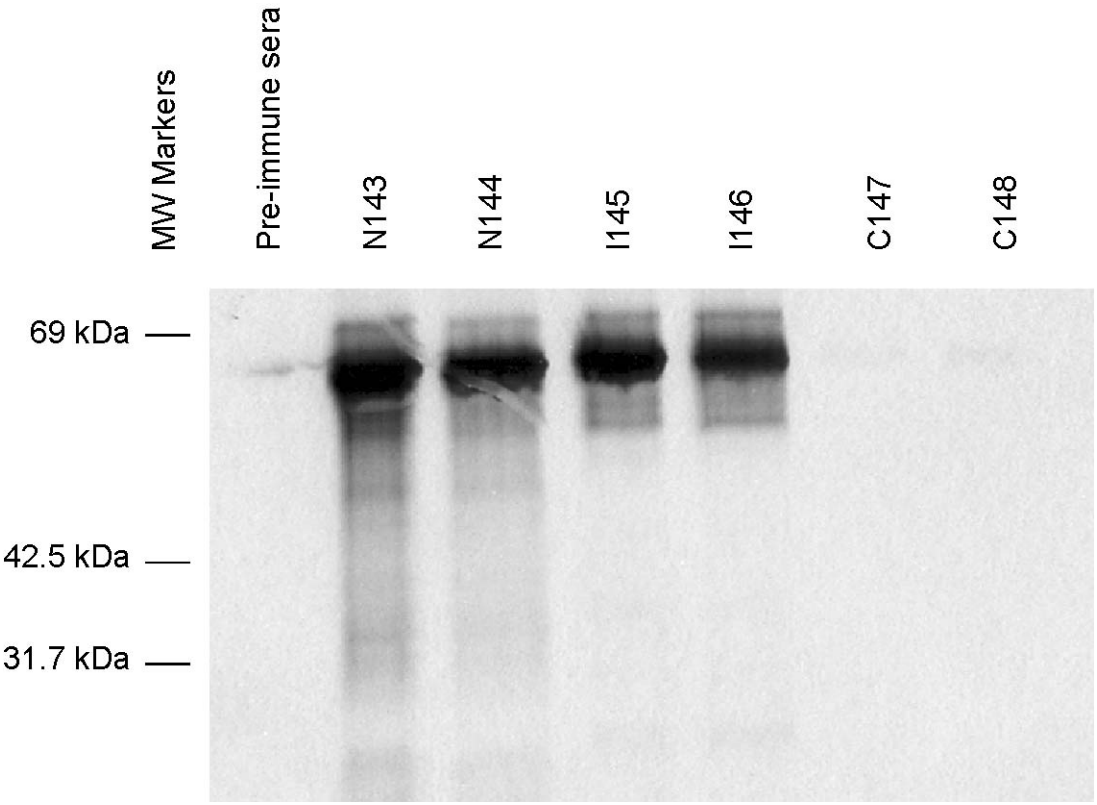


Figure 2-5: Tbx6-specific antibodies were used for immunoprecipitation of Tbx6 *in vitro*

Translated, radiolabelled Tbx6 was used for immunoprecipitation with antisera from the third test bleed of six rabbits used in the generation of Tbx6-specific antibodies. Protein was removed from beads and separated on 12% SDS-polyacrylamide gels. After 24 hours exposure, one major protein band was detected at approximately 60 kDa for both rabbits inoculated with the soluble GST-NTbx6 and GST-ITbx6 antigens, however the use of antisera from the insoluble GST-CTbx6 antigens did not result in immunoprecipitation of radiolabelled Tbx6. The use of pre-immune sera from the source animal used for generation of anti-T6N143 served as a negative control in immunoprecipitation reactions.

2.4 DURING MESODERM FORMATION, TBX6 EXISTS PREDOMINANTLY AS A 58 KD PROTEIN

Mouse *Tbx6* functions in the specification and the patterning of presomitic mesoderm (Chapman and Papaioannou, 1998; White et al., 2003). Using *in situ* hybridization, *Tbx6* transcripts were previously localized to the primitive streak and tailbud, the sources of mesoderm and the presomitic mesoderm (Chapman et al., 1996a). *Tbx6* expression was down regulated as this paraxial mesoderm underwent segmentation to form the somites, meaning that *Tbx6* had a limited window of transcriptional expression, commencing at e7.0 and continuing until e12.5 (Chapman et al., 1996a). In the mouse, by e10.5, the embryo had already formed approximately 30 somites; continued extension of the embryo axis occurred via the tailbud where *Tbx6* was expressed. This relatively narrow window of expression was particularly intriguing and suggested that *Tbx6* was tightly regulated to maintain its mRNA in the spatial and temporal pattern. Because nothing was known about protein expression *in vivo*, we assumed that Tbx6 protein was expressed in a similar pattern to its mRNA. Therefore, the primitive streak/tail bud would be an excellent source of Tbx6 protein.

To determine the molecular mass of the endogenous Tbx6 protein and whether any differentially processed versions existed, protein lysates were prepared from e10.5 tailbud tissue and from caudal somitic tissue (known *Tbx6* mRNA negative cells). Briefly, these samples were dissected, lysed, and portions of the lysates were separated on SDS-polyacrylamide gels and immunoblotted with the N-terminal and Internal Tbx6 polyclonal antibodies. Both Tbx6 antibodies detected a single band at approximately 58 kDa in the tailbud sample, but not in the caudal somite tissue as expected from transcriptional data for *Tbx6* expression (Figure 2-6). A commercially available actin antibody was used as a positive loading control. Since both Tbx6 antibodies detected a single protein band within the tail bud samples only, we concluded that Tbx6 expression was concentrated in tailbud tissue but not caudal somitic tissue, as we expected based on previous *in vivo* mRNA expression results. In addition, we concluded for the tissues investigated that there were no other forms of Tbx6 *in vivo* and the predominant form of Tbx6 was the 58 kDa protein originally predicted by Chapman et al. (1996). These conclusions were made only for the expression of Tbx6 at e10.5 in the tail bud, because other forms may exist at different stages of development or in adult tissue.

2.5 TBX6 PROTEIN LOCALIZES TO THE PRIMITIVE STREAK AND PARAXIAL MESODERM DURING MOUSE EMBRYOGENESIS

Tbx6 mRNA localized to the PS and the PSM beginning at e7.0 and later to the tailbud at e10.5 until e12.5 (Chapman et al., 1996a). Therefore, we performed whole mount immunocytochemistry using the N-terminal (anti-T6N144) and Internal (anti-T6I146) antibodies to localize Tbx6 protein between e7.5 and e12.5 of development. Both antibodies detected Tbx6 protein in the PS and migrating wings of PAM at e7.5, and in the PS and PSM at e8.5 and e9.5 (Figure 2f-h, k-l). At e10.5, Tbx6 protein localized to the tailbud, which took over as the source of mesoderm from the PS (Figure 2i&j). After e12.5, the tailbud was no longer capable of generating somites and, at this time, Tbx6 protein was barely detectable (data not shown). Preimmune sera was used as a negative control and showed no specific staining in the embryos (data not shown). In addition to highlighting the specificity of the antibodies, these results demonstrated that the Tbx6 protein was localized to the same tissue as the mRNA.

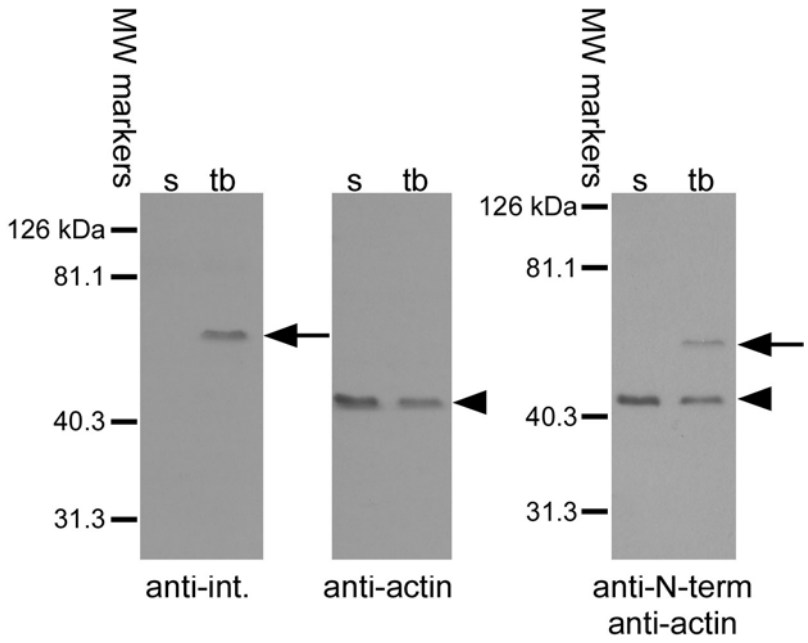


Figure 2-6: Western blot analysis confirmed that Tbx6 existed primarily as a 58 kDa protein in presomitic mesoderm tissue

Figure published in White and Chapman (2005). Protein lysates prepared from e10.5 tailbuds (tb) and caudal somites (s) were resolved by SDS-PAGE and subjected to Western blot analysis using the N-terminal (anti-T6N143), Internal (anti-T6I145), and anti-actin antibodies. Both Tbx6 antibodies detected a single band at approximately 58 kDa in the tailbud samples (arrow), but not the caudal somite samples. The predicted molecular masses of the Tbx6 signals were determined by the distances traveled through the gels in relation to prestained protein markers. The Western blot probed with the anti-Internal antibody was stripped and subsequently probed with the anti-actin antibody, while the N-terminal and anti-actin antibodies were used simultaneously. The anti-actin antibody detected a 42 kDa band in both the tailbud and somite samples (arrowhead) and served as both a loading control and a known protein molecular mass reference.

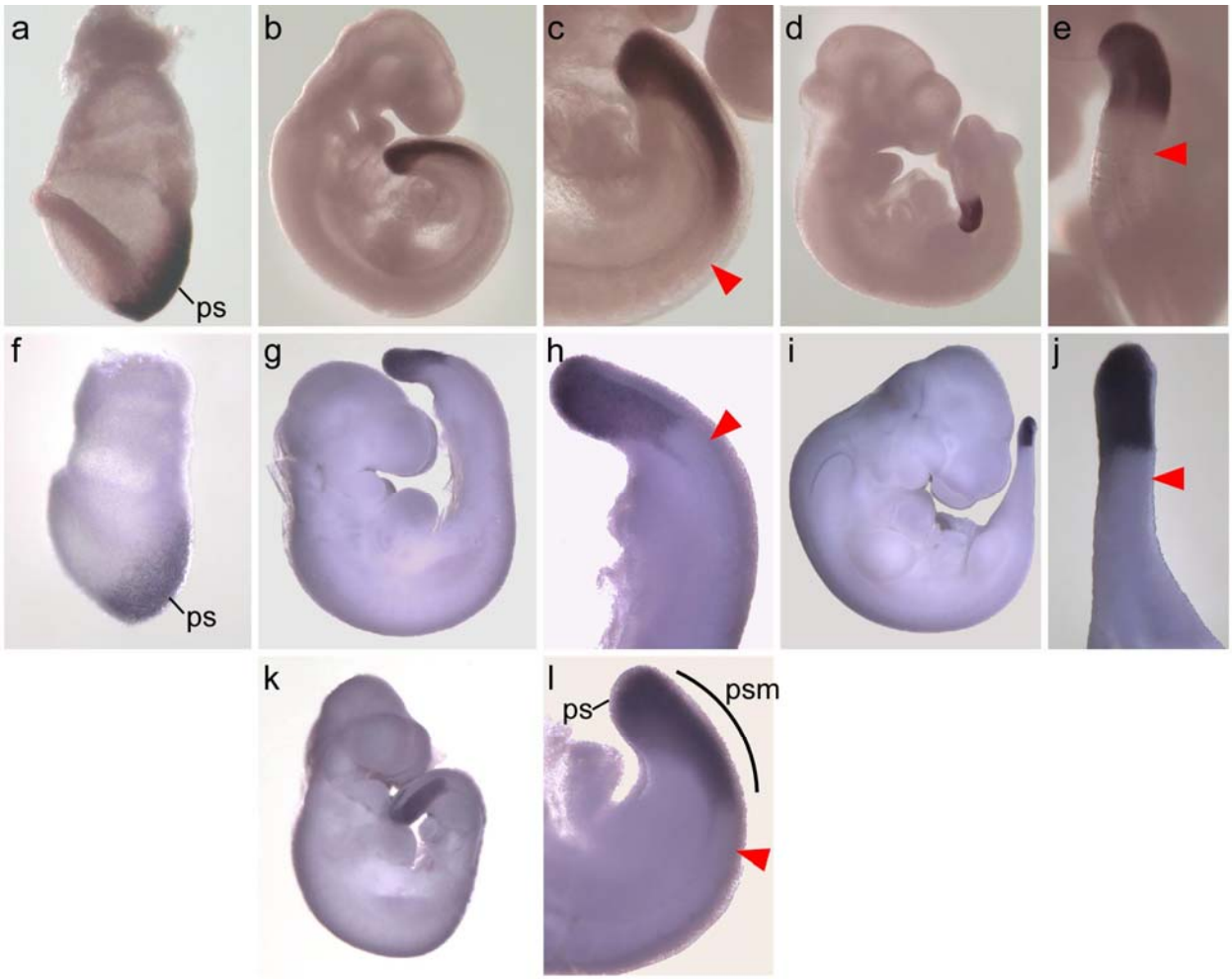


Figure 2-7: Localization of Tbx6 to the primitive streak and presomitic mesoderm

Figure published in White and Chapman (2005). Localization of *Tbx6* mRNA by whole mount *in situ* hybridization (**a-e**) and Tbx6 protein by whole mount immunocytochemistry using the Internal (**f-j**) and N-terminal (**k-l**) Tbx6 antibodies. *Tbx6* mRNA was localized to the primitive streak (PS) and presomitic mesoderm (PSM) at e8.0 (a.), the PS (located at the tip of the tail) and PSM at e9.5 (b., high magnification shown in c.), and in the tailbud at e10.5 (d., high magnification shown in e.). The Internal antibody detected Tbx6 protein in the PS and migrating wings of paraxial mesoderm (PAM) at e7.5 (f.), the PS and PSM at e9.5 (g-h.), and the tailbud at e10.5 (i-j.). The N-terminal antibody detected Tbx6 protein in the PS and PSM at e9.5 (k-l.). (h, j, l.) High magnifications of e9.5 and e10.5 tail regions with Tbx6 protein localized to the PSM but not in the formed somites. The red arrowheads (c, e, h, j, l.) indicate the boundary between the PSM and the somitic mesoderm.

2.6 DISCUSSION

2.6.1 Generation of Tbx6-specific antibodies

In all, six different rabbits were inoculated with three different antigens. The antisera from two rabbits inoculated with GST-NTbx6 (anti-T6N143, anti-T6N144) detected Tbx6, as both antisera specifically detected Tbx6 in an immunoprecipitation reaction (Figure 2-4), in addition to Western blot and immunohistochemistry. The same occurred for the antisera from the two rabbits inoculated with GST-ITbx6 (anti-T6I145, anti-T6I146). However, the antisera from the two rabbits inoculated with the insoluble GST-CTbx6 (anti-T6C147, anti-T6C148) failed to immunoprecipitate Tbx6 and were not used in further experiments. These N-terminal and Internal Tbx6 antibodies were used interchangeably throughout later experiments in this thesis, and they will be referred to simply as N-terminal and Internal Tbx6 antibodies since there have been no differences observed between the two antisera of each antigen.

2.6.2 Tbx6 exists primarily as a 58 kDa protein in the presomitic mesoderm

The Tbx6 antibodies recognized a single 58 kDa protein in lysates prepared from e10.5 tailbuds (Figure 2-6), suggesting either that a single Tbx6 protein was made *in vivo* or that if different isoforms existed, they generated either proteins of similar molecular mass or proteins at low levels and were below the level of detection using Western blot analysis. Aside from the original *Tbx6* cDNA sequence (Accession # U57331) from Chapman, et al. (1996a), other *Tbx6* cDNA sequences have also been deposited at Genbank and other sequencing centers, such as Ensembl, that were predicted to generate proteins of approximately 47 kDa; this included the human *TBX6* cDNA. Although the N-terminal antibody would detect these smaller proteins, we did not observe a 47 kDa protein in the tailbud lysate, which suggested a single form of Tbx6 existed in the tailbud tissue. This does not rule out other isoforms at different stages or in different tissues, however we currently believe that these differences were due to errors in sequence reads when using PCR-based sequencing, as was used at most sequencing facilities.

A shortened protein form of Tbx6 existed in sequencing databases. However there was no published visual record of this protein in the literature. Although results were not provided

here, initial attempts to clone *Tbx6* using PCR into expression vectors failed because the PCR products generated from the original *Tbx6* cDNA sequence (Accession # U57331) were mutated, introducing the same frameshift mutation in a high GC region of *Tbx6* that was reported by most sequencing facilities; these constructs generated a truncated, radiolabelled protein when *in vitro* translated. However, as described in Chapter 2, using restriction digests to clone Tbx6 from the original *Tbx6* cDNA source construct resulted in the expression of a 58 kDa protein. These results provided more evidence that this alternative Tbx6 protein was not real, but rather was an error generated during the process of dye termination sequencing/PCR.

This small detail has major impacts when studying a protein for which minimal information is known. How do you start to study a protein when you do not even know the real sequence? How do you know the protein you express in a laboratory actually models the protein *in vivo*? This subtle difference is important because sequencing centers use the cDNA of characterized animals (often the mouse) to predict the coding regions of other genomes, namely human, rat, chimpanzee, etc; if a starting sequence is wrong, then all predicted sequences are wrong. Secondly, removing the C-terminal half of Tbx5 (AA 238 – 518) significantly enhanced the protein's DNA binding affinity compared to full length Tbx5 (Ghosh et al., 2001). This suggests that truncated proteins may function differently than full length protein; in this case, the truncated protein interfered with DNA binding, and may affect other DNA binding proteins that function within the same region(s) to ultimately change the expression of downstream targets. The Tbx6 protein we are using for our experiments in this thesis is the same molecular mass as the protein found *in vivo*.

Using computer predictions, there are two predicted motifs within Tbx6, namely a PEST sequence at AA 320-331 and a leucine zipper motif at AA 436-458 (Hulo et al., 2006; Rechsteiner and Rogers, 1996). These motifs are further described in Section 5.1.1, however the smaller Tbx6 variant results in changes to these motifs. The smaller Tbx6 variant, believed by others to be the true Tbx6 sequence, not only truncates the protein at AA 436, but also has protein sequence differences starting at AA 330 (Figure 2-1). The Tbx6 short variant contains a potential PEST sequence from AA 320 to 351, which is expanded in comparison to the large Tbx6 variant, but contains no leucine zipper motif. If the Tbx6 short variant is a bona fide protein, this suggests that it may be much less stable and be degraded more rapidly than the larger 58 kDa Tbx6. Further, since the C-termini of the proteins vary so significantly, the Tbx6

short variant could serve a very different function either because it is missing the leucine zipper (if this motif proves real) or because the activation/depression domain of the T-box proteins are often found in the C-terminus (Carlson et al., 2001; Kispert et al., 1995a). Again, these are interesting projects to pursue.

2.6.3 Conclusion

We have now successfully generated antibodies against two non-conserved regions of the Tbx6 protein. Altogether, our results indicated that the Tbx6-specific antibodies were useful for Western blots, immunoprecipitation of *in vitro* translated Tbx6 protein, and confirmed that the major endogenous Tbx6 protein in tailbud tissue matched the predicted 58 kDa molecular mass. We also used these antibodies to localize the protein in embryos during development. Both the N-terminal and Internal antibodies stained the PSM and tailbud, but not the segmented PAM. This co-localization of the Tbx6 protein with the *Tbx6* mRNA suggested that production of the protein itself was also tightly regulated during development.

3.0 THE T-BOX TRANSCRIPTION FACTOR, TBX6, PREFERENTIALLY RECOGNIZES THE T-BOX BINDING SITE

3.1 INTRODUCTION

The T-box genes were named after the founding member, *T* or *Brachyury*, which was originally identified as a mouse mutant that displayed a short tail (Dobrovolskaia-Zavadskaia, 1927). At the protein sequence level, the only region of homology among all family members is the approximately 150 AA region known as the T-box domain that encodes a DNA binding domain. Initial binding site studies showed that T bound to a 24 base pair, almost perfect palindromic sequence, and later studies demonstrated that T also bound to repeated half-sites (Casey et al., 1998; Kispert and Herrmann, 1993; Sinha et al., 2000). Given that the T-box conserved region encoded the DNA binding domain, it was not surprising that many of the characterized T-box protein binding sites were also somewhat conserved. Results from several laboratories showed that the T-box proteins bound to a core 5'-AGGTGT-3' with sequences flanking this core sequence being more variable (Conlon et al., 2001; Kispert and Herrmann, 1993; Sinha et al., 2000; Tada et al., 1998). Many of the T-box proteins bound as monomers to the half-site, however other T-box proteins, specifically T and human TBX1, bound as true dimers to the palindromic site (Conlon et al., 2001; Kispert et al., 1995a; Muller and Herrmann, 1997; Sinha et al., 2000; Tada et al., 1998). Spacing and orientation of the half-sites was also critical for determining binding site specificity of the different T-box proteins (Conlon et al., 2001).

3.1.1 Aims of this study

Because little was known of the binding properties for Tbx6, I sought to determine if the murine Tbx6 recognized the *in vitro* consensus palindromic and half-site binding sequences identified

for T. After establishing that Tbx6 bound both sequences, I determined a consensus binding site sequence for Tbx6 using a PCR-based binding site selection procedure. Next, I expressed Tbx6 transiently in tissue culture cells to determine the effect of Tbx6 on a luciferase reporter gene under the control of a minimal promoter, while containing an enhancer that included the various forms of the Tbx6 binding site(s). Lastly, I sought to determine the effect of murine T and Tbx6 together on the luciferase expression system. Specifically, I wanted to test whether T and Tbx6, which are coexpressed in the PS, interacted *in vivo* to affect transcription of downstream targets.

3.2 TBX6 RECOGNIZES THE BRACHYURY CONSENSUS SEQUENCE AND HALF-SITE SEQUENCE *IN VITRO*

An *in vitro* consensus binding site for T had been identified as a 24 base pair, almost perfect palindromic sequence (Kispert and Herrmann, 1993), however T as well as other T-box proteins also bound to the half-site sequence to activate reporter constructs *in vivo*. While some T-box proteins bound to the palindromic sequence as separate monomers *in vitro*, others required dimerization, such as T and TBX1 (Coll et al., 2002; Muller and Herrmann, 1997; Sinha et al., 2000). Electrophoretic mobility shift assays (EMSAs) are the primary technique used to establish whether proteins bind to specific target sequences. In particular for T-box proteins, EMSAs can be used to determine whether the protein binds the T palindromic binding site as a dimer, like T, or as a monomer. Proteins that bind as monomers will shift both the half-site and palindromic sequence probes, whereas proteins that bind as true dimers will shift only the T palindromic sequence, because they require the presence of two 'half-site' sequences. Therefore, I tested whether Tbx6 could recognize the T palindromic and half-site sequences. This work was published in White and Chapman (2005).

I created double-stranded, radioactive probes from oligonucleotides (oligos) based upon both the palindromic site and half-site of T and used these for EMSAs (Figure 3-1a). An *in vitro* translated T7-tagged Tbx6 protein was used for all EMSAs and was generated using the same construct previously described for the immunoprecipitation of radiolabelled Tbx6 (Section 2.3).

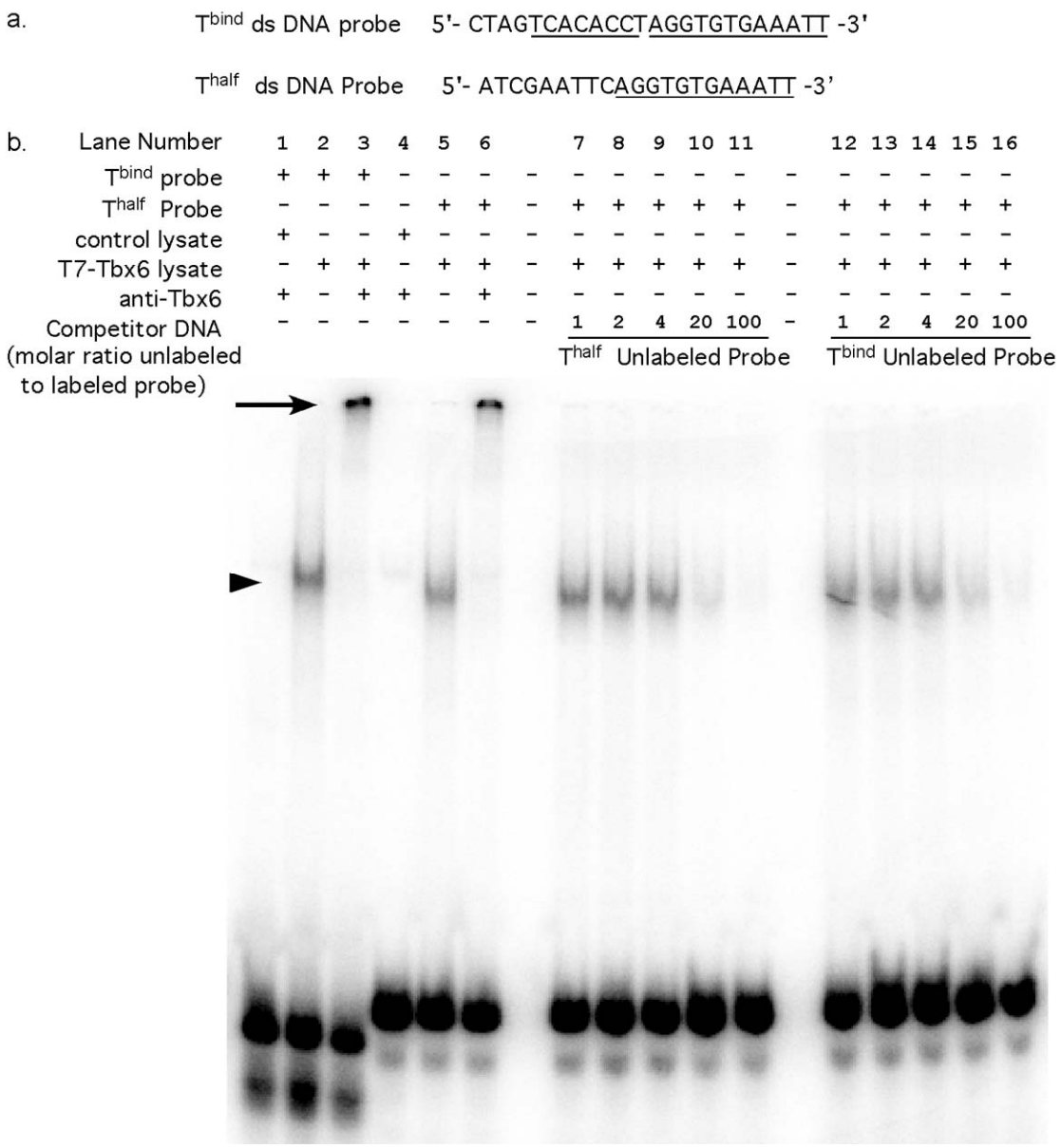


Figure 3-1: Tbx6 recognizes the Brachyury consensus sequence

Figure modified from White and Chapman (2005). (**a**) DNA probe sequences for the T consensus site (T^{bind}) and the half-site (T^{half}). The T^{bind} site is an almost perfect palindrome of the T^{half} site. Sequences matching the T^{half} site are underlined. (b) *In vitro* translated T7-Tbx6 or unprogrammed control lysates were used for EMSAs of the T^{bind} and the T^{half} probes. One major band was observed for both T^{bind} and T^{half} probes (arrowhead, lanes 2,5). Super-shifts of the protein/DNA complex with α-N-terminal Tbx6 antibodies to the top of the wells (arrow) confirmed that binding was Tbx6-specific in lanes 3,6. Increasing concentrations of unlabeled T^{half} (lanes 7-11) or T^{bind} (lanes 12-16) probes effectively competed binding of Tbx6 to the T^{half} probe at a 1:100 molar ratio in a similar manner.

Tbx6 bound both the T consensus site (T^{bind}) and the T consensus half-site (T^{half}) by EMSA (lanes 2,5 in Figure 3-1b). This binding was specific to Tbx6, because no binding was observed in a control lysate and the binding occurred in the presence of a non-specific competitor (lanes 1,4 in Figure 3-1b and Methods). Addition of the anti-Tbx6 N-terminal antibody resulted in supershifting of the protein/DNA complexes to the top of the gel and confirmed that Tbx6 was specifically responsible for the observed band (lanes 3,6 in Figure 3-1b). It should be noted that both purified N-terminal antibodies (anti-T6N143, anti-T6N144) and purified internal antibodies (anti-T6I145, anti-T6I146) were able to supershift the Tbx6/DNA complexes in EMSAs (Figure 3-2). Furthermore, Tbx6 binding could be competed by the addition of increasing concentrations of unlabeled palindromic and half-site probes (lanes 7-16 in Figure 3-1b). Competition was observed by the addition of 100-fold molar excess of cold probe, regardless of whether the palindromic or half-site had been used in the reactions.

These experiments in Figure 3-1b used a low molar concentration of Tbx6 protein relative to DNA probe. If the palindromic sequence were in fact being recognized by multiple Tbx6 subunits, either the band in the palindromic probe lane (lane 2) would be of slower mobility than in the half-site lane (lane 5) as multiple Tbx6 subunits would bind the palindromic site, or multiple bands would exist in the palindromic probe lanes. However, this was not the case for the experiments in Figure 3-1b as the T palindromic and half-site probes resulted in bands at approximately the same location (lanes 2,5), suggesting instead that only one subunit of Tbx6 was recognizing the palindromic sequence. Therefore, only one subunit of Tbx6 must be bound to the DNA probes in all of the reactions. This was not the case for the experiment depicted in Figure 3-2, as two bands of different mobilities appeared when these experiments were performed with higher levels of Tbx6; therefore, under certain conditions, Tbx6 was capable of binding both 'half sites' of the palindromic probe.

Experiments were designed to vary the Tbx6 protein concentration (with a constant DNA probe concentration) to determine whether recognition of the two half-sites within the palindromic probe were dependent upon Tbx6 protein concentration. As shown in Figure 3-3, a slower mobility band formed when increasing concentrations of Tbx6 were added to a limited

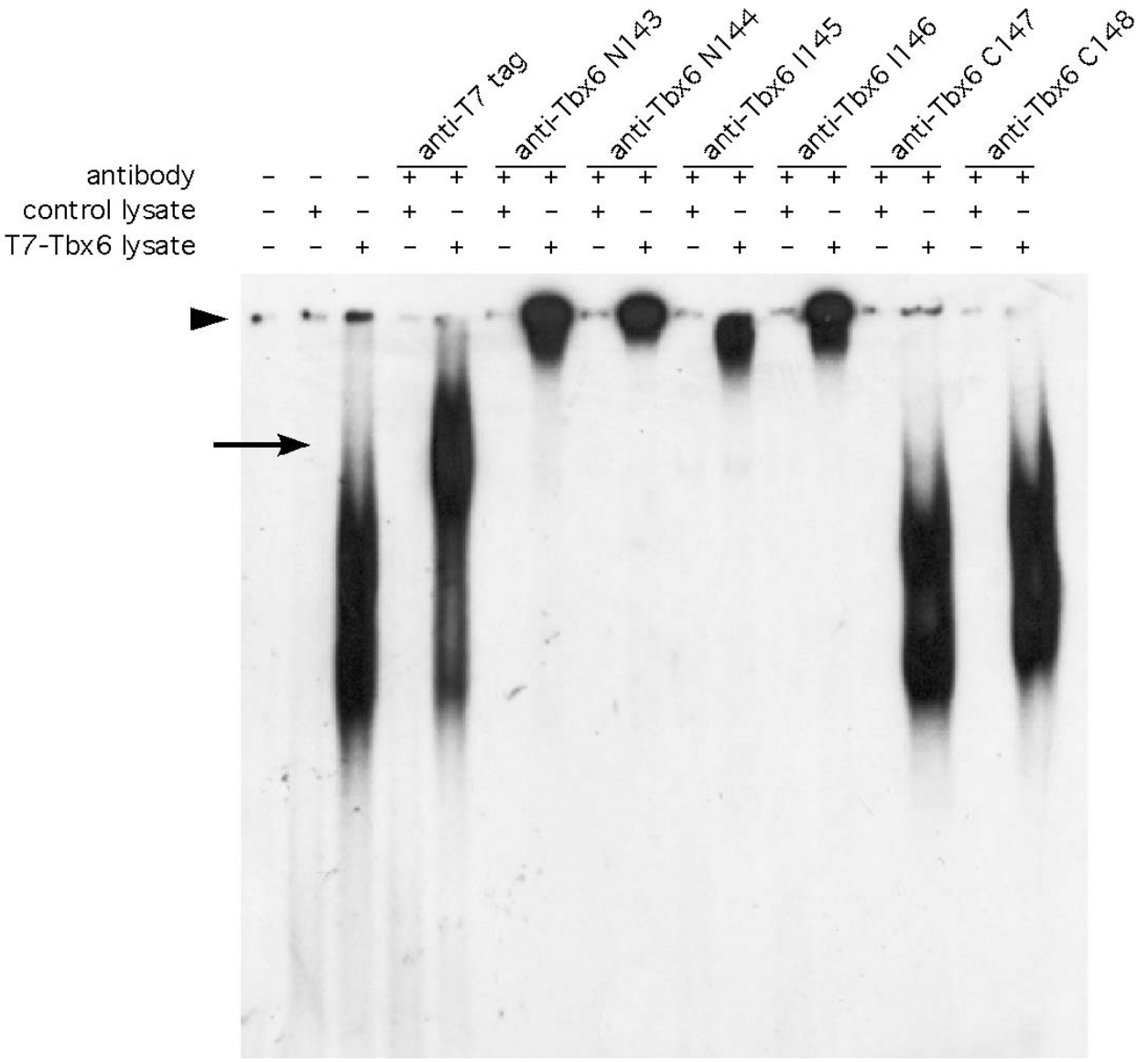


Figure 3-2: Tbx6 N-terminal and internal antibodies recognize Tbx6 in EMSAs

In vitro translated T7-Tbx6 or unprogrammed control lysates were used for EMSAs of the T palindromic consensus sequence probe. The addition of various antibodies to the EMSA reactions resulted in the supershifting of the protein/DNA complex to the top of the wells (arrowhead) for both N-terminal Tbx6 antibodies (N143, N144) and for both internal Tbx6 antibodies (I145, I146). However, no supershift was observed using two antisera raised against the C-terminus of Tbx6 (C147, C148). A supershift of the protein/DNA complex was also observed using the monoclonal anti-T7 antibody that detected the tag on the T7-Tbx6 protein (arrow).

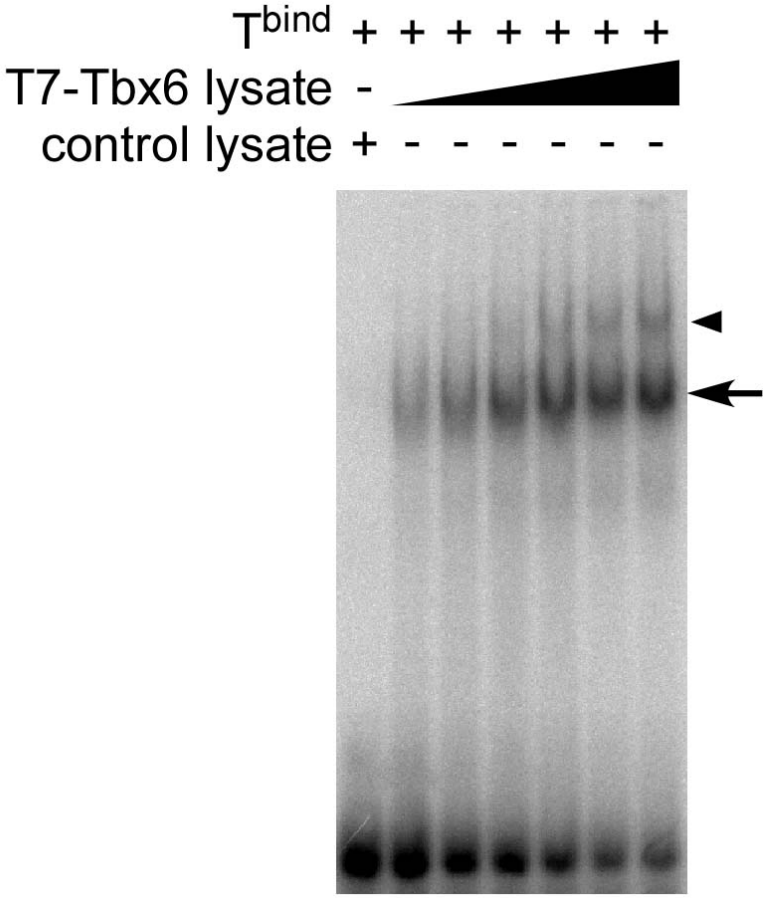


Figure 3-3: Tbx6 can bind to the T palindromic sequence with one or two subunits

Figure modified from White and Chapman (2005). EMSA using the T^{bind} probe with increasing concentrations of Tbx6 protein showed the presence of a second slower migrating band (arrowhead) in addition to the faster migrating band (arrow); these data demonstrated that Tbx6 recognized the two half-sites sequences in a concentration dependent manner.

supply of probe (arrowhead). Thus, these results suggested that Tbx6 bound readily to the half-site sequence, and multiple Tbx6 subunits recognized the palindromic site at the same time in a concentration dependent manner *in vitro*.

Whether these subunits are actually true monomers of Tbx6 protein is unknown. All T-box proteins were classified by the conserved T-box DNA binding domain, and multiple T-box

protein studies suggested that the specificity of T-box proteins for their DNA targets was based upon the sequence of the T-box domain (Coll et al., 2002; Conlon et al., 2001; Muller and Herrmann, 1997; Sinha et al., 2000). Crystal structure analysis of T had shown that T was a monomer in solution and formed a true dimer only after a single protein first bound to the DNA target. T had never been detected by EMSAs to bind a single half-site sequence; this suggested that dimerization was necessary for maintaining stable binding (Muller and Herrmann, 1997). On the other hand, the crystal structure for Tbx3 revealed that two Tbx3 proteins bound to the palindromic sequence, however the lack of interaction between the two monomers suggested that Tbx3 bound as two independent monomers. Tbx3 had not been tested to bind half-site probes in EMSA experiments, but had been shown to bind palindromic sequences in EMSA experiments, resulting in two distinct bands much like that seen in Figure 3-2 and Figure 3-3 (Carlson et al., 2001; Coll et al., 2002). Therefore, I predict that the T-box of Tbx6, like Tbx3, binds DNA as a monomer and Tbx6 does not require dimerization to bind the palindromic T binding site sequence, unlike T that forms a true dimer upon binding its palindromic sequence.

3.3 IDENTIFICATION OF AN *IN VITRO* TBX6 BINDING SITE

Although the above studies showed that Tbx6 bound the T consensus and half-site sequences *in vitro*, it was possible that Tbx6 could bind to other sequences as well. To identify an *in vitro* consensus binding site sequence for Tbx6, we performed a PCR-based binding site selection similar to that described for identification of the T binding site (Kispert and Herrmann, 1993). This work was published in White and Chapman (2005). Briefly, this procedure consisted of incubating an *in vitro* translated, non-radioactive T7-Tbx6 protein with a 76mer that contained a random mix of 26 nucleotides flanked by two known sequences that (1) were recognized by PCR primers for amplification of selected sequences, and (2) failed to contain any sequences similar to the T consensus sequence (see Figure 3-4). Tbx6 only bound to a subset of those random sequences, i.e. those that matched Tbx6-specific binding sites. Tbx6/DNA complexes were isolated using immunoprecipitation to remove non-specific sequences while enriching for Tbx6-specific sequences. Although none of the antibodies recognized the T-box domain, it was possible that a specific antibody could interfere with the ability of Tbx6 to bind target sequences.

To rule out any sequence bias due to the antibodies used, we performed the selection scheme using three different antibodies: the T7 monoclonal antibody that recognized the T7-tag of the T7-Tbx6 fusion protein, and the N-terminal and Internal Tbx6 polyclonal antibodies. After immunoprecipitation of Tbx6/DNA complexes, the DNA was then isolated and subjected to PCR, amplifying resulting sequences for further selection. The cycle was repeated, and after each round of selection, a portion of resulting sequences were used to further enrich for Tbx6-specific sequences. To monitor the efficiency of the selection scheme, EMSAs were used to ensure that the pool of selected sequences truly represented sequences that Tbx6 bound *in vitro*. For these EMSA experiments, a portion of the selected sequences were radiolabelled and used as a probe; shifting of the protein/DNA complexes indicated that the selected sequences must have contained some sequences that were specific to Tbx6.

By the sixth round for the T7 monoclonal antibody (Figure 3-5) and the fourth round of selection for both polyclonal antibodies (Figure 3-6), a band was observed in the T7-Tbx6 programmed lysate that was not observed in the unprogrammed lysate. Specificity of binding was examined by adding a Tbx6-specific antibody that supershifted the Tbx6/DNA complexes (arrowheads in Figure 3-5, 3-6). These types of experiments had been performed for several T-box proteins, and similar to my experiments, four to six rounds of selection were required to obtain specific shifts of protein/DNA complexes on EMSAs (Conlon et al., 2001; Ghosh et al., 2001; Kispert and Herrmann, 1993). All three binding site selection experiments had between 30-40 individual clones sequenced to determine a consensus binding site. A majority of the sequences (86 of 97) were aligned around the conserved 5'- GGTGT -3' sequence also present in the T half-site. As shown in Table 3-1, these aligned sequences revealed a consensus binding site of 5'-AGGTGTBRNNNN-3' for Tbx6, where position 7 was T, G or C, but not A (designated as B) and position 8 was A or G (designated as R). Interestingly, palindromic sequences and sequences exactly matching the T half-site (5'-AGGTGTGAAATT-3') were not obtained. However, eleven of the 97 sequences had an insertion or deletion of one nucleotide between positions 1 and 4, resulting in sequences that were only very weakly shifted in EMSAs with Tbx6, in comparison to the T half-site sequence, and was therefore not included in determining a consensus sequence (data not shown).

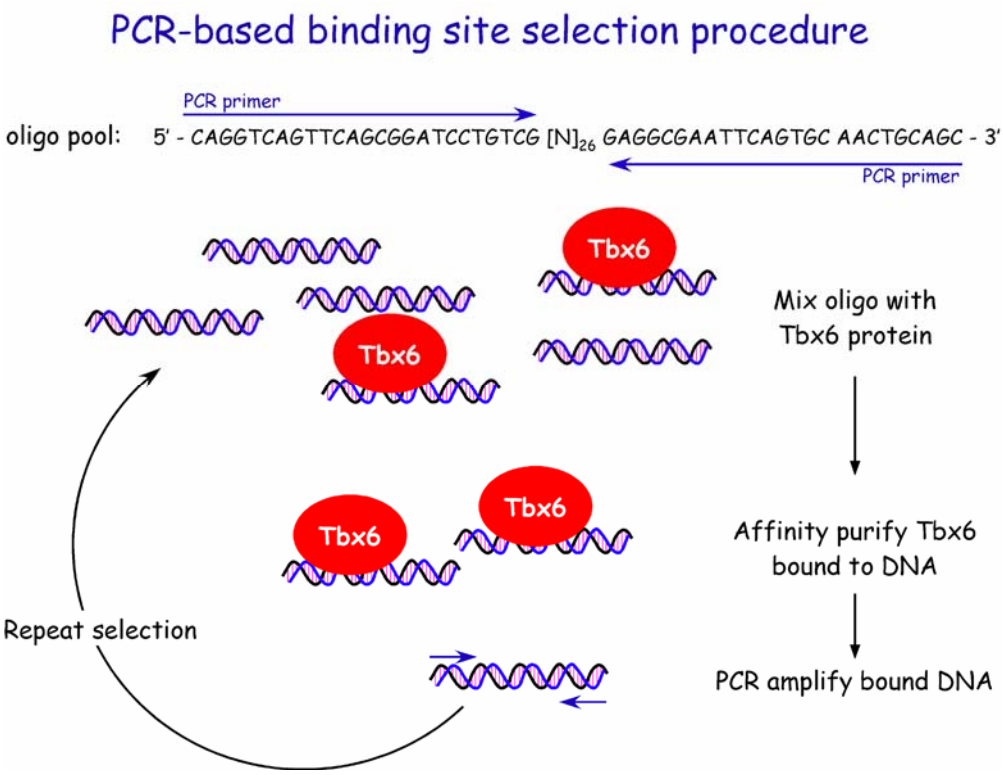


Figure 3-4: Diagram of Tbx6 binding site selection procedure

To identify an *in vitro* binding site for Tbx6, we performed a PCR-based binding site selection similar to that described to identify a consensus binding site for T (Kispert and Herrmann, 1993). This procedure used cycles of immunoprecipitation and PCR to enrich for Tbx6-specific sequences. A double-stranded oligo, containing a random $[N]_{26}$ sequence flanked by known sequences for PCR amplification, was used as the probe. Sequences bound to Tbx6 were purified away from non-specific sequences, which were lost during the immunoprecipitation. This cycle was repeated until Tbx6 protein/DNA complexes were detected using EMSAs.

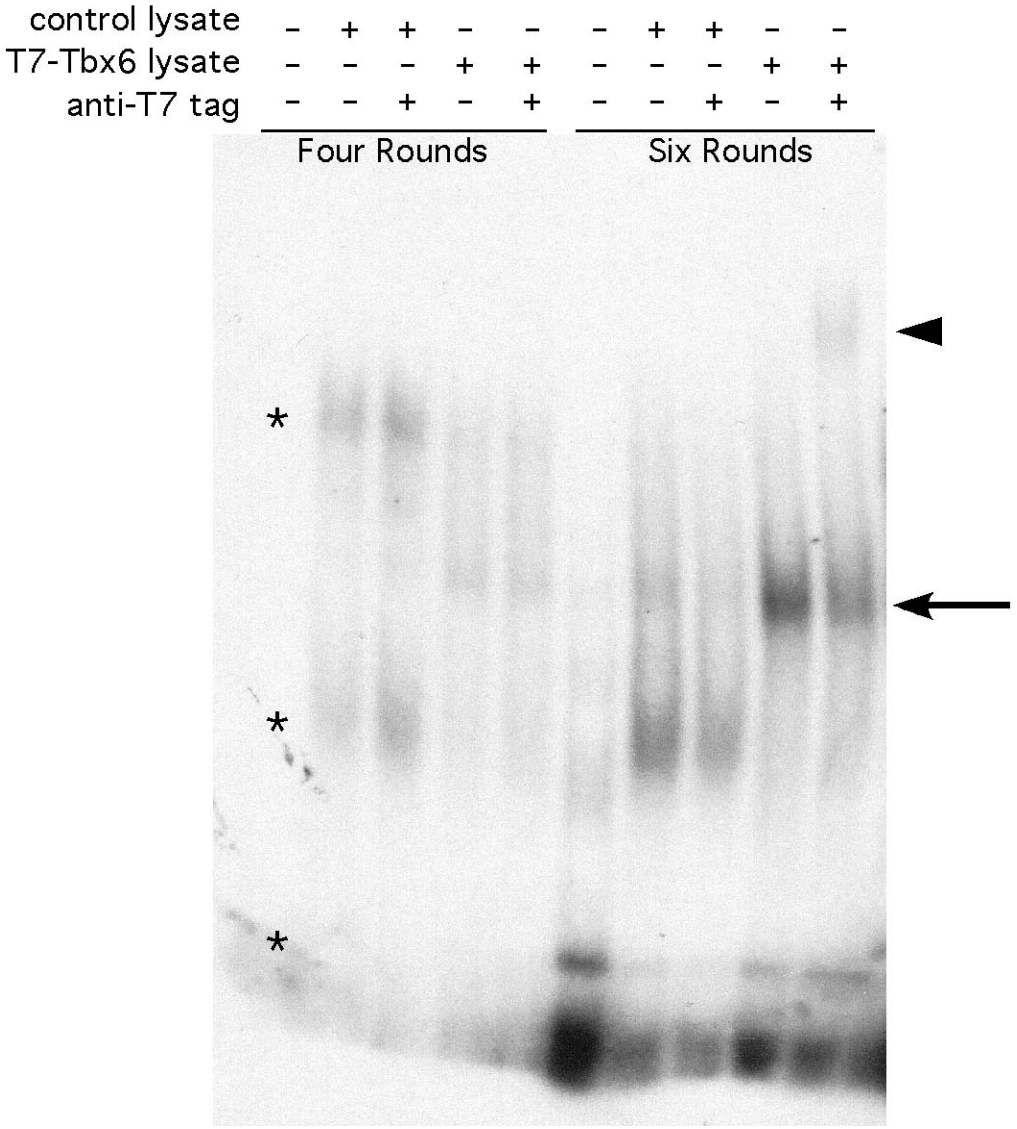


Figure 3-5: Representative EMSA showing the enrichment for Tbx6 sequences during the binding site selection procedure using the T7 monoclonal antibody.

In vitro translated T7-Tbx6 or unprogrammed control lysates were used for EMSAs of the enriched Tbx6-specific sequences using the T7 monoclonal antibody. Non-specific binding occurred in the control lysate lanes after four and six rounds of selection (asterisk, note bands present in control lysate that do not shift with antibody). Tbx6-specific bands were barely observed after four rounds of selection, but were apparent after six rounds of selection (arrow). Super-shifts of the Tbx6/DNA complexes with anti-T7 antibody resulted in the appearance of a slower mobility band (arrowhead), confirming that the band was due to Tbx6.

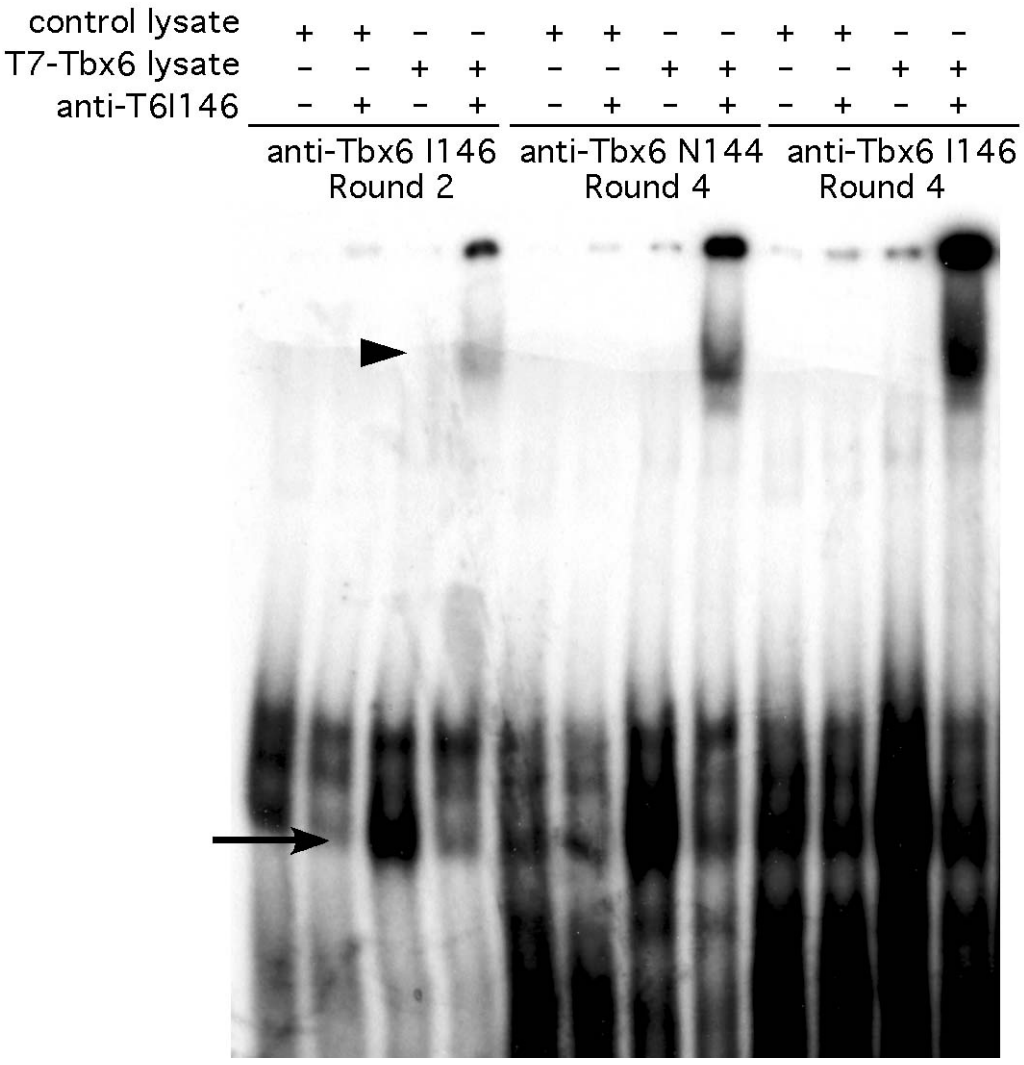


Figure 3-6: Representative EMSA showing the enrichment for Tbx6 sequences during the binding site selection procedure using the Tbx6 N-terminal and Internal polyclonal antibodies.

In vitro translated T7-Tbx6 or unprogrammed control lysates were used for EMSAs of the enriched Tbx6-specific sequences using the anti-T6N144 and anti-T6I146 antibodies. Non-specific binding occurred in all lysate lanes after two (I146) and four (N144, I146) rounds of selection. Tbx6-specific bands were barely observed after two rounds of selection for the Internal Tbx6 antibody, but were more apparent after four rounds of selection with both antibodies (arrow). The Tbx6/DNA complexes overlapped non-specific bands, however supershifts of the Tbx6/DNA complexes with anti-T6I146 purified antibody resulted in the appearance of slower mobility bands (arrowhead), confirming that the bands were due to Tbx6.

Table 3-1: Tbx6 consensus sequence as determined by total percentages of nucleotides found in selected sequences obtained through the binding site selection procedure

Position #	1	2	3	4	5	6	7	8	9	10	11	12
% Nucleotide A	100	4	0	0	0	2	3	72	49	41	27	31
% Nucleotide T	0	0	1	95	0	83	35	6	14	27	41	15
% Nucleotide C	0	2	0	5	0	14	26	0	16	14	17	19
% Nucleotide G	0	94	99	0	100	1	36	21	21	18	15	35
Tbx6 Consensus	A	G	G	T	G	T	B	R	N	N	N	N
T half-site	A	G	G	T	G	T	G	A	A	A	T	T
Tbx5 Consensus	R	G	G	T	G	T	B	R	N	N		

Binding site selection experiments identified a Tbx6 consensus binding site with a core sequence of 5'-AGGTGTBR-3' (White and Chapman, 2005). These findings were similar to those found for all of the T-box proteins thus far examined (Coll et al., 2002; Conlon et al., 2001; Ghosh et al., 2001; Hiroi et al., 2001). Tbx6 selected some preferences at particular positions, specifically selecting sequences containing anything but A at position 7; in addition, Tbx6 selected sequences containing A at position 1, like T, whereas Tbx5 bound to either A or G (Figure 3-1). Also, Tbx6 appeared to select sequences containing either A or G, selecting G approximately 20% of the time, similar to Tbx5 (Figure 3-1); on the other hand, T always selected sequences containing A at position 8 (Ghosh et al., 2001; Kispert and Herrmann, 1993). Outside of these 8 positions, Tbx6 enriched for sequences containing A at position 9 and 10, and for T at position 11, similar to T and Tbx5 (Figure 3-1). As discussed in the Introduction to this thesis, evolutionary alignments of the various T-box regions placed Tbx6 in its own subfamily, and this subfamily was most closely related to the Tbx2-Tbx5 subfamily (Section 1.8.2). The data presented in Section 3.1 to Section 3.3 supported the notion that the T-box of Tbx6 recognized DNA sequences as monomers, similar to Tbx3 and Tbx5.

3.4 EXPRESSION OF MURINE T AND TBX6 IN CELL CULTURE

When these experiments were being designed, there was little known about how Tbx6 functioned as a transcription factor. However, Brachyury (T) and its homologs in *Xenopus* and zebrafish had been extensively studied. Seminal work with T and its orthologs determined that T was a nuclear protein *in vivo*, and full-length T could activate a bacterial chloramphenicol acetyltransferase (CAT) reporter gene when co-expressed transiently in tissue culture (Cunliffe and Smith, 1994; Kispert and Herrmann, 1994; Kispert et al., 1995a; Schulte-Merker et al., 1992). Deletion analysis of T in tissue culture further showed that T contained two transactivating domains and two repression domains (Kispert et al., 1995a). Similar experiments have been successfully completed for human T, TBX1, TBX2, and TBX5, the *Xenopus* Xbra, Eomes and VegT, and murine Tbx3 (Carlson et al., 2001; Conlon et al., 2001; Sinha et al., 2000). More recently, interactions between several T-box proteins and other transcription factors had been examined primarily using transient reporter assays in tissue culture (Brown et al., 2005; Koshiba-Takeuchi et al., 2006). Therefore, I sought to express both Tbx6 and T in tissue culture to (1) confirm that full length Tbx6 and T were nuclear localized proteins *in vitro*, (2) determine how Tbx6 affected transcription of a luciferase reporter *in vitro*, and (3) determine if Tbx6 affected the ability of T in this *in vitro* luciferase assay system, and vice versa. The latter experiment was based on both the overlapping expression domains of Tbx6 and T in the embryo and the potential for these transcription factors to have similar target genes. All data presented in the remainder of Chapter 3 represent unpublished data.

3.4.1 Design and generation of T and Tbx6 cell culture expression vectors

As a first step towards achieving these goals, I generated cell culture expression vectors to express tagged versions of full-length murine T and Tbx6 for use in immunofluorescence and luciferase assays. To do so, I used PCR to amplify the full-length murine *T* cDNA. The construct was cloned into pET21a using engineered restriction sites creating the coding region for a protein containing a T7 N-terminal tag, the full-length *T* cDNA, and a C-terminal His tag; this cloned construct was sequenced and matched our original sequence for the *T* cDNA. The T7-tagged versions of T and Tbx6 were shuttled through intermediate vectors and finally cloned

into the tissue culture expression vector pcDNA 3.1/V5-His. Both constructs retained the N-terminal T7 tag, however the T construct also contained a C-terminal V5-His tag. The T7-Tbx6 was the same fusion protein used in previous experiments, however under the control of a tissue culture promoter (Figure 3-7a).

To verify that the tagged Tbx6 and T protein expression constructs produced proteins of the correct molecular mass, constructs were *in vitro* transcribed and translated in a rabbit lysate system in the presence of ^{35}S-methionine; unprogrammed lysate was used as a negative control. The molecular mass of the T protein was estimated by comparing the primary band against the known ~60 kDa T7-Tbx6 protein and prestained markers (Figure 3-7b). Both proteins with their respective tags exhibited a similar molecular mass. Computer predictions estimated a molecular mass for the T fusion protein to be approximately 53 kDa, which included the 437 AA T protein with a short 7 AA T7 tag and the V5 epitope and 6XHis C-terminal tag. The T protein produced was slightly larger (60 kDa) than the predicted 53 kDa. However proteins may have other modifications, such as phosphorylation and glycosylation, and other variations in apparent molecular mass had been reported on gels (Hames, 1998).

Ultimately, these constructs were designed for use in tissue culture, therefore the T construct was transfected into COS-7 cells and 293T cells to verify that the T fusion protein of predicted molecular mass was produced *in vivo*. 24 hours after transfection, cells were isolated, suspended in SDS-PAGE lysis buffer and boiled. Samples were separated on a 10% SDS-polyacrylamide gel, and were subjected to Western blot analysis using an anti-T polyclonal goat antibody (Santa Cruz Biotech) to detect T. The untransfected cell lysate lanes and the cells transfected with the Tbx6 fusion construct served as negative controls. Blots were stripped and probed with anti-actin as a loading control. The anti-T Western blot detected several bands, however only one of these bands was present solely in the T transfected lane (Figure 3-7c); the major protein produced was again approximately 55-60 kDa. The additional bands on the blot were likely due to non-specific binding of the T antibody to other proteins in the cell lysate because these bands were present in control lanes. Therefore, this construct produced protein that was approximately the molecular mass expected for the T fusion protein.

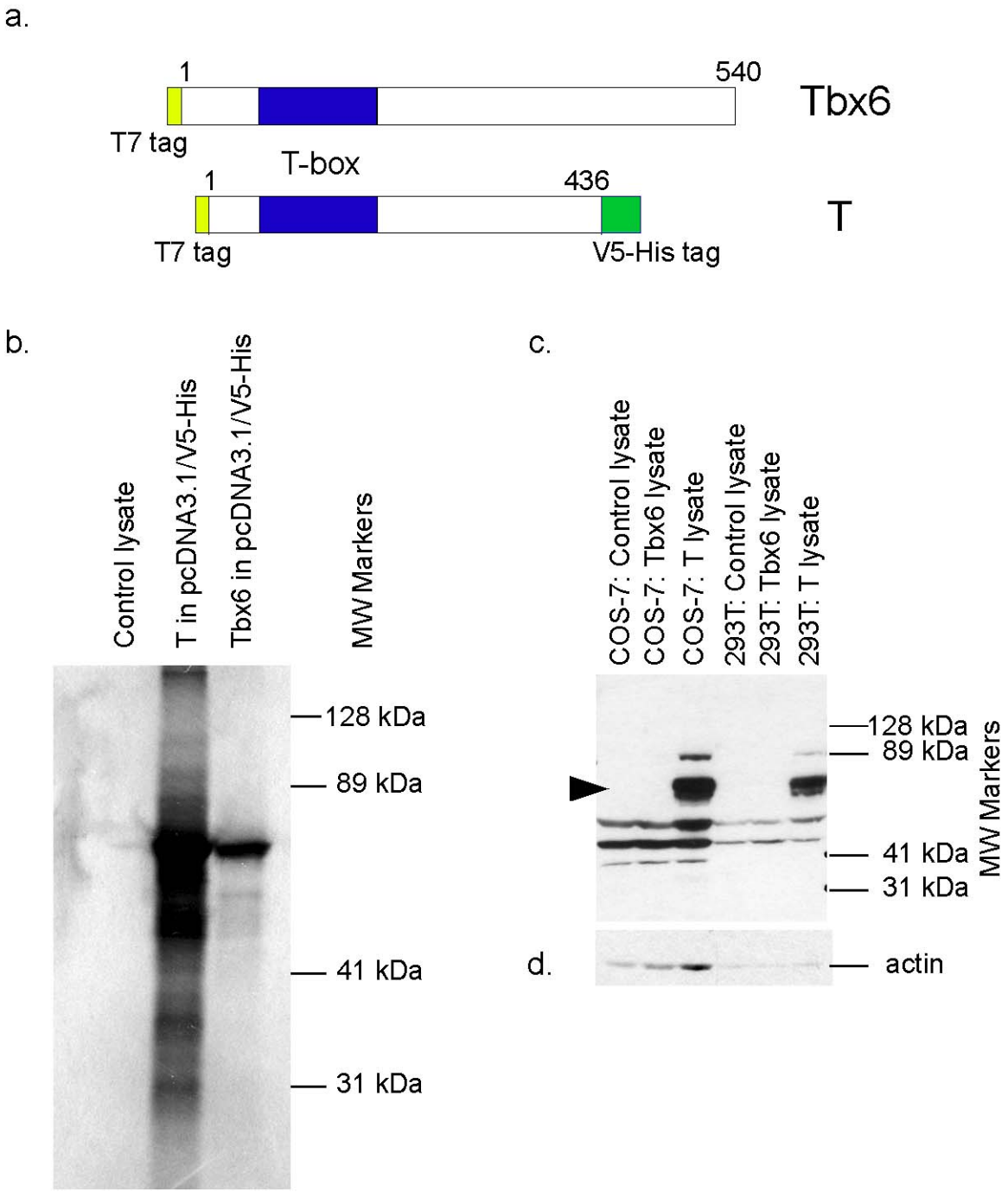


Figure 3-7: Generation of Tbx6 and T tissue culture expression constructs

(a) Schematic of the Tbx6 and T full-length fusion proteins for cell culture studies. Both Tbx6 and T contain a N-terminal T7 tag, while T also contained a C-terminal V5-His tag. (b) *In vitro* transcribed and translated T and Tbx6 proteins from the expression constructs; both proteins with tags were approximately 60 kDa. (c) Protein lysates from COS-7 and 293T cells transfected with Tbx6 and T expression constructs were subjected to Western blot analysis using the anti-T polyclonal goat antibodies. The untransfected and Tbx6-transfected cell lysates served as negative controls. (d.) Western blot probed with anti-actin as a loading control.

3.4.2 Tbx6 and T are primarily nuclear proteins when expressed in mammalian cell lines.

Full length Tbx6 and T were known to be primarily nuclear proteins *in vivo* (Kispert and Herrmann, 1994; Kispert et al., 1995a; White and Chapman, 2005). Therefore, these proteins were expected to be nuclear when expressed in mammalian cell lines. Tbx6 and T were transiently transfected individually into two different mammalian cell lines (COS-7 and 293T); after 24 hours, cells were fixed and stained with the following: Tbx6 and T specific antibodies, TO-PRO-3 to detect the nucleus, and Phalloidin to detect F-actin.

Tbx6 is a nuclear protein in both COS-7 (Figure 3-8) and 293T (Figure 3-9) cell lines. Cells expressing Tbx6 were stained with pre-bleed antisera (negative control), purified anti-Tbx6 N-terminal, or purified anti-Tbx6 Internal antibodies; these antibodies were detected with an anti-rabbit Alexa 488 secondary antibody (Molecular Probes). To visualize the nuclear and cytoplasmic compartments, all cells were stained with the chemical dyes rhodamine-phalloidin (to stain F-actin, red) and TO-PRO-3 (nuclear stain, blue). Tbx6 co-localized with the nuclear stain, while no staining was apparent in cells stained with pre-bleed antisera (see panel Ai, compared to Bi, Ci in Figures 3-8, 3-9). Transfection of the T expression vector into COS-7 and 293T cell lines confirmed that T was also a nuclear protein (Figure 3-10). Cells expressing T were stained with a commercially available anti-T polyclonal goat antibody, and this antibody was detected with a Rhodamine donkey anti-goat secondary antibody (placed in green channel). To visualize the nuclear and cytoplasmic compartments, cells were co-stained with the chemical dyes FITC-phalloidin to stain F-actin (placed in red channel) and TO-PRO-3, a nuclear stain (blue). T co-localized with the nuclear stain.

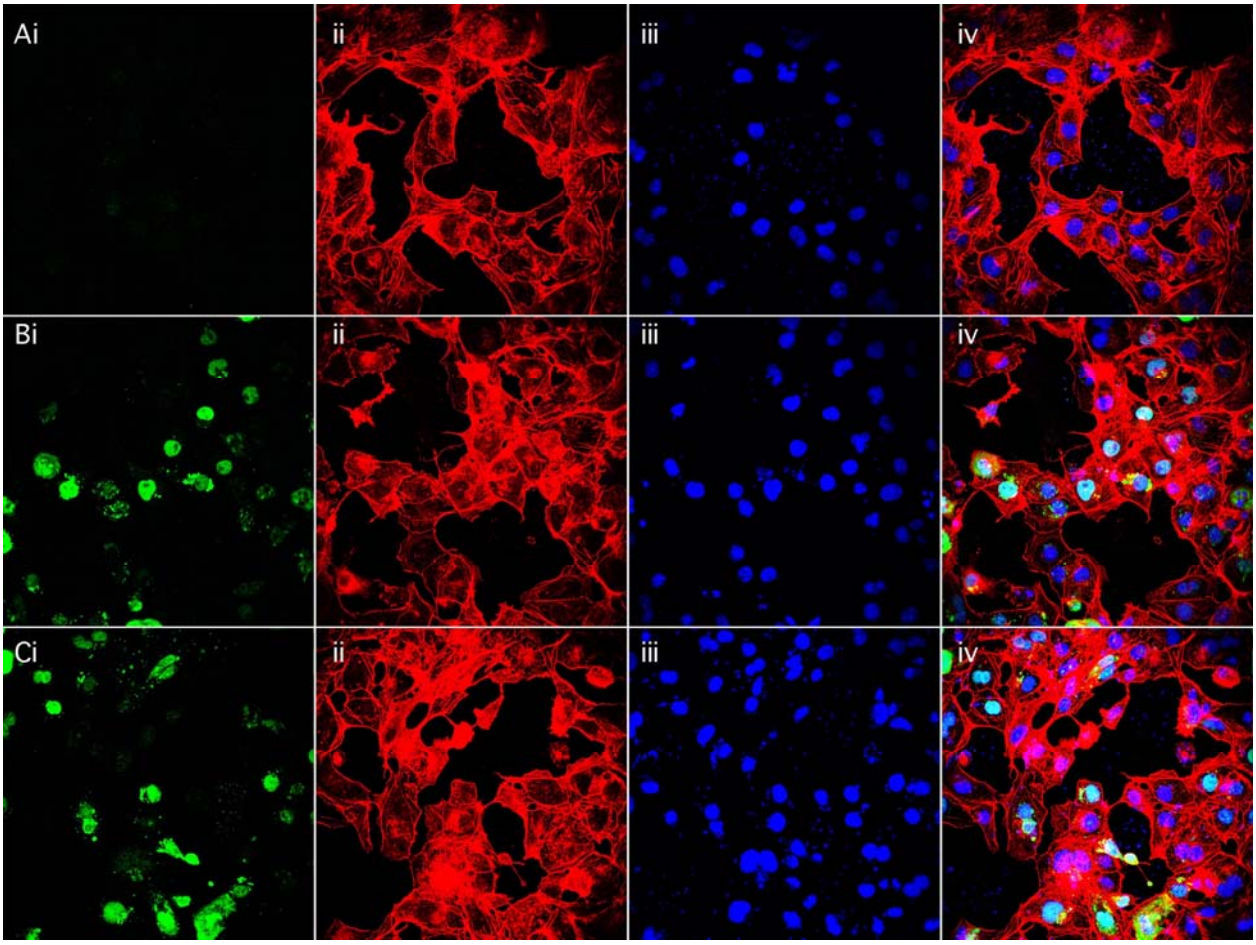


Figure 3-8: Nuclear localization of transfected T7 tagged-Tbx6 in COS-7 cells

Transiently transfected COS-7 cells with the Tbx6 expression construct. (Ai.-iv.) Pre-bleed antibody staining. (Bi.-iv.) Staining with a Tbx6 antibody raised against the N-terminus of Tbx6. (Ci.-iv.) Staining with a Tbx6 antibody raised against an Internal region of Tbx6. (i.) Pre-bleed / Tbx6 staining (green). (ii.) Staining with Rhodamine-Phalloidin to visualize F-actin (red). (iii.) Staining with TO-PRO-3 nuclear stain (blue). (iv.) Merged images showing the co-localization of Tbx6 with the nuclear stain.

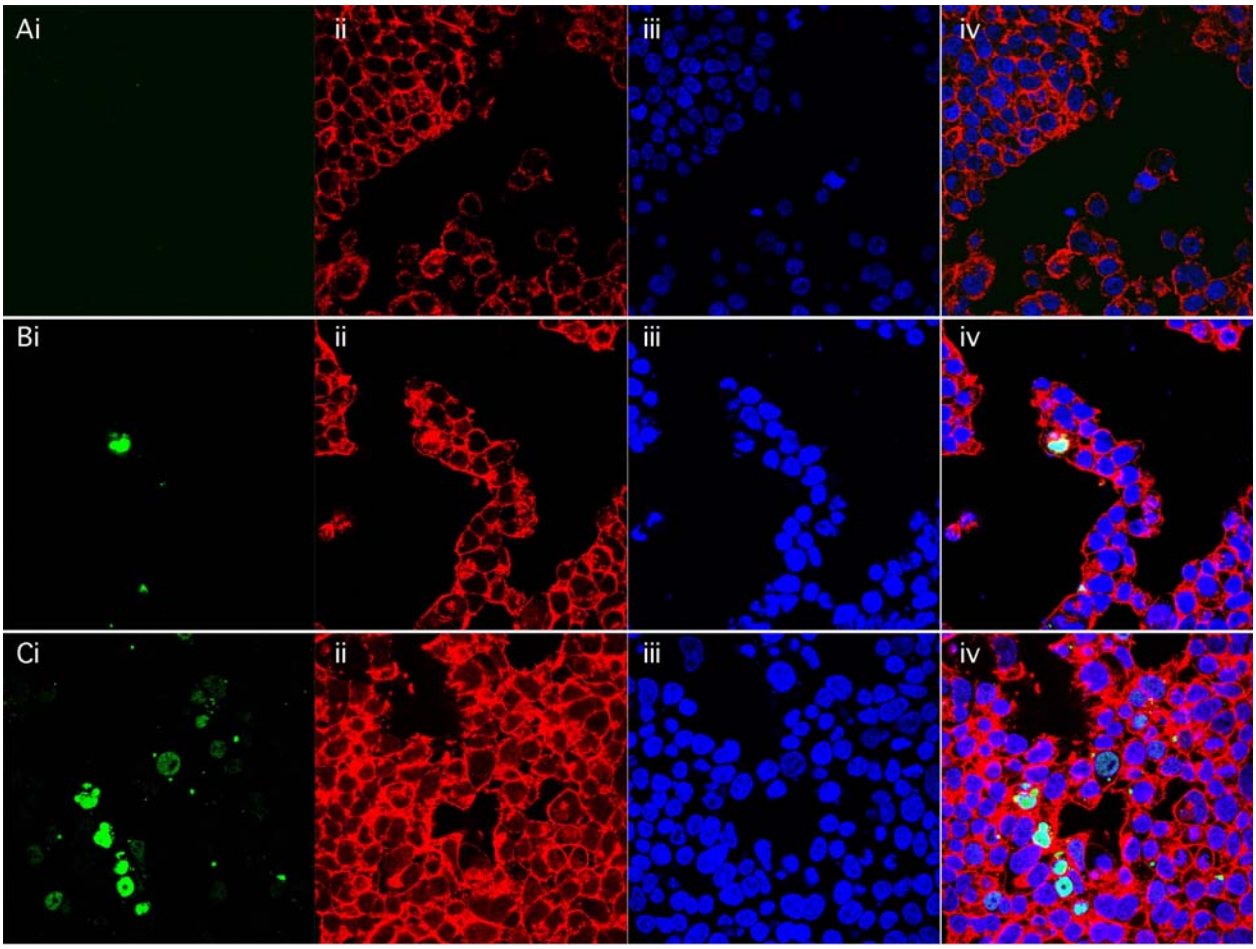


Figure 3-9: Nuclear localization of transfected T7 tagged-Tbx6 in 293T cells

Transiently transfected 293T cells with the Tbx6 expression construct. (Ai.-iv.) Pre-bleed antibody staining. (Bi.-iv.) Staining with a Tbx6 antibody raised against the N-terminus of Tbx6. (Ci.-iv.) Staining with a Tbx6 antibody raised against an Internal region of Tbx6. (i.) Pre-bleed / Tbx6 staining (green). (ii.) Staining with Rhodamine-Phalloidin to visualize F-actin (red). (iii.) Staining with TO-PRO-3 nuclear stain (blue). (iv.) Merged images showing the co-localization of Tbx6 with the nuclear stain.

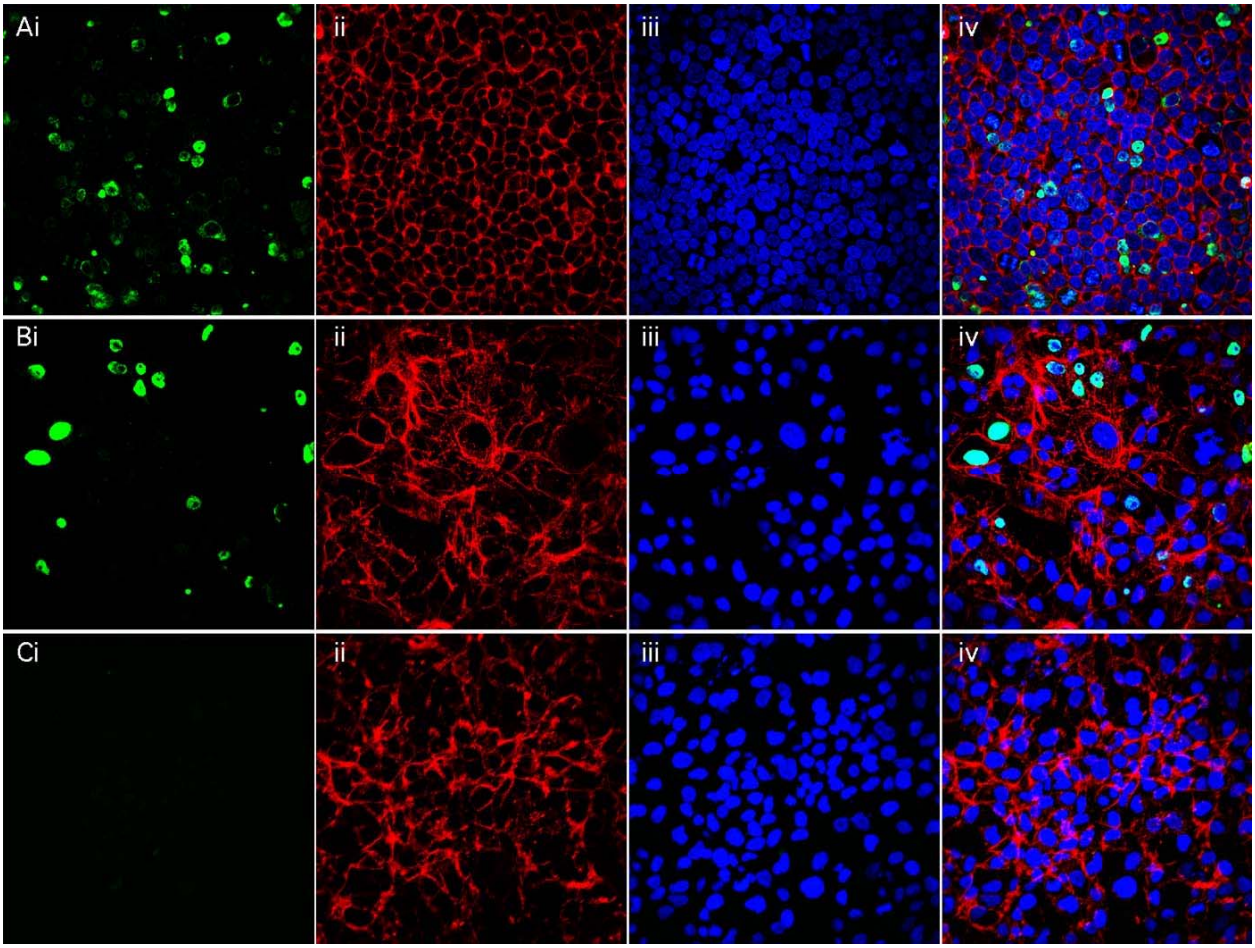


Figure 3-10: Nuclear localization of transfected T in cell culture

Transiently transfected 293T (A) and COS-7 (B,C) cells with the T expression construct. (Ai. and Bi.) Staining of T with the commercial anti-T antibody (green). (Ci.) Staining with the secondary anti-goat antibody only. (ii.) Staining with Phalloidin to visualize F-actin (red), and (iii.) TO-PRO-3 nuclear stain (blue). (iv.) Merged images showing the co-localization of T with the nuclear stain.

3.5 TBX6 FUNCTIONS AS A WEAK ACTIVATOR OF LUCIFERASE IN A TRANSCRIPTION ASSAY

While no palindromic sequences had been identified in the regulatory regions of T *in vivo* targets, multiple half-sites were found (Casey et al., 1998; Casey et al., 1999; Hilton et al., 2003; Tada et al., 1998). The importances of the individual half-sites were analyzed by mutating each site separately and then assaying the activity of the mutated enhancers either *in vitro* or *in vivo*. Results from these studies showed that in some cases, one binding site was necessary and sufficient for activity of the enhancer, while in other cases multiple half sites appeared to be required. In addition, studies by Conlon et al. (2001) showed that orientation and spacing of the T half-sites were critical for determining the binding specificity for Xbra, VegT and Eomesodermin. It is likely that each T-box target gene may show differences in the number, spacing and orientation of half-sites and that all of these factors contribute to the ability of the T-box protein(s) to modulate transcription of target genes *in vivo*. Therefore, to determine how Tbx6 affected transcription of downstream targets, I used an *in vitro* luciferase transcription assay. In addition to determining how Tbx6 alone affected transcription, I also tested how Tbx6 affected transcription of the reporter gene in combination with T. These combined studies are important because of the overlapping expression domains of T and Tbx6 during embryogenesis.

3.5.1 Design of a luciferase transcriptional assay system for T-box proteins

A luciferase transcriptional assay was used to determine how Tbx6 affected transcription of a reporter gene when that reporter gene was under the control of a minimal promoter and an enhancer with intact or mutated Tbx6 binding site(s). These assays were performed in COS-7 and 293T cells. The COS-7 cell line is an African green monkey kidney fibroblast-like cell line, while the 293T (HEK 293 T/17) cell line is a human fetal kidney epithelial cell line. Both of these lines were used in the study of numerous T-box transcription factors (Brown et al., 2005; Bruneau et al., 2001; Hofmann et al., 2004; Kispert et al., 1995a; Koshiba-Takeuchi et al., 2006). The luciferase transcriptional assay utilized three plasmids: a β-galactosidase control plasmid, protein expression plasmid(s), and a luciferase reporter plasmid. The β-galactosidase vector was used as a transfection efficiency control for each transfection to normalize any error due to cell

plating, transfection, and cell removal/lysis. The Tbx6 and T expression plasmids were described in Section 3.4, along with an empty control plasmid, pcDNA3.1/V5-His-MCS (MCS). Two different firefly luciferase source vectors were used in these assays: pGL3-Promotor (pGL3P, Promega), a vector that expressed the *luc* luciferase cDNA under the control of a SV40 minimal promoter, and another vector based on the pGL4.10 (Promega) promoterless plasmid. The pGL4 plasmid was modified to delete one putative Tbx6 binding site, at approximately +400 bp from the start of the *luc2* luciferase cDNA (pGL4M). Using this pGL4M plasmid, I developed plasmids that expressed luciferase under the control of the β-globin [pGL4M(βg)] and hsp70 [pGL4M(hsp)] minimal promoters. The sequences of all plasmids were confirmed by DNA sequencing. In general, the pGL4M(βg) vector had the lowest background activity, followed by the pGL4M(hsp) vector, and lastly the pGL3P vector (data not shown). The Luciferase Assay System (Promega) was used to detect the activity of luciferase, and the β-galactosidase Assay System (Promega) was used to detect the expression of β-galactosidase. Each set of experiments were performed in individual wells of 24-well plates, with one well reserved for no transfection (negative control/blank). Another well was reserved for transfection of luciferase under the control of a constitutively active SV40 promoter/enhancer; this construct would result in robust expression of the luciferase reporter, serving as a positive control for assaying luciferase activity. The β-galactosidase reactions were tested alongside known dilutions of purified β-galactosidase to serve as a positive control in the β-galactosidase assays.

3.5.2 Tbx6 weakly activates transcription of a luciferase reporter under the control of a T palindromic and half-site enhancers.

When this assay system was originally designed, it was unknown if full-length Tbx6 would function on its own to activate or repress transcription. Therefore, I cloned the T consensus sequence (Tbind), T half-site sequence (Thalf), and a mutated version of the T half-site (Tmut) into three different luciferase vectors that consisted of different luciferase variants, promoters, and vector backbone sequences (Table 3-2). We had shown that Tbx6 bound both the T palindromic and T half-site sequences, but not the mutated half-site in EMSA experiments (Figure 3-1, and not shown). Therefore, if Tbx6 functioned as an activator or repressor, the

Table 3-2: Luciferase constructs used in Tbx6 luciferase assays

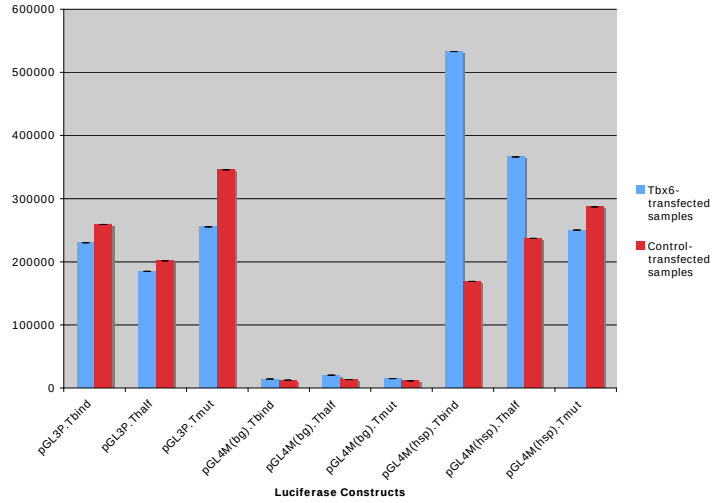
Luciferase Reporter	Binding Site (Enhancer)	Minimal Promoter	Construct Name
pGL3-Pro	T palindromic binding site	SV40	pGL3P.Tbind
pGL3-Pro	T half-site	SV40	pGL3P.Thalf
pGL3-Pro	Mutated T half-site	SV40	pGL3P.Tmut
pGL4M	T palindromic binding site	βglobin	pGL4M(βg).Tbind
pGL4M	T half-site	βglobin	pGL4M(βg).Thalf
pGL4M	Mutated T half-site	βglobin	pGL4M(βg).Tmut
pGL4M	T palindromic binding site	hsp70	pGL4M(hsp).Tbind
pGL4M	T half-site	hsp70	pGL4M(hsp).Thalf
pGL4M	Mutated T half-site	hsp70	pGL4M(hsp).Tmut

presence of a palindromic or half-site binding sequence would affect the strength of activation or repression of a reporter in cell culture. In addition, if Tbx6 functioned as a repressor, then a high background would be necessary to detect the drop in reporter transcription. These commercial reporters were designed to minimize background transcriptional noise; however the pGL3P vectors had higher levels of background and therefore were used to monitor repressor activity.

In 24-well plates, COS-7 and 293T cells were transfected with Tbx6 (or empty MCS expression construct) and each of the nine constructs listed in Table 3-2 as separate samples. The samples were processed, and the average luciferase units were determined from triplicate reads of each sample using the Luciferase Assay System (Promega). The data were graphed with the standard error from the triplicate reads of each sample (Figure 3-11). These data could not be directly compared because the data were not normalized for transfection efficiency using the β-galactosidase transfection control. However, three trends were apparent: (1) the reporter constructs resulted in different levels of activity when in the presence or absence of Tbx6, (2) the pGL4M(hsp)-derived constructs appeared to have higher induced responses to Tbx6, and (3) the reporter constructs responded differently depending on the cell line used. Two independent trials were completed and the resulting average luciferase data from those trials were similar.

To determine the effect of Tbx6 on the various luciferase reporters in the system, the data obtained in Figure 3-11 needed to be adjusted for the transfection efficiency of the various samples in the different wells with the β-galactosidase control; as mentioned previously, the

a.



b.

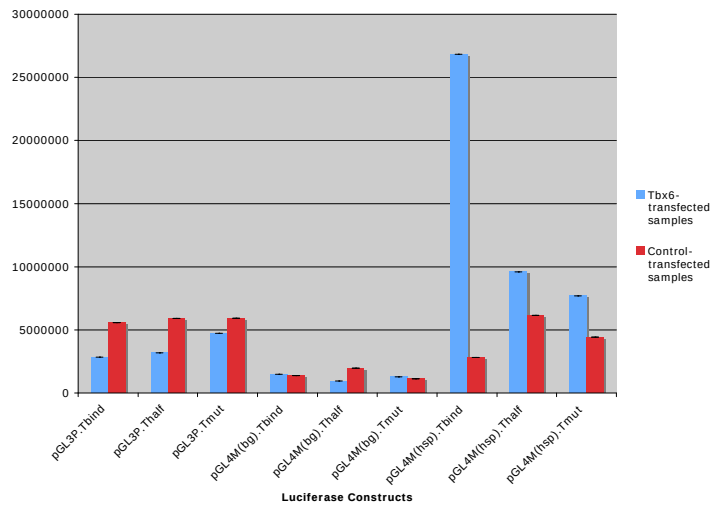


Figure 3-11: Luciferase data from one independent trial testing the effect of Tbx6 on various T binding site sequences

COS-7 (a) and 293T (b) cells were transfected with various luciferase reporters and either a Tbx6 or control expression construct to measure the effect of Tbx6 on various T binding site sequences. Reporters with either a T palindromic, T half-site, or a mutated T half-site sequence were transfected with Tbx6 or empty expression construct (as listed on the X-axis, legend). The total amount of DNA per transfection was kept constant (800 ng). The values given represented the raw luciferase data collected from one such trial and summarized the average luciferase units, unadjusted from the β-galactosidase control, for triplicate readings of each experimental sample as determined using the Luciferase Assay System (Promega); confidence intervals were reflected on data sets above. Note the y-axis was not scaled the same in the two graphs above due to the luciferase levels between different cell lines.

β-galactosidase control corrected for differences between transfection efficiency and sample concentration/processing. As observed in Figure 3-11, control transfected samples resulted in fairly significant luciferase activity; the control experiments represented the level of activity of the various reporter constructs when expressed in tissue culture without Tbx6 present. Therefore, the data were presented as a ratio of the adjusted luciferase units for the Tbx6 transfected sample over the control (empty expression vector) transfected sample. Figure 3-12 summarizes these normalized luciferase units (ratio of the Tbx6-transfected samples over their corresponding empty vector transfections) obtained from the same trials shown in Figure 3-11. Comparisons of ratios (Figure 3-12) should eliminate any variability due to the concentration and basal level activity of the reporter constructs used in the different sample wells; these ratios represented the fold activation due solely to the presence of Tbx6. As Tbx6 did not bind the mutated T half-site sequence by EMSAs, Tbx6 should have no effect on the reporter with mutated binding sites and therefore we expected that the ratio for all Tmut enhancers would be approximately one.

On average, Tbx6 typically resulted in a 2 to 3 fold activation when using the pGL4M(hsp).Thalf construct compared to samples transfected with pGL4M(hsp).Tmut, while 3 to 10 fold activation was observed when using the pGL4M(hsp).Tbind construct compared to samples transfected with pGL4M(hsp).Tmut. These results were cell line dependent; the COS-7 cell line typically resulted in lower fold activation (approximately 2 fold activation for pGL4M(hsp).Thalf and 3 fold activation for pGL4M(hsp).Tbind) and 293T resulted in higher fold activation (about 3 fold activation for pGL4M(hsp).Thalf and approximately 10 fold activation for pGL4M(hsp).Tbind). In this particular trial, the COS-7 cells transfected with the pGL4M(hsp).Tbind construct had a ratio of 2.607 (Tbx6 to control MCS expression vector), while the pGL4M(hsp).Thalf construct and pGL4M(hsp).Tmut construct resulted in ratios of 1.755 and 0.824, respectively; the 293T cells transfected with the pGL4M(hsp).Tbind construct had a ratio of 11.743 (Tbx6 to control MCS expression vector), while the pGL4M(hsp).Thalf construct and pGL4M(hsp).Tmut construct resulted in ratios of 3.619 and 1.247, respectively.

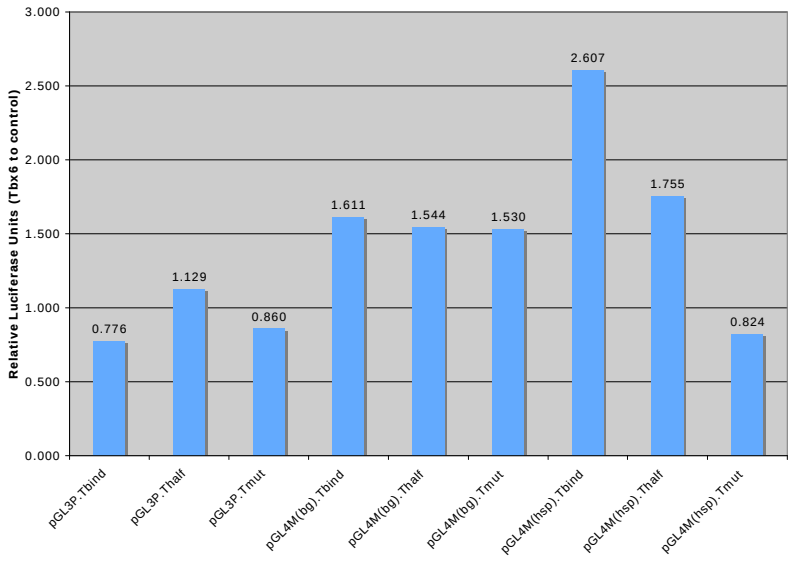
Therefore, in COS-7 cells transfected with Tbx6, the T palindromic sequence sample resulted in a 3.16 fold net activation (2.61 divided by 0.82) compared with a reporter containing a mutated half-site, while the T half-site resulted in a 2.13 fold net activation (1.755 divided by 0.824) compared with a reporter containing a mutated half-site. Using similar calculations for

the 293T cells transfected with Tbx6, the T palindromic sequence sample resulted in a 9.4 fold net activation (11.743 divided by 1.247) compared with a reporter containing a mutated half-site, while the T half-site resulted in a 2.9 fold net activation (3.619 divided by 1.247). No activation was observed using the pGL4M(βg) series of experiments, i.e. expression of luciferase in the presence of Tbx6 was about the same in all cell lines, as determined using normalized luciferase units. The pGL3P series of vectors appeared to vary between the two experimental trials, typically detecting weak activation of the pGL3P.Thalf construct only. Altogether, these data showed weak activation by Tbx6 using the pGL4M(hsp) luciferase constructs that contained either the T palindromic or half-site sequences.

There were several possible reasons why Tbx6 produced different levels of activation when co-expressed with the different minimal promoters and luciferase reporters. Tbx6 appeared to be stable when transfected for up to 48 hours since the protein was detected in the nucleus using immunofluorescence (Figures 3-8, 3-9). However, the stability of the luciferase reporter was never tested. The pGL4M vector backbone was derived from the pGL4.10 vector through site-directed mutagenesis of the *luc2* luciferase gene. This site-directed mutant changed the nucleotide sequence of the vector but created a silent mutation and therefore would not have changed the amino acid sequence of the protein. Although the experiments were performed following the procedures of the manufacturer, it is possible that the luciferase being produced from the different constructs may be partially degraded by the cell before the assays were completed, especially since the luciferase produced from the pGL3-based vectors contained a different amino acid sequence than the pGL4-based vectors. In addition, both COS-7 and 293T cell lines contained the SV40 T antigen. The presence of the T antigen will affect minimal promoters, and this may explain the high background level in the pGL3P constructs, containing a SV40 minimal promoter (Table 3-2).

Interestingly, similar experiments with Tbx5, primary cardiac myocytes yielded 100 times greater activation than non-cardiac cell types (Bruneau et al., 2001), suggesting that co-factors present in these cardiac cells affected the ability of Tbx5 to activate transcription. This could be the case with Tbx6 as well, as the above experiments were performed in COS-7 and 293T, two kidney cell lines, rather than a PS or PSM cell line, which are currently unavailable.

a.



b.

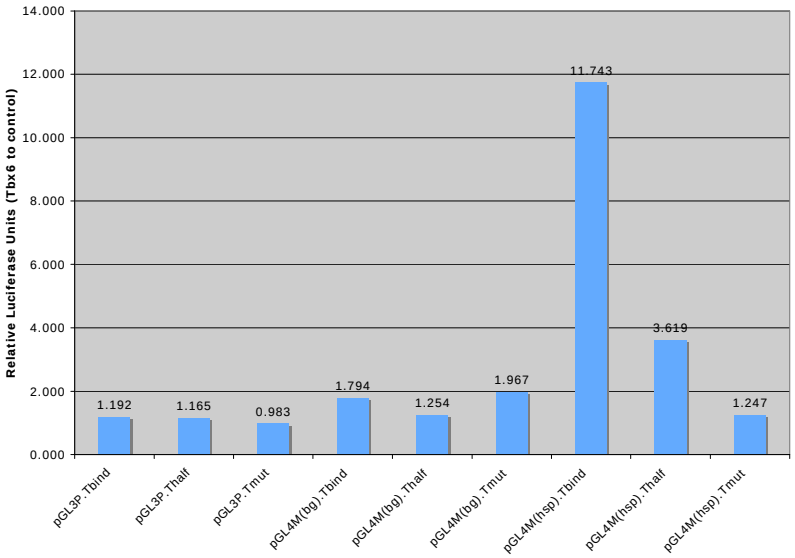


Figure 3-12: Tbx6 weakly activates transcription of a luciferase reporter

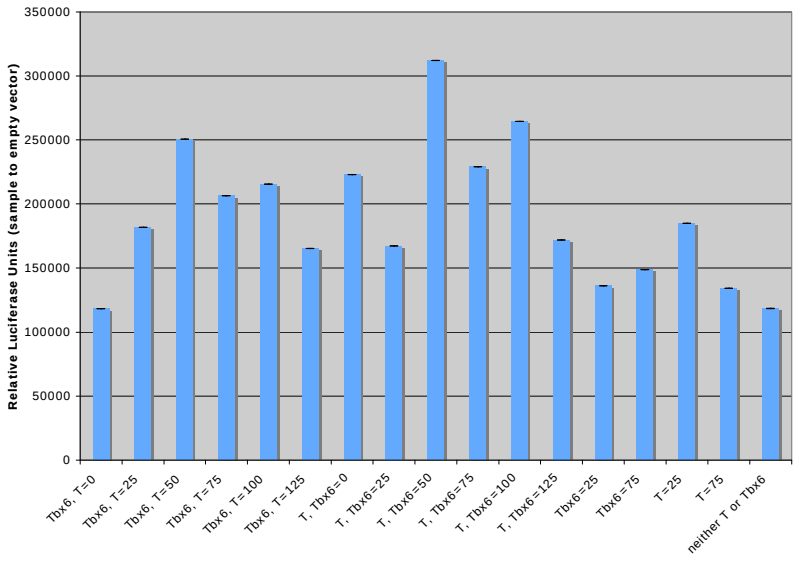
COS-7 (a) and 293T (b) cells were transfected with various luciferase reporter constructs and either a Tbx6 or control expression construct to measure the effect of Tbx6 on various T binding site sequences. Reporters with either a T palindromic, T half-site, or a mutated T half-site sequence were transfected with Tbx6 or the empty expression construct. The total amount of DNA per transfection was kept constant (800 ng). The values given represented the ratio of Tbx6-transfected samples (over the corresponding empty expression construct sample) to ensure that activation was due solely to the addition of Tbx6 to the reaction, as samples transfected with empty expression vector resulted in background luciferase activity (see Figure 3-11). As described in the text, Tbx6 weakly activated the pGL4M(hsp) vector, while providing minimal to no effect using the pGL4M(βg) and pGL3P constructs.

3.6 THE EFFECT OF T ON TBX6 TRANSCRIPTIONAL ACTIVITY IN CELL CULTURE

Tbx5 and Tbx20 were co-expressed in the *Xenopus* heart and physically interacted via sequences in their N-termini (Brown et al., 2005). Interestingly, luciferase transcriptional assays showed that co-transfection of Tbx5 in combination with either high or low doses of Tbx20 resulted in synergistic activity; however Tbx5, when co-transfected with intermediate doses of Tbx20, resulted in a repressive effect (Brown et al., 2005). This suggested that the relative levels of the two transcription factors were critical for controlling transcriptional activity. In this way, co-expression of T-box proteins could create yet another level to regulate target gene expression. My previous experiments using Tbx6 failed to show a robust activation of the luciferase reporter using the T palindromic site. It was therefore possible that instead of activating transcription, Tbx6 may function to down-regulate T targets by either actively repressing their transcription through the recruitment of transcriptional repressors or more passively by interfering with T binding to enhancer sequences through binding competition of target sequences. Tbx6 and T are co-expressed in the PS, but also have unique areas of expression. Using Tbx5/Tbx20 as a model for T-box activity, we investigated whether co-expression of Tbx6 and T could affect target gene expression using the luciferase reporter system described above. To examine whether Tbx6 was functioning as a repressor when both T and Tbx6 were co-expressed, I chose to use the pGL3P.Tbind construct, which exhibited the highest background of all the reporter plasmids and consequently a reduction of T activation would be easy to detect.

To measure the effect of Tbx6 and T co-expression on the pGL3P.Tbind reporter, one experimental construct was added at 100 ng to the transfections, while the other T-box expression construct was varied by addition of increasing levels of construct DNA from 0 ng to 125 ng total DNA (in 25 ng increments). The total amount of DNA per transfection was kept constant (800 ng). In addition to the controls described above, I included the pGL3P.Tbind without T or Tbx6 to establish the baseline for luciferase expression. The samples were processed, and the average luciferase units were determined from triplicate reads of each sample using the Luciferase Assay System (Promega). The data were graphed with the standard error from the triplicate reads of each sample (Figure 3-13). These data could not be directly

a.



b.

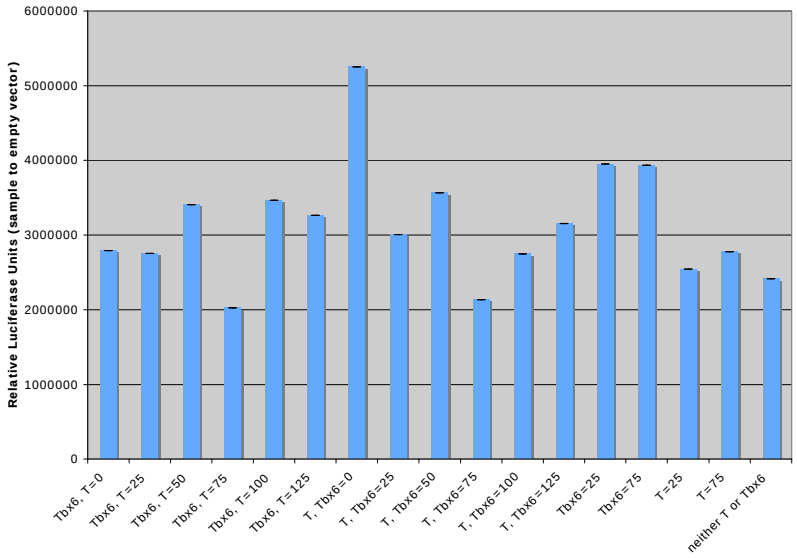


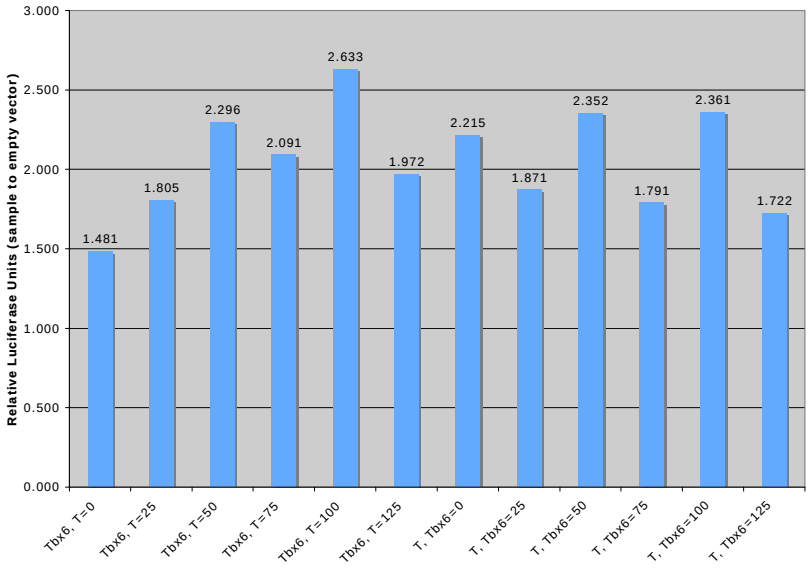
Figure 3-13: Luciferase data from one independent trial testing the effect of Tbx6 and T in co-transfection studies

COS-7 (a) and 293T (b) cells were transfected with various concentrations of T and Tbx6 to measure the effect of both proteins on the pGL3P.Tbind reporter. 100 ng of one T-box expression construct was added to the transfections, while the amount of DNA for the other T-box expression construct was varied from 0 to 125 ng total DNA (in 25 ng increments). The total amount of DNA per transfection was kept constant (800 ng). The values given represented the raw luciferase data collected from one such trial and summarized the average luciferase units, unadjusted from the β-galactosidase control, for triplicate readings of each experimental sample as determined using the Luciferase Assay System (Promega); confidence intervals were reflected on data sets above. Note the y-axis was not scaled the same in the two graphs above due to the luciferase levels between different cell lines.

compared because the data were not normalized for transfection efficiency using the β-galactosidase transfection control data. However, the data shown in Figure 3-13 suggested several trends, including (1) the pGL3P.Tbind reporter construct responded differently depending on the cell line used, as seen before, and (2) there were no major changes of activation or repression initially detected. Three independent experimental trials were completed and the resulting average luciferase data from all trials were similar.

To measure the effect of both Tbx6 and T on the pGL3P.Tbind reporter, the data obtained in Figure 3-13 needed to be adjusted for the transfection efficiency of the various samples in the different wells with the β-galactosidase control. Therefore, the data were presented as ratios of the adjusted luciferase units for the experimental transfected samples (with various combinations of T and/or Tbx6 expression constructs) over the control (empty MCS expression construct only) transfected sample. As described previously, these ratios should eliminate any variability due to the concentration and activity of the pGL3P.Tbind reporter construct because of high background activity without any T or Tbx6 present. These ratios (Figure 3-14) represented the fold activation due solely to the presence of Tbx6 and/or T. T had previously been shown to activate reporters containing the T palindromic site (Kispert et al., 1995a). In this set of experiments, Tbx6 did not result in activation of the pGL3P.Tbind reporter construct. The presence of both proteins did not change these behaviors. On its own, Tbx6 had no effect on the reporter construct, while T on its own activated the reporter construct. Titrating in increasing amounts of Tbx6, while keeping the concentration of T constant, and vice versa, did not significantly alter the expression of the reporter. This does not necessarily mean that these two proteins were not affecting each other. The sensitive nature of T-box protein concentration, as evidenced by mutations in several T-box genes leading to haploinsufficiency (including T as reviewed in Chapter 1), may not be recapitulated in these tissue culture assays using synthetic enhancers for Tbx6 and T. For example, Brown et al. (2005) used a bona fide enhancer to confirm that Tbx5 and Tbx20 acted synergistically in heart development, not a synthetic sequence enhancer. Because the signal to noise ratio was so high for the pGL3P.Tbind construct, these experiments could be repeated using the pGL4M(hsp).Tbind construct that resulted in approximately 10 fold activation of reporter in Tbx6 to determine if the failure to detect an interaction was due solely to the construct used within the experiment.

a.



b.

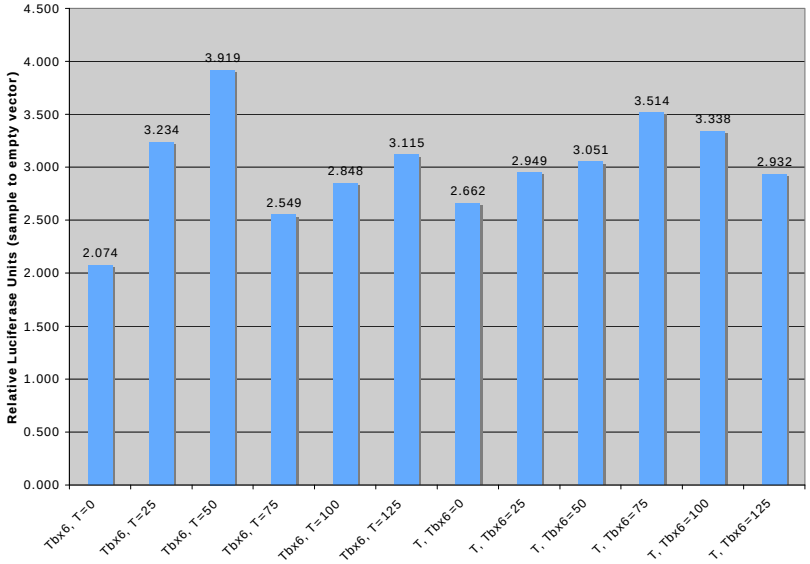


Figure 3-14: There is no correlation between the presence of Tbx6 and T in co-transfection studies

COS-7 (a) and 293T (b) cells were transfected with various concentrations of T and Tbx6 to measure the effect of both proteins on the pGL3P.Tbind reporter. 100 ng of one T-box expression construct was added to the transfections, while the amount of DNA for the other T-box expression construct was increased 0 to 125 ng total DNA (in 25 ng increments). The total amount of DNA per transfection was kept constant (800 ng). The values given in the graphs represent the ratio of the various transfection samples to an empty expression construct to ensure that activation was due only to expression of the specific proteins.

3.7 DISCUSSION

The T-box family members are related through a conserved DNA binding domain, the T-domain, suggesting that they may bind similar sequences. The crystal structure of two T-box proteins, T and Tbx3, had been determined. These structures showed that Tbx3 bound to the T palindromic consensus sequence as a monomer, meaning that two Tbx3 molecules bound independently to the sequence; however T dimerized upon binding to the palindromic sequence (Coll et al., 2002; Muller and Herrmann, 1997). As a first step toward identifying target genes of Tbx6, we tested whether Tbx6 bound to the consensus sequence identified for T. Our results confirmed earlier studies showing that human TBX6 bound to the T palindromic site *in vitro* (Papapetrou et al., 1999), and determined that Tbx6 readily bound the T half-site sequence as well (Figure 3-1b). At low protein concentrations (relative to probe concentration), binding to the palindromic site was equally competed using either the unlabeled T palindromic site or half-site, suggesting that Tbx6 bound to the palindromic sequence as a monomer (Figure 3-1b). However, using a truncated form of the human TBX6 (containing AA 1-228, including the DNA binding domain) and the T palindromic site, Papapetrou et al. (1999) observed two mobility complexes by EMSAs and suggested that Tbx6, like T, bound to the palindromic sequence as a dimer. We therefore repeated our EMSAs using increasing concentrations of Tbx6 protein and a low concentration of the palindromic probe; in these experiments, a second, slower mobility complex was also observed in addition to the faster migrating complex (Figure 3-3). These results suggested that although two Tbx6 subunits bound the palindromic sequence simultaneously, Tbx6 required only the half-site to be present when binding DNA. The data therefore suggest that Tbx6 should readily recognize the half-site in enhancer regions of target genes.

All proteins containing T domains had been aligned to discern the evolutionary changes occurring within this protein family; Tbx6 had been placed within its own subfamily (Papaioannou and Goldin, 2003). Because there was no information about the binding of other subfamily members, we determined the consensus Tbx6 binding site *in vitro* (White and Chapman, 2005). Tbx6 selected the following consensus sequence: 5'- AGGTGTBRNNN -3' (Table 3-1). Tbx6 selected some preferences at particular positions, specifically selecting sequences containing anything but A at position 7 (Table 3-1). Also, Tbx6 appeared to select sequences containing either A or G, selecting G approximately 20% of the time, similar to Tbx5;

on the other hand, T always selected sequences containing A at position 8 (Ghosh et al., 2001; Kispert and Herrmann, 1993). Tbx6 bound the half-site and to binding sites selected that were similar to those of Tbx5, suggesting that the DNA binding characteristics described for the Tbx2 subfamily may be applicable for the Tbx6 subfamily. The binding preferences that we had identified could now be applied to predict putative binding sites within the enhancers of Tbx6 target genes.

Luciferase transcriptional assays revealed that Tbx6 functioned as a weak activator on its own. There had been two recent reports suggesting that Tbx6 was a weak activator (Hofmann et al., 2004; Yasuhiko et al., 2006). When testing a transcription factor that weakly activates transcription on its own, there are two possible experiments that can be completed: (1) add a powerful activation or repression domain to the protein (VP-16 activator and *Drosophila* engrailed repressor), or (2) co-express the protein of interest with its co-activator. The authors of these papers tried to address both paths to observe how Tbx6 selected target sequences. Hofmann et al. (2004) showed that a Tbx6-VP16 fusion protein, when co-expressed with the LEF-1 transcription factor, activated luciferase under the control of the *Dll1* 'msd' enhancer; however Tbx6 (without the VP-16 activator) and LEF-1 were unable to activate transcription. While these results were not entirely satisfying and raised questions about whether Tbx6, or another T-box protein, was necessary to regulate the 'msd' enhancer, the authors showed that the T-box binding sites were necessary by mutating the T-box binding sites *in vitro* and *in vivo*, as determined by a loss of reporter activity (Hofmann et al., 2004). Curiously, Hofmann et al. (2004) used a truncated version of Tbx6 (which may account for the inability of Tbx6 to activate the 'msd' enhancer). As previously mentioned in this thesis, removing the C-terminal half of Tbx5 (AA 238 – 518) significantly enhanced the protein's DNA binding affinity compared to full length Tbx5 (Ghosh et al., 2001). This suggested that truncated proteins may bind DNA sequences tightly, interfering with other DNA binding proteins that function within the same regions to affect the expression of downstream targets. Altogether, the authors concluded that a T-box protein, specifically T or Tbx6, was necessary to activate the *Dll1* 'msd' enhancer, although they never published data testing T in their *in vitro* studies, stating only that T resulted in minimal activation of luciferase reporters on its own, like Tbx6 (Hofmann et al., 2004).

In Yasuhiko et al. (2006), the authors had determined that Tbx6 activated *Mesp2*, a Notch signaling pathway gene, using both a consensus Tbx6 binding site and miscognate binding site.

Tbx6 bound a portion of the promoter/enhancer of *Mesp2* in EMSA experiments, producing a slower mobility complex although only one consensus Tbx6 binding site was present; further analysis showed that the miscognate site 5'- ATTCTGACCC -3' was bound by Tbx6 only when Tbx6 bound to the cognate 5'- AGGTGTGGAT -3' sequence immediately downstream of it (Yasuhiko et al., 2006). Thus Tbx6 may have the ability to bind miscognate sites *in vivo* after binding a cognate site. This ability was previously discussed for T and other proteins that weakly dimerize upon binding DNA sequences (Coll et al., 2002). However, the function of both sites *in vivo* was not determined because the authors instead chose to delete all sites (including a third Tbx6 binding site occurring further downstream) in their luciferase reporter system *in vitro* and their transgenic *in vivo* studies (Yasuhiko et al., 2006). Future tests aimed at determining the importance of miscognate binding sites would clarify whether the observed binding of Tbx6 to these sequences was an artifact of the EMSA process.

Activation of *Mesp2* was accomplished by the expression of Tbx6 and the Notch signaling pathway, either the Notch intracellular domain (ICN) or an activated RBP-Jκ-VP16; the use of the Notch ICN resulted in constant activation of Notch signaling, and therefore active RBP-Jκ *in vivo*. This suggested that Tbx6 only activated targets when the Notch (or another signaling pathway) was active within the same cell, as Yasuhiko et al. (2006) also showed that Tbx6 on its own had only weak activating ability. Ultimately, our luciferase transcription assays suggested that full-length Tbx6 weakly activated transcription of a synthetic construct containing a Tbx6 half-site or palindromic site for the pGL4M(hsp) derived constructs (Chapter 3.5). This assay system utilizing the pGL4M(hsp) derived constructs and Tbx6 could be used to map activation and repression domains in Tbx6. Transcriptional reporter assays were performed for T and allowed the mapping of two activation and two repression regions in the C-terminus of T (Kispert et al., 1995a). This assay system could be adapted in the future to include luciferase constructs with multiple T palindromic or half-site binding sites as several T-box proteins required multiple binding site sequences, either palindromic or half-sites, to induce strong activation when using synthetic enhancer constructs in luciferase assays (Carlson et al., 2001; Kispert et al., 1995a).

Studies have shown that Tbx5 and Tbx20, two T-box proteins with overlapping expression, physically interacted within their N-terminal regions and they effect the ability of each other to function within a luciferase assay transcription assay (Brown et al., 2005). On its

own, T functioned to activate transcription of various reporters (Conlon et al., 2001; Kispert et al., 1995a). Tbx6 functioned here as a weak activator (Chapter 3.5). Therefore, I sought to determine whether Tbx6 and T, which are co-expressed in the PS, can affect the transcriptional activity of each other. Titrating in one T-box expression construct, while keeping the concentration of the other expression construct constant, did not significantly alter the native activity of either T-box protein. This does not necessarily mean that these two proteins were not affecting each other. The sensitive nature of T-box protein concentration, as evidenced by mutations in several T-box genes leading to haploinsufficiency (including T as reviewed in Chapter 1), may not be recapitulated in tissue culture assays using synthetic enhancers for Tbx6 and T. Brown et al. (2005) used a bona fide enhancer to confirm that Tbx5 and Tbx20 acted synergistically in heart development. Others in the lab are investigating whether Tbx6 and T physically interact through co-immunoprecipitation and yeast two-hybrid experiments as well as testing whether truncated versions of Tbx6 ultimately interfere with T function *in vivo*. Altogether, these experiments seek to further understand the relationship between the two T-box proteins, T and Tbx6.

4.0 NOTCH SIGNALLING FUNCTIONS BOTH UPSTREAM AND DOWNSTREAM OF TBX6

4.1 INTRODUCTION

Notch signaling plays many vital roles throughout development; the signaling pathway itself is evolutionarily conserved (functioning in organisms ranging from flies to humans) and serves to exchange signals between neighboring cells through the Notch receptors and DSL ligands (reviewed in Artavanis-Tsakonas et al., 1999). Among the many processes studied involving Notch signaling, somitogenesis relies on the Notch/Delta signaling pathway for proper boundary formation between somites (reviewed in Saga and Takeda, 2001). As detailed in the Introduction (Section 1.3), somitogenesis begins when mesodermal cells ingress through the PS, migrating laterally to form the PSM and undergo segmentation to form the somites (reviewed in Gossler and Hrabe de Angelis, 1998). The paraxial mesoderm cells form bilateral bands of mesenchymal tissue on either side of the neural tube and are specified as they emerge from the PS. These cells then proceed through somitogenesis, first by forming an epithelial somite, followed by cellular differentiation (Gossler and Hrabe de Angelis, 1998).

The Notch ligand, *Delta-like 1 (Dll1)* is expressed in the PS and PSM in a domain that overlaps with *Tbx6* (Bettenhausen et al., 1995; Chapman et al., 1996a). Like other genes that are normally expressed in the PSM, *Dll1* is not expressed in the *Tbx6* null mutant embryos (Chapman and Papaioannou, 1998), suggesting that it might be a target of Tbx6. Alternatively, its expression may be lost simply because the PSM is not present in the mutant. At the time these studies were beginning, it was also known that: 1) reduction of *Tbx6* expression in embryos below heterozygous levels led to fusions of the ribs and vertebrae, a characteristic shared by Notch signaling mutants, 2) *Tbx6* genetically interacted with the classic mouse mutant *rib-*

vertebrae (rv), and 3) the classic mouse mutant *rv* mapped close to *Tbx6* (Beckers et al., 2000b; White et al., 2003).

4.1.1 Aims of this study

Because the mouse mutant *rv* mapped close to *Tbx6*, I set out to determine if *rv* was a mutation in *Tbx6* by examining the sequence of *Tbx6* mRNA expressed in *rv* homozygous mutant embryos; this was done to determine whether mutations in the coding region of *Tbx6* were responsible for the *rv* mutant phenotype. The cDNAs cloned from *rv/rv* mutant embryos were identical to the known wild type sequence for *Tbx6*. During these studies, we uncovered a polymorphism that appeared to map upstream of the *Tbx6* start of transcription; I next focused on this region to uncover the nature of the *rv* mutation. Examination of the PSM enhancer of *Tbx6* revealed a conserved binding site for RBP-Jκ, a Notch transcription factor. I therefore tested whether RBP-Jκ functioned upstream of *Tbx6* both *in vitro* and *in vivo*. Lastly, I sought to determine whether *Dll1*, a Notch signaling ligand expressed in the PSM, was a direct target of Tbx6 using *in vitro* EMSA assays.

4.2 *RV* IS A *TBX6* HYPOMORPHIC MUTATION

rv was determined to be a spontaneous recessive mutation characterized by fusions of ribs and vertebrae, shortened trunk, kinked tail, formation of a single kidney, and reduced fertility (Beckers et al., 2000b; Nacke et al., 2000; Theiler and Varnum, 1985). Interestingly, *rv* had been mapped to position 62 cM on mouse chromosome 7 (Beckers et al., 2000b), which was close to the mapped location of *Tbx6* at position 61 cM (Chapman et al., 1996a). *Tbx6* homozygous embryos were recognized by an enlarged tailbud, ectopic lateral neural tubes, and embryonic lethality at approximately e12.5 (Chapman and Papaioannou, 1998). Our lab attempted to rescue this *Tbx6* mutant using a transgene containing the whole coding region of *Tbx6*, along with cis-acting regulatory sequences necessary to drive expression in a *Tbx6*-specific pattern (Section 1.5.5). Although this rescue was incomplete due to the low level of *Tbx6* expression from the

transgene, embryos formed somites that later gave rise to fused vertebrae and ribs and an enlarged tail region with multiple outgrowths; these were all characteristic of the *rib-vertebrae* homozygous mutant embryos (White et al., 2003). If the rib-vertebrae phenotypes were due to a mutation in *Tbx6*, we reasoned that it would not be a null mutation in *Tbx6* but rather a mutation that resulted in decreased Tbx6 function and therefore a hypomorphic mutation.

4.2.1 *Tbx6* transcripts appear normal in *rv*/*rv* embryos

There were several different ways that alterations in a gene had previously been shown to cause a hypomorphic mutation, including spontaneous hypomorphic mutations that resulted in production of a truncated protein, such as the *normoblastosis* mutation in mice, splicing errors of native proteins, such as the alternative splicing of *daf-4* in *C. elegans*, and for various reasons in incomplete transgene rescue, such as *cd-18* in the mouse and also as previously mentioned for *Tbx6* (Birkenmeier et al., 2003; Gunther and Riddle, 2004; White et al., 2003; Wilson et al., 1993a). As a first step to determine whether the *rv* phenotype was due to a mutation in *Tbx6*, I examined *Tbx6* transcripts to test if *Tbx6* in *rv*/*rv* embryos was different from *Tbx6* found in wild-type embryos.

To detect errors in *Tbx6* mRNA transcripts, total mRNA was isolated from an *rv/rv* embryo and primers specific for the full-length cDNA of *Tbx6* were used for reverse transcription PCR (RT-PCR). One major product of approximately 1.6 kb was produced, gel purified, and cloned into a sequencing vector. The sequences of five independent clones were compared to the known wild-type sequence of *Tbx6*. The presence of one major band of the expected size for *Tbx6* in the *rv/rv* sample suggested that if the *rv* mutation was indeed in *Tbx6*, it was unlikely to affect splicing since the cDNA appeared full length. Aside from the GC rich regions of Tbx6 that previously were difficult to sequence, the *rv/rv Tbx6* cDNA sequences were identical to that of our wild-type *Tbx6* cDNA. As previously described in Chapter 2 (Section 2.4.3), we believe that sequencing of *Tbx6* cDNA sequences resulted in sequencing errors when using PCR-based sequencing and cloning due to high GC regions within stretches of the cDNA sequence. Therefore, the *rv* mutation appeared to have a normal *Tbx6* cDNA sequence.

4.2.2 The *rv* mutation is caused by an insertion in the *Tbx6* enhancer

Concurrent work in the Chapman lab using genomic Southern blot hybridization analysis and probes spanning the *Tbx6* genomic region successfully identified a polymorphism in *rv*/*rv* genomic DNA. Using various restriction digests of wild type *Tbx6* and *rv*/*rv* DNA, members of the lab narrowed down the mutated region to the 5' end of the *Tbx6* locus; these experiments suggested that there was an insertion in this region (Figure 4-1a). Therefore, I amplified the genomic region from -1526 to +1116 (relative to the start of *Tbx6* transcription) using PCR from *rv*/*rv* genomic DNA. As predicted by Southern blot analysis, this region contained an insertion of 185 bp into a known *Tbx6* enhancer region. Interesting, sequencing analysis revealed that this inserted DNA was actually a duplicated region of the *Tbx6* genome containing a portion of the first exon after the initiating methionine, the entire first intron, and a portion of the second exon (see figure 4-1b, as modified from White et al. (2003)). The *rv* mutation was therefore caused by the insertion of this duplicated and inverted region 243 bp upstream of the *Tbx6* coding region. Although the insertion never changed the sequence of the *Tbx6* transcript, additional results from the Chapman lab confirmed that this mutation lowered the level of *Tbx6* transcription *in vivo*, resulting in the *rv/rv* hypomorphic allele of *Tbx6*, $Tbx6^{rv}$, as published in White et al. (2003).

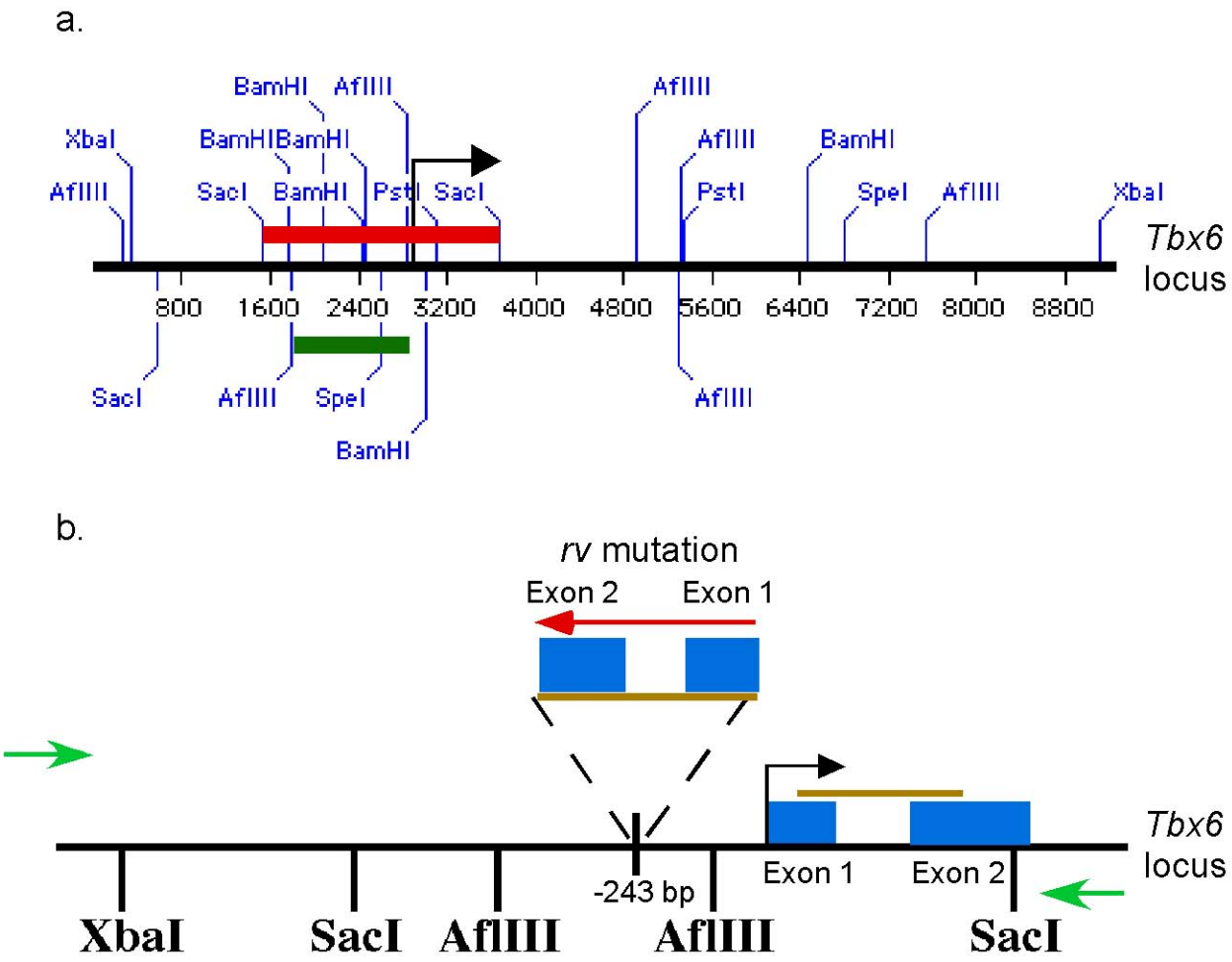


Figure 4-1: The *rv* mutation results from a partial duplication of *Tbx6*

(a.) Restriction enzyme map of the Tbx6 locus used for genomic Southern blot analysis of wild type and *rv*/*rv* DNA that narrowed down the location of the *rv* mutation at the *Tbx6* locus. Polymorphisms were observed in digests of genomic DNA with AflIII (green line) and SacI (red line), which predicted a small (~200 bp) insertion. (b.) Figure modified from White et al. (2003). Diagram of the 5' end of the *Tbx6* genomic locus showing the positions of the exons (blue boxes), start of transcription (black arrow), and the nature and position of the *rv* lesion. A brown line indicates the *Tbx6* genomic region that is duplicated and inverted (red arrow) in the *rv* mutation. Green arrows represent primers used to ultimately amplify this portion of the *Tbx6* genome. The *rv* mutation is caused by the insertion of this duplicated and inverted region 243 bp upstream of the *Tbx6* coding region.

4.3 NOTCH SIGNALLING IS REQUIRED FOR TBX6 REPORTER GENE EXPRESSION

4.3.1 Introduction

The *rv* mutation was caused by an insertion into a known *Tbx6* PSM enhancer (White et al., 2005), and the location of the *rv* insertion negatively affected *Tbx6* expression without deleting any of the existing sequence. We therefore sought to determine which sequences within this known *Tbx6* PSM enhancer might be important binding sites for known transcription factors. As detailed in Chapter 1, *Tbx6* is expressed in the PS and PSM and is down-regulated as the PAM segments to form the somite (Chapman et al., 1996a). Studies in the Chapman lab have focused on identifying the factors responsible for controlling the spatial and temporal expression of genes involved in somitogenesis, in particular *Tbx6*. Initially these studies focused on identifying the cis-acting regulatory regions of *Tbx6;* other members in the lab used genomic regions of *Tbx6* to drive a *lacZ* reporter gene with the goal of determining key genomic regions responsible for controlling *Tbx6* expression (see Chapter 1.5.5). Three of these *Tbx6* reporter constructs (generated by others in the Chapman lab) laid the groundwork for my studies on the control of *Tbx6* expression by Notch signaling. Figure 4-2 (modified from White et al. (2005)) summarizes their experiments.

We compared the 2.3 kb region located just upstream of the start of *Tbx6* transcription, encompassing the PSM enhancer, from the mouse with those of the known mammalian orthologs from the Ensembl genomic databases for the human, rat, and dog genomes using the mVISTA visualization module for the AVID genomic alignment program (Bray et al., 2003; Frazer et al., 2004; Mayor et al., 2000). This comparison revealed a large region of homology (greater than 75% nucleotide identity) as shown in the comparison between mouse and human sequences (see Figure 4-3). Within this region, and conserved in human, rat, mouse and dog, was a binding site for RBP-Jκ, the main, if not the only transcription factor functioning in the Notch-signaling pathway in mice (Tun et al., 1994). The binding site for RBP-Jκ was located 312 bp upstream of the *Tbx6* transcription start site (Figure 4-3). Notch signaling is known to play a critical role in somite formation and patterning (see Chapter 1.7). Therefore, I set out to determine whether the binding site for RBP-Jκ was necessary for proper *Tbx6* expression by testing whether the

Drosophila ortholog of RBP-Jκ, Suppressor of Hairless (Su[H]), bound *in vitro* to probes matching this sequence or sequences harboring a mutation of the RBP-Jκ binding site using EMSA experiments. Ultimately, I tested whether loss of this RBP-Jκ site affected expression of reporter constructs *in vivo*.

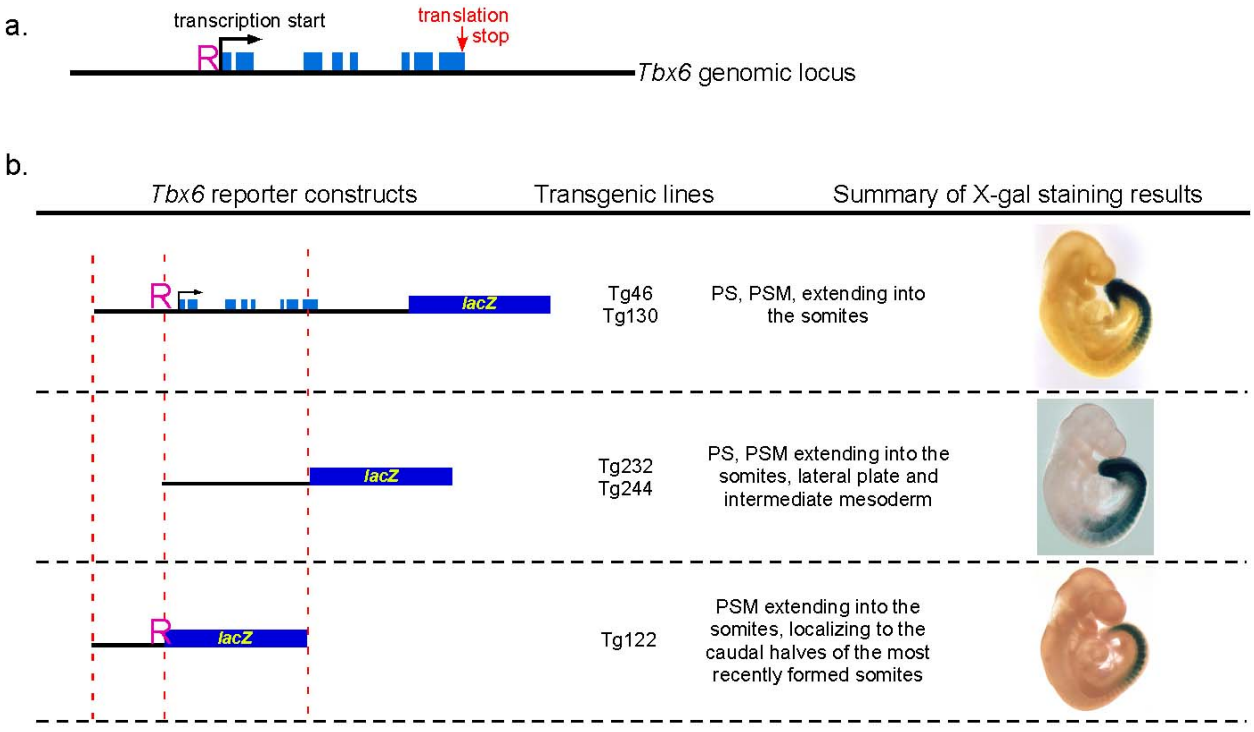


Figure 4-2: Summary of select cis-acting *Tbx6* regulatory regions

Figure modified from White et al. (2005). (a.) Schematic of the *Tbx6* genomic locus with the positions of the start of transcription (black arrow), exons (blue boxes) and stop codon (pink arrow) shown. The position of the RBP-Jκ binding site (R) is indicated. (b.) *Tbx6* transgenic constructs and summary of β-galactosidase staining results at e9.5. The position of the RBP-Jκ binding site (R) is indicated in the relevant constructs.

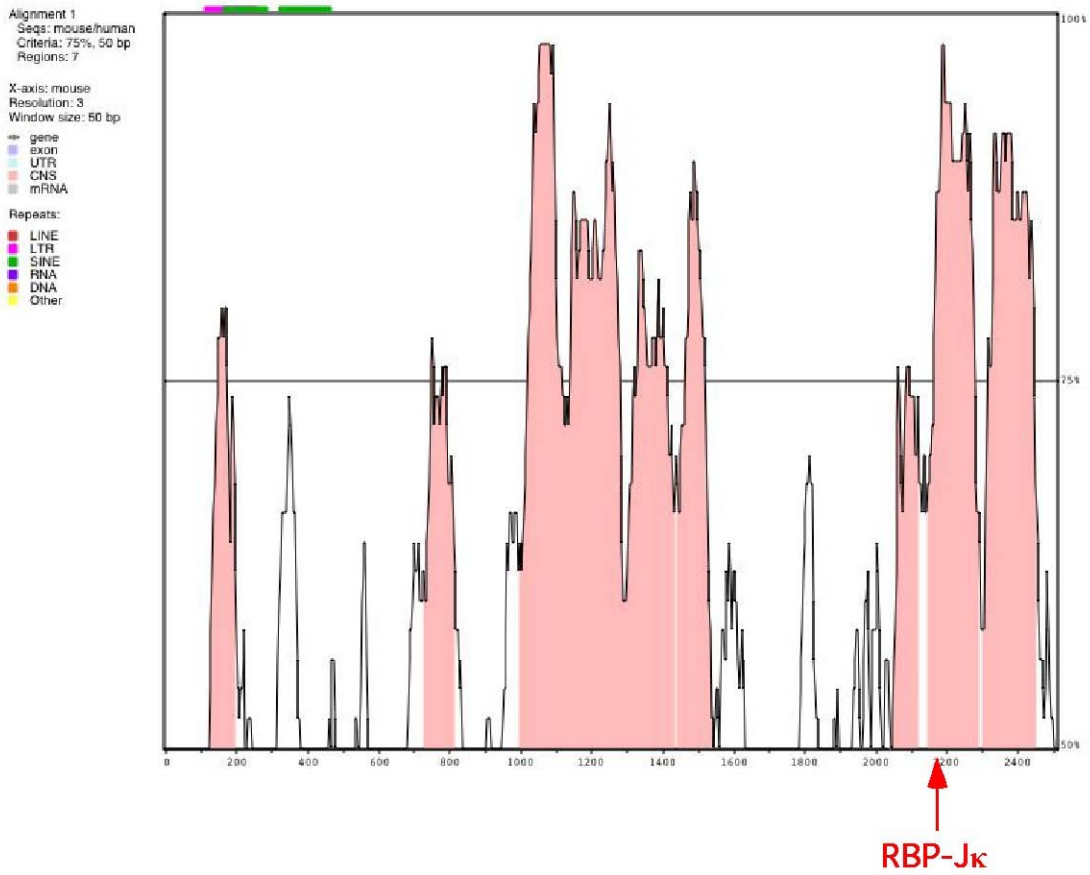


Figure 4-3: A conserved predicted binding site for RBP-Jκ in the aligned enhancer for mammalian orthologs of Tbx6

The 2.3 kb genomic sequence upstream of the mouse *Tbx6* transcription start site compared to relevant sequences upstream of the *Tbx6* gene from human, rat and dog using the mVISTA/AVID alignment program revealed three regions of greater than 75% homology; the above figure was generated comparing the mouse and human sequences (pink highlighted regions) with the position of the RBP-Jκ binding site indicated by the red arrow. RBP-Jκ is approximately 310 bp upstream of the start of *Tbx6* transcription.

4.3.2 Su[H], a RBP-Jκ ortholog, binds to sequences matching the *Tbx6* enhancer

Since the *Drosophila* Su[H] and mammalian RBP-Jκ bound to the same consensus sequence *in vitro* (Brou et al., 1994; Singson et al., 1994; Tun et al., 1994), I expressed and purified a GST-Su[H] fusion protein in bacteria for use in EMSA experiments; the GST-Su[H] construct was a gift from C. Xu and R. Carthew and was previously used to express full-length GST-Su[H]. To investigate whether GST-Su[H], and ultimately RBP-Jκ, bound discriminately to the *Tbx6* enhancer but not a mutated form of this region, I created double-stranded, radioactive probes from double stranded, annealed oligos matching a 36 bp *Tbx6* genomic sequence that contained the predicted RBP-Jκ binding site ($Su[H]^{bind}$). I also generated a mutant form of this probe ($Su[H]^{mut}$, Figure 4-4a) that was identical to $Su[H]^{bind}$, except that it contained a single guanine to cytosine substitution. An identical mutation in the Su[H] binding site had previously been shown to prevent binding of Su[H] *in vitro* (Bailey and Posakony, 1995).

I verified that Su[H] bound the predicted RBP- Jκ binding site by performing EMSAs. The GST-Su[H] fusion protein clearly shifted the $Su[H]^{bind}$ probe and addition of unlabeled $Su[H]^{bind}$ probe effectively competed this binding, while binding of Su[H] was not affected by the addition of an 80-fold excess of unlabeled, mutated probe (Figure 4-4b, modified from White et al. (2005)). Su[H] thus directly bound a probe matching the sequence found within the known *Tbx6* enhancer. These results therefore further supported our hypothesis that RBP-Jκ, the main transcription factor of the Notch signaling pathway, was involved in the regulation of *Tbx6* expression.

4.3.3 Mutation of the predicted RBP-Jκ binding site in a *Tbx6* PSM::*lacZ* transgene results in loss of reporter gene expression

To test the importance of RBP-Jκ in the regulation of *Tbx6* expression *in vivo*, we mutated the RBP-Jκ binding site in the $Tbx6^{Tg122/lacZ}$ construct by changing the guanine residue to a cytosine as described above and generated three transgenic lines with this $Tbx6^{Tg122\text{-}Su[H]mut/lacZ}$ construct (Figure 4-5b). The $Tbx6^{Tg122/lacZ}$ was normally expressed in the PSM, commencing at e9.5 and persisting in the caudal halves of the newly formed somites (Figure 4-5d-e). $Tbx6^{Tg122\text{-}Su[H]mut/lacZ}$

transgenic embryos were dissected at e9.5 to e12.5 and assayed for β-galactosidase activity. One line showed no staining, while the other two lines showed very limited staining. Mutation of this single base in the transgene effectively eliminated expression of the *lacZ* reporter from the PAM, except for approximately 20 cells that were scattered in the somites of the hindlimb bud region from e9.5 to e11.5 (Figure 4-5f-h). Given these results, we proposed that RBP-Jκ was critical for expression of *Tbx6* in the PSM. However, there was mounting evidence that the Notch signaling pathway may also be functioning downstream of Tbx6. Therefore, my focus now turns to *Dll1*, a Notch ligand, that we hypothesize is a direct downstream target of Tbx6.

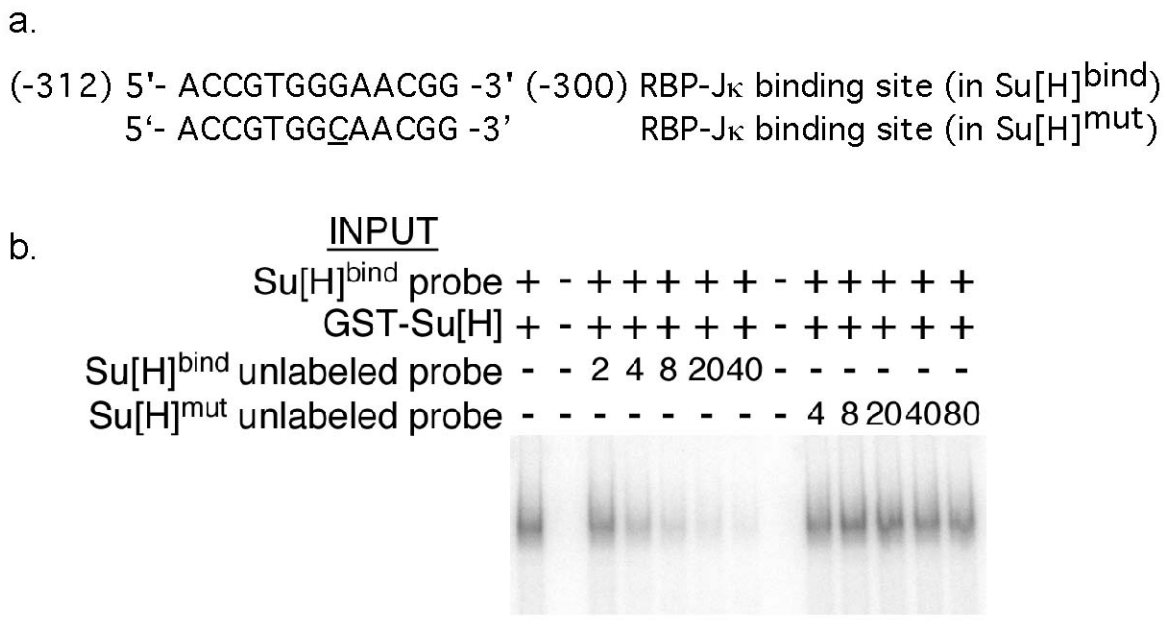


Figure 4-4: RBP-Jκ binds to sequences found in the *Tbx6* enhancer

Figure modified from White et al. (2005). (a.) Sequences of the RBP-Jκ binding site found in the *Tbx6* enhancer and the mutated site (guanine to cytosine, underlined 'C') used to generate mutant probes. (b.) EMSAs used a GST-Su[H] fusion protein to shift the radiolabelled 36 bp Su[H]^{bind} probe. This binding was clearly competed by the addition of excess unlabeled Su[H]^{bind} probe, but not by the addition of excess unlabeled Su[H]^{mut} probe containing the single G to C mutation. The numbers indicate the molar ratio of excess unlabeled probe added.

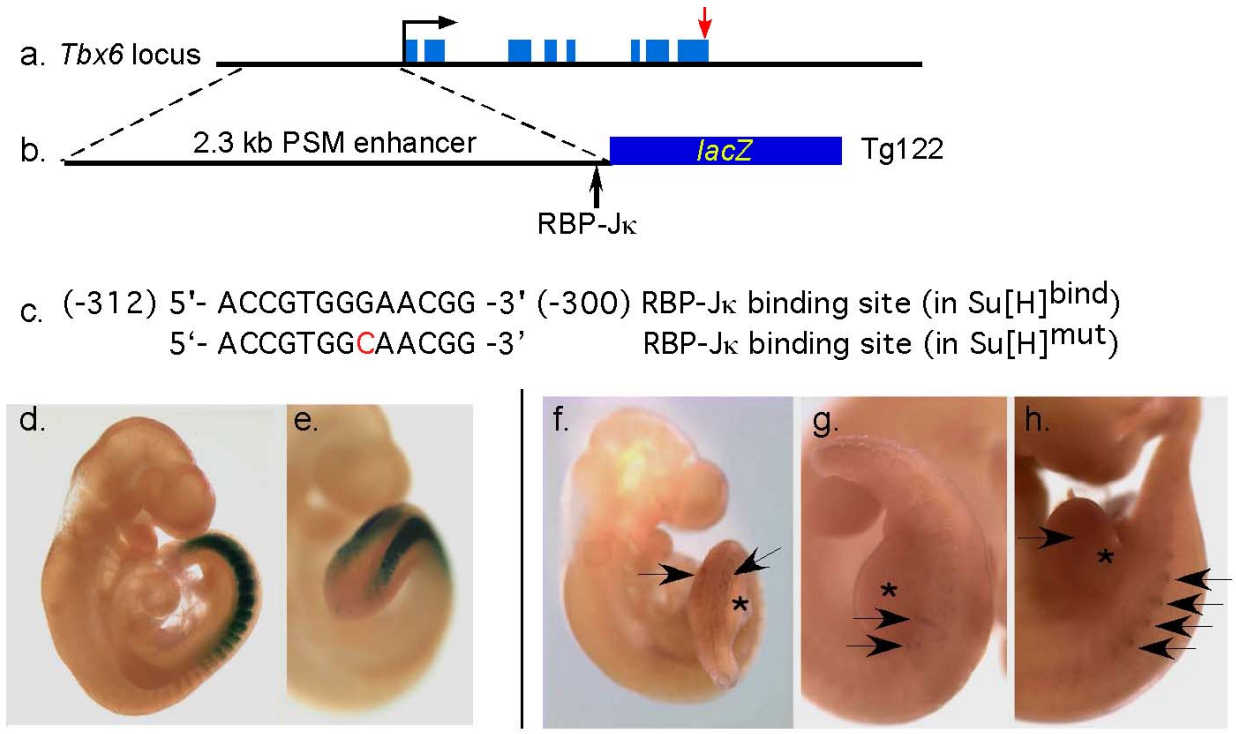
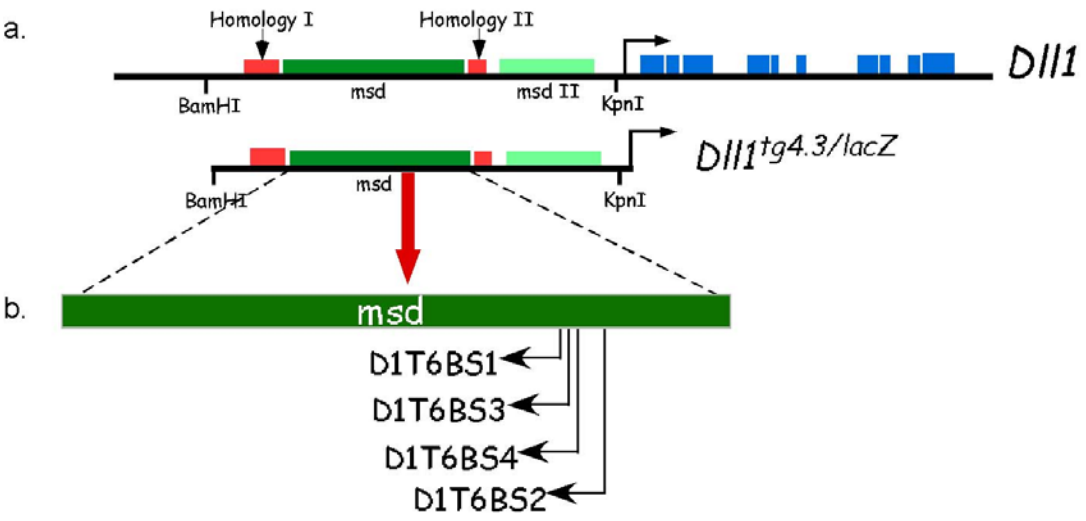


Figure 4-5: Paraxial mesoderm expression of *Tbx6* is dependent on the transcription factor RBP-Jκ.

Figure modified from White et al. (2005). (a.) Schematic of the *Tbx6* genomic locus with the positions of the start of transcription (black arrow), exons (blue boxes) and stop codon (red arrow). (b.) The 2.3 kb PSM enhancer from the *Tbx6* genomic region and the *lacZ* reporter cassette used to generate the $Tbx6^{Tg122/lacZ}$ transgene is shown in relation to the *Tbx6* genomic locus. The location of the RBP-Jκ binding site is labeled. (c.) Sequence of the RBP-Jκ binding site found in the $Tbx6^{Tg122/lacZ}$ transgene and the sequence of the mutated site (guanine to cytosine, red 'C') used to generate the $Tbx6^{Tg122\text{-}Su[H]mut/lacZ}$ transgene. (d-e.) The tail region of an e9.5 $Tbx6^{Tg122/lacZ}$ transgenic embryo with X-gal staining present in the PSM and formed somites, becoming localized to the caudal halves of the somites. (f-h.) Tail regions of $Tbx6^{Tg122\text{-}Su[H]mut/lacZ}$ transgenic embryos dissected at e9.5 (f.), e10.5 (g.) and e11.5 (h.) and stained with X-gal. The hindlimb bud is indicated by an asterisk (*) and the arrows indicate the sparse cells staining positive for β-galactosidase activity.

4.4 TBX6 BINDS TO DNA SEQUENCES WITHIN A KNOWN *DLL1* PSM ENHANCER *IN VITRO*

The Notch ligand *Dll1* is expressed in the PS and PSM, later becoming localized to the caudal halves of the somites (Bettenhausen et al., 1995). *Dll1* is also expressed in a number of other tissues, including the neural tube, but for this work, we were specifically concerned with *Dll1* expression in the PS and PSM. *Dll1* enhancer elements capable of driving *lacZ* reporter expression in *Dll1*-specific domains had previously been identified (Beckers et al., 2000a). Figure 4-6a contains a schematic of the major *Dll1* enhancer elements that are described below; the figure summarizes work completed by Beckers et al. (2000a). The $Dll1^{tg4.3/lacZ}$ enhancer consisted of a 4.3 kb upstream region (relative to the *Dll1* translation start site). This enhancer region contained several *Dll1* enhancer elements, including two mesoderm enhancer elements, 'msd' and 'msdII', and two regions of homology to *DeltaD*, the zebrafish *Dll1* homolog, that functioned as enhancers for neural development (Beckers et al., 2000a). The 1.5 kb $Dll1^{tg'msd'/lacZ}$ enhancer functioned as a PAM enhancer and was active in the PSM from approximately e7.5 onward; interestingly, this enhancer did not completely mimic endogenous *Dll1* expression as it was not active in nascent somites (Beckers et al., 2000a). A second mesoderm-specific enhancer, 'msd II' or $Dll1^{tg1.6/lacZ}$, consisted of 1.6 kb upstream of the *Dll1* start of translation, and was active in the PAM as well as in the ventral neural tube at e10.5 (Beckers et al., 2000a). Given our previous results, specifically the loss of *Dll1* expression in the *Tbx6* mutant and the genetic interaction between *Dll1* and *Tbx6*, we sought to further define the role of Tbx6 in regulating *Dll1* expression.



c.

position #		1	2	3	4	5	6	7	8	9	10	11	12
Tbx6 consensus		A	G	G	T	G	T	B	R	N	N	N	N
D1T6BS1		A	G	G	T	G	T	T	G	C	T	G	T
D1T6BS2		A	G	G	T	G	T	G	A	T	T	C	T
D1T6BS3		A	G	G	T	G	T	C	c	T	C	A	C
D1T6BS4	A	g	G	G	T	G	T	a	c	C	T	G	A
T half-site		A	G	G	T	G	T	G	A	A	A	T	T

Figure 4-6: Putative Tbx6 binding sites within the *Dll1* msd enhancer

Entire figure modified from White and Chapman (2005). (a) Schematic of the *Dll1* genomic locus and *Dll1*$^{Tg4.3}$ enhancer summarizing data from Beckers et al. (2000a). The positions of the start of transcription (black arrow) and exons (blue boxes) are shown. The 'msd', 'msd II', and two regions of homology are marked in the *Dll1* enhancer. (b) The 1.5 kb 'msd' enhancer from the *Dll1* genomic region is expanded and contains four putative Tbx6 binding sites (D1T6BS1-4); their positions and orientations (arrows) are indicated. (c) The sequences of the four putative Tbx6 sites are given along with the Tbx6 consensus site and the T half-site. Lower case letters indicate bases not matching the Tbx6 consensus sequence, B represents G, T or C, and R represents A or G.

Using the Ensembl mouse genome sequencing database, I determined that the *Dll1* enhancer contained four putative Tbx6 binding sites clustered within 100 bp of each other in the *Dll1*$^{tg'msd'/lacZ}$ enhancer (Figure 4-6b), however no sites matching the Tbx6 consensus sequence were found in the 'msdII' (*Dll1*$^{tg1.6/lacZ}$) enhancer. Within the *Dll1*$^{tg'msd'/lacZ}$ enhancer, the putative Tbx6 binding sites D1T6BS1, D1T6BS2, D1T6BS3, and D1T6BS4 were located at -1997, -1920, -1982 and -1968 respectively, relative to the start of transcription for *Dll1*, with all sites found in the same orientation relative to each other; D1T6BS1 and D1T6BS2 matched the Tbx6 consensus sequence, while D1T6BS3 and D1T6BS4 contained slight mismatches (Figure 4-6). We generated double stranded, radioactive oligo probes containing each of these sites for EMSAs to test whether they were bona fide Tbx6 binding sites. As expected, Tbx6 bound to D1T6BS1 and D1T6BS2, both of which matched the Tbx6 consensus sequence (Figure 4-7). Addition of the N-terminal Tbx6 antibody supershifted these protein/DNA complexes to the top of the gel and confirmed that the shift was specific to the Tbx6 protein in the lysate. Tbx6 failed to bind D1T6BS3 or D1T6BS4, both of which contained mis-matches in the Tbx6 consensus sequence. D1T6BS3 had a C at position 8 replacing the A or G in the Tbx6 consensus sequence. D1T6BS4 contained an extra G within the Tbx6 "core" binding site of 5'-AGGTGT-3', and even if the extra G were permitted, it also contained an A in position 7. Tbx6 binding to these sites *in vitro* was very weak as determined by the presence of a very faint band in DNA/protein complex by EMSAs (Figure 4-5b). Therefore, Tbx6 had the ability to bind putative Tbx6 binding sites within a known *Dll1* mesoderm enhancer *in vitro*. Thus, together with genetic studies, namely loss of *Dll1* expression in *Tbx6* mutants and the genetic interaction between *Tbx6* and *Dll1*, we predicted that *Dll1* was a direct downstream target of Tbx6 (Chapman and Papaioannou, 1998; White and Chapman, 2005).

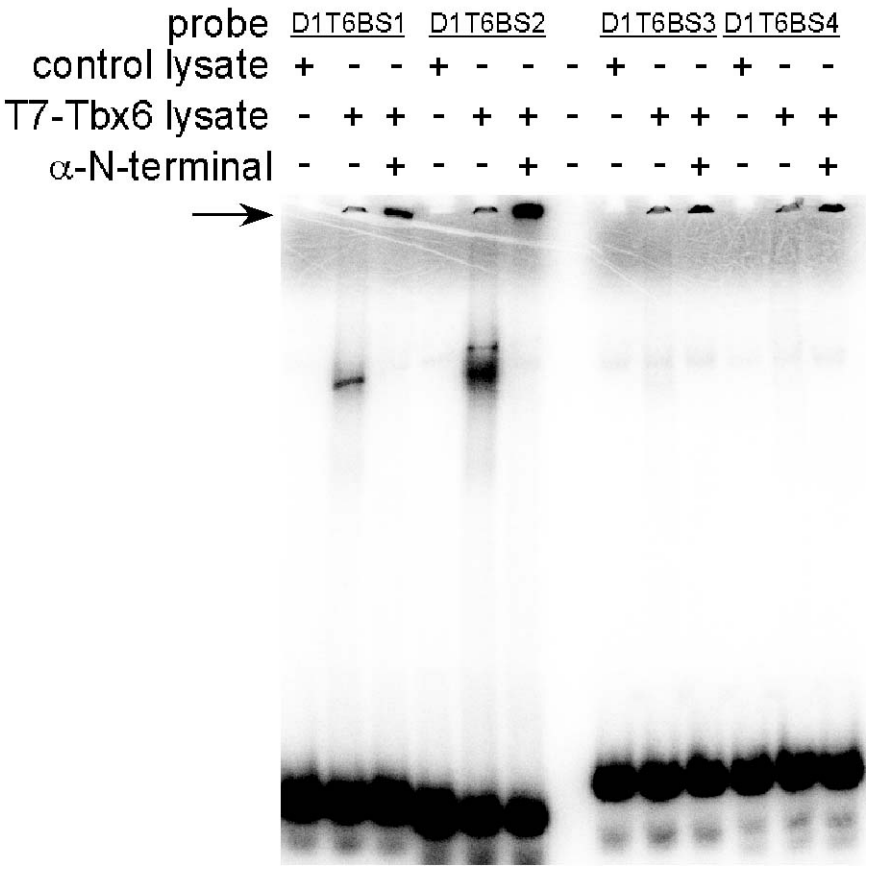


Figure 4-7: Tbx6 binds to *Dll1* enhancer sequences *in vitro*

Figure published in White and Chapman (2005). *In vitro* translated T7-Tbx6 or unprogrammed control lysate was used for EMSAs of the putative Tbx6 binding sites. EMSAs demonstrated Tbx6 binding to D1T6BS1 and D1T6BS2 with specificity of binding confirmed by supershifting the protein/DNA complex with anti-N-terminal Tbx6 antibody (arrow). Tbx6 binding to D1T6BS3 and D1T6BS4 was very weak in comparison. A faint background band was present in all lanes including the negative control unprogrammed lysate.

4.5 DISCUSSION

4.5.1 Upstream effectors of *Tbx6 in vivo*

Our lab had previously identified a 2.3 kb PSM enhancer located from 2.5 kb to 200 bp upstream of the start of *Tbx6* transcription that could drive reporter gene expression in the PSM (White et al., 2005). Herein, I present data indicating that *Tbx6* lies downstream of Notch signaling. I compared the above mouse *Tbx6* enhancer to those of the known *Tbx6* mammalian orthologs from human, rat and dog. These comparisons revealed three separate regions of homology, containing greater than 75% nucleotide similarity (Figure 4-3). Located within one of these homologous regions at -312 to -300 bp, relative to the *Tbx6* transcription start site, was a binding site for RBP-Jκ, the main, if not the only transcription factor functioning in the Notch-signaling pathway in mice (Tun et al., 1994). I verified that this site was indeed a binding site for RBP-Jκ using its *Drosophila* ortholog Su[H] by performing EMSA experiments (Figure 4-5b). We then showed that the binding site for RBP-Jκ was critical for expression of *Tbx6* in the PSM. Normally, the $Tbx6^{Tg122/lacZ}$ construct was expressed in the PSM and the somites beginning at e9.5; however, nearly complete abrogation of reporter gene expression was observed in embryos harboring the $Tbx6^{Tg122\text{-}Su[H]mut/lacZ}$ transgene, which differed only by containing the single G to C nucleotide mutation in the predicted RBP-Jκ site (White and Chapman, 2005). Because the $Tbx6^{Tg122/lacZ}$ construct was only robustly expressed in the PSM after e8.5, we could only assess the role of RBP-Jκ in maintaining *Tbx6* expression *in vivo*, but not whether RBP-Jκ was required for the initial expression of *Tbx6* in the PSM and what role RBP-Jκ played in controlling *Tbx6* expression in the PS. To answer these questions, I generated a construct, the $Tbx6^{Tg46/lacZ\text{–}Su[H]}$ construct, containing the single nucleotide mutation in the RBP-Jκ site; the *Tg46* reporter, without the RBP-Jκ mutation, mirrored endogenous Tbx6 expression both temporally and spatially. This $Tbx6^{Tg46/lacZ\text{–}Su[H]}$ construct is currently awaiting microinjection.

4.5.2 Downstream effectors of *Tbx6 in vivo*

In a *Tbx6* mutant, expression of *Dll1* was no longer detectable at e9.5 while *T* was expressed throughout the expanded tail region (Chapman and Papaioannou, 1998). Despite still being expressed in the *Tbx6* mutant, T was unable to activate *Dll1* expression. We therefore favor the hypothesis that Tbx6, as opposed to T, is the T-box gene functioning directly upstream of *Dll1* in the PSM. Based partially upon work described in this thesis, although T and Tbx6 could bind identical sequences *in vitro*, the selected Tbx6 binding sites showed specific preferences that were different from that of T; in addition, Tbx6 bound sequences as a monomer and T required an imperfect palindromic sequence (Kispert and Herrmann, 1993; White and Chapman, 2005). Work in this thesis also suggested that T and Tbx6 together failed to affect target gene expression when co-expressed in a reporter assay (Chapter 3.7). Herein, I present data showing Tbx6 binds directly to *Dll1* 'msd' enhancer sequences. As mentioned previously in this chapter, transgenic studies determined that two mesoderm-specific enhancers existed for *Dll1* within its known 4.3 kb enhancer region upstream of the *Dll1* translation start site (Beckers et al., 2000a). The 1.5 kb 'msd' enhancer functioned as a PAM enhancer; it drove expression of a *lacZ* reporter in the PSM from the PS stage onward, but was never active in the somites, another *Dll1* expression domain (Beckers et al., 2000a). There were four putative T-box binding sites clustered within 100 bp of each other in the 'msd' enhancer (Figure 4-6a), with no sites found in the tg1.6 'msd II' mesodermal enhancer (Beckers et al., 2000a). Tbx6 bound to the two sequences D1T6BS1 and D1T6BS2 *in vitr*o, while Tbx6 weakly recognized the D1T6BS3 and D1T6BS4 miscognate sites using EMSAs.

A report by Hofmann et al. (2004) also identified one of these sites, D1T6BS2 (their site T5), as a T-box binding site and showed that mutation of this site in either the full 4.3 kb enhancer or in the 'msd' enhancer resulted in the loss of reporter gene expression in the tailbud. They never tested the D1T6BS1 as a T-box binding site; thus, the importance of this binding site for the regulation of *Dll1* expression was never determined *in vivo*. While T bound *in vivo* to sequences containing the same 5'-AGGTGT-3' core sequence, binding site selection studies for T showed that there were preferences for G in position 7 and that position 8 was always A in the T consensus sequence (Casey et al., 1998; Kispert and Herrmann, 1993). D1T6BS2 matched the T half-site, however D1T6BS1 contained a T in position 7 and a G in position 8 and therefore

was unlikely to be a binding site for T. On the other hand, our studies for Tbx6 showed that position 7 was anything but A and that position 8 was either A or G (see Chapter 3). Casey et al. (1998) never observed Xbra (the *Xenopus* T ortholog) bound to a sequence with a G substituted at position 8. Therefore, our *in vitro* and *in vivo* data supported the hypothesis that *Tbx6* was the T-box protein that functioned directly upstream of *Dll1* to control expression of *Dll1* in the PSM.

4.5.3 Chapter conclusions

We have shown that the spontaneous mutation, *rv*, reduced *Tbx6* expression below heterozygous levels and resulted from a mutation in *Tbx6* (White et al., 2003). No difference in sequence was detected between the *Tbx6* transcripts expressed in *rv/rv* embryos compared to the known *Tbx6* cDNA sequence. Therefore, we determined that the *rv* mutation consisted of an insertion of 185 bp into the 5' enhancer region of *Tbx6*; this inserted DNA actually represented an inverted insertion of a duplicated portion of *Tbx6*, including part of the first exon, all of the first intron, and a portion of the second exon, while the *Tbx6* gene was still intact (Figure 4-1b). The *rv* mutation thus represented a hypomorphic mutation of *Tbx6* by reducing *Tbx6* expression (White et al., 2003).

Because the *rv* mutation was caused by an insertion into a known *Tbx6* PSM enhancer, we hypothesized that there were sequences within this *Tbx6* enhancer important for controlling the temporal and spatial expression of *Tbx6*. We used EMSA experiments and site directed mutagenesis to show that RBP-Jκ, a Notch signaling pathway transcription factor, was responsible for maintaining *Tbx6* expression in the PSM. Lastly, we previously knew that *Dll1* expression was absent in a *Tbx6* mutant and that *Tbx6* and *Dll1* genetic interacted; this led us to test whether *Tbx6* was directly upstream of *Dll1* (Chapman and Papaioannou, 1998; White et al., 2003). We showed that Tbx6 bound to sequences within a known mesoderm enhancer of *Dll1* (White and Chapman, 2005). The data presented in this chapter suggested that Notch signaling was likely to be functioning both upstream and downstream of *Tbx6* in the pathway that led to somite formation.

5.0 CONCLUSIONS AND FUTURE DIRECTIONS

Experiments detailed in this thesis described how Tbx6 functioned as a DNA binding protein and transcription factor, how RBP-Jκ maintained *Tbx6* expression in the PSM, and how *Tbx6* recognized downstream targets like *Dll1*. Therefore, much has been learned about how Tbx6 functions, however many new questions have arisen from this work.

5.1 TBX6 PROTEIN SEQUENCE CAN BE USED TO FURTHER CHARACTERIZE THE FUNCTION OF TBX6

The Tbx6 protein has been confirmed to be 58 kDa in tailbud tissue, with no other proteins detected by Western blot analysis using two Tbx6-specific antibodies (Chapter 2.4). This confirmed that our initial Tbx6 protein sequence, predicting the larger 58 kDa Tbx6 protein, was correct at this developmental stage. Studies can now be pursued to further characterize Tbx6 functional domains. There are two predicted domains within Tbx6, namely a PEST sequence at AA 320-331 and a leucine zipper motif within its C-terminus (AA 436-458).

Using the PESTFIND program, one potential PEST sequence was found in the Tbx6 protein (AA 320 to 331); PEST sequences are sequences enriched in proline (P), glutamic acid (E), serine (S), and threonine (T), which target the PEST-containing protein for rapid degradation (reviewed in Rechsteiner and Rogers, 1996). In theory, this could explain the tight temporal and spatial control of Tbx6 expression that restricts Tbx6 protein to the PSM while preventing its presence in the somites. Other reports have recently found that transcription factors were regulated by PEST sequences. The microphthalmia-associated transcription factor (MITF) was thought to promote the entrance of neural crest cells into the melanocyte differentiation pathway (Opdecamp et al., 1997). Recent work suggested that the PEST region of MITF was regulated

by the ubiquitin-proteasome system because a serine to alanine mutation within the PEST region prevented degradation of MITF (Xu et al., 2000). In addition, the human AT-hook transcription factor, AKNA, and the C_2H_2 zinc finger transcription factor, EKLF1/KLF1, were also regulated through PEST sequences (Quadrini and Bieker, 2006; Sims-Mourtada et al., 2005). Since Tbx6 protein expression mirrored that of the tightly regulated mRNA, particularly its down regulation upon somite formation, future studies could focus on the importance of this PEST sequence for Tbx6 degradation.

The Tbx6 protein sequence was analyzed using the MacVector program to detect various protein patterns based on the PROSITE database (Hulo et al., 2006); computer prediction revealed the presence of a putative leucine zipper located at AA 436 to 458 (Busch and Sassone-Corsi, 1990; Landschulz et al., 1988). The leucine zipper motif is interesting because most T-box proteins have no other domains or motifs aside from the T-box that can be predicted by examining their protein sequence. A leucine zipper is a protein motif involved in protein-protein interactions. The presence of this motif in Tbx6 may hint at possible interacting partners that would have functional implications for Tbx6 in the developing embryo. This would be an interesting outcome as one could envision either (1) Tbx6 forming homodimers when binding DNA, or (2) Tbx6 binding to another protein containing a leucine zipper motif, either blocking Tbx6 homodimerization or forming Tbx6-another protein heterodimers thereby regulating Tbx6 function. I have already cloned the portion of the *Tbx6* cDNA matching the region predicted to have a leucine zipper motif, along with a site directed mutant of this motif, in-frame with GST. These constructs will prove useful in future studies for the determination of whether the leucine zipper motif is used for homodimerization of Tbx6.

Most characterized T-box proteins do not contain predicted domains/motifs outside of the DNA binding domain. Characterization of these proteins has focused on identifying the NLS and activation/repression domains primarily through protein deletion strategies in tissue culture (Carlson et al., 2001; Kispert et al., 1995a; Zaragoza et al., 2004). Similar "structure/function" experiments with Tbx6 may reveal why Tbx6 functions only as a weak activator in the luciferase assays (Chapter 3.5 and Yasuhiko et al., 2006). In addition to identifying activation/repression domains in T-box proteins, it may now be possible to identify interacting proteins; these interacting proteins may aid in the activation or repression of downstream targets. In addition to determining whether the leucine zipper motif functions in Tbx6, a yeast two-hybrid system could

function to identify interacting proteins and was used to identify Tbx5 as an interacting partner of Nkx2-5 (Hiroi et al., 2001). Other members of the lab are currently taking this approach with Tbx6.

5.2 SEVERAL MAJOR SIGNALING PATHWAYS ARE THOUGHT RESPONSIBLE FOR THE CORRECT *TBX6* EXPRESSION

This document had shown that RBP-Jκ, the Notch signaling transcription factor, was necessary for maintenance of *Tbx6* expression in the PSM. As discussed in Chapter 4.5.1, I have already generated a construct to test whether RBP-Jκ was necessary for initial expression of Tbx6 in the PSM. This construct, the $Tbx6^{Tg46/lacZ–Su[H]}$ construct, contained the single nucleotide mutation in the RBP-Jκ site, but within the *Tg46* reporter. The *Tg46* reporter mirrored endogenous Tbx6 expression both temporally and spatially. The $Tbx6^{Tg46/lacZ–Su[H]}$ construct is currently awaiting microinjection.

With the advent of genome sequencing databases, on-line transcription factor databases, and sequence search engines, it is possible to obtain sequence information for your favorite gene from a large array of organisms. We can obtain *Tbx6* genomic sequences from human, mouse, rat, chimpanzee, and dog, and use this information to identify putative upstream regulators of *Tbx6* expression. Alignment of Tbx6 genomic sequences reveals several transcription factor binding sites located within the conserved regions of the *Tbx6* PSM enhancer for HMG box transcription factors. The HMG box transcription factors, TCF1/LEF1, are responsible for canonical Wnt signaling (reviewed in van Noort and Clevers, 2002). Interestingly, a TCF/LEF binding site was also found in a conserved region upstream of *T* and has been shown to be necessary for maintenance of *T* expression in the PS through Wnt3a. Although initial expression of *Tbx6* is not dependent on Wnt3a, once the *Wnt3a* phenotype becomes apparent, *Tbx6* expression does require *wnt3a* signaling (Arnold et al., 2000; Yamaguchi et al., 1999). Therefore, Wnt signaling through TCF1/LEF1 putative binding sites may be important for the proper maintenance of *Tbx6* expression.

5.3 DOWNSTREAM TARGETS OF TBX6 ARE AFFECTED IN MUTATIONS OF *TBX6*

Mice homozygous for the spontaneous mutation *rib-vertebrae* exhibit fusions of ribs and vertebrae and other phenotypes similar to incomplete rescue of the *Tbx6* phenotype by transgenics (Beckers et al., 2000b; White et al., 2003). We have shown that the *rv* mutation was caused by an insertion into a known *Tbx6* PSM enhancer that resulted in decreased *Tbx6* expression (Chapter 4.2). Thus *rv* represented a hypomorphic mutation of *Tbx6* (Chapter 4.2 and White et al., 2003). Reducing *Tbx6* transcripts below heterozygous levels resulted in altered expression of several genes, including several Notch signaling pathway members, including *Notch1*, *Dll1*, *Dll3*, *Mesp2*, and *Lfng* (Beckers et al., 2000b; White et al., 2003). This suggested that *Tbx6* had some function in the control of one or more of these target genes. Interestingly, we have shown that *Tbx6* genetically interacted with *Dll1* and Beckers et al. showed similar results for the *rv, Dll1* double heterozygous mutants (Beckers et al., 2000b; White et al., 2003). In a *Tbx6* mutant, expression of *Dll1* was no longer detectable at e9.5 while *T* was expressed throughout the expanded tail region (Chapman and Papaioannou, 1998). Others have hypothesized that T was required for *Dll1* expression in the PAM (Hofmann et al., 2004). Therefore, it would follow that in a *Tbx6* mutant, where *T* was still expressed, *Dll1* would also be expressed. Despite expression in the *Tbx6* null mutant, T was unable to activate *Dll1* expression. Therefore, we favor the hypothesis that *Tbx6* is the T-box gene functioning directly upstream of *Dll1* in the PSM.

As detailed in the discussion of Chapter 4, *Dll1* contained two mesoderm enhancer elements, $Dll1^{tg'msd'/lacZ}$ and $Dll1^{tg1.6/lacZ}$ (Beckers et al., 2000a). The 1.5 kb $Dll1^{tg'msd'/lacZ}$ enhancer functioned as a PAM enhancer and was active in the PSM from the PS stage onward, but the 'msd' enhancer was never active in the somites, another region of *Dll1* expression (see Figure 4-6a). Within this region, I uncovered four putative Tbx6 binding sites and determined that Tbx6 was capable of binding two of these sequences *in vitro*. It was published that one of these sequences was partially responsible for activating *Dll1* within the PSM, however the authors stated that either T or Tbx6 was responsible for activation of *Dll1* (Hofmann et al., 2004). Although much genetic data supported Tbx6 as the activator for *Dll1*, the authors claimed that activation was dependent on Wnt signaling in concert with signaling of a T-box

protein (Hofmann et al., 2004). In addition to the four putative T-box binding sites, the 'msd' enhancer also contained binding sites for TCF1/LEF1 that were required for the dual-signal outcome (Hofmann et al., 2004). Since T and Tbx6 have overlapping expression in the PS, either of these T-box proteins could be critical for *Dll1* mesoderm expression. As detailed above, the inability of T to maintain *Dll1* expression in the *Tbx6* mutant suggests that Tbx6 and not T is required for *Dll1* expression.

Also of interest is recent work from the Saga lab on the regulation of *Mesp2* expression, as both Tbx6 and Notch converge on this pathway. In these studies, Tbx6 was shown to bind miscognate sequences *in vitro* and *in vivo* in certain contexts (Yasuhiko et al., 2006). This was an important finding because it suggested that Tbx6 could stably bind multiple sites as long as one site matched our identified Tbx6 consensus sequence described in Chapter 3 (White and Chapman, 2005). Therefore, this could mean that Tbx6 formed a weak dimer that became stable when bound to DNA, but still allowed the protein to release the DNA to prevent tight binding on miscognate sites; this was originally predicted for T (Coll et al., 2002). Therefore, Tbx6 may perform a similar function at the *Dll1* 'msd' enhancer. I have designed constructs that contain site-directed mutations in each of the four putative Tbx6 binding sites, which could be tested in luciferase assays. However, as Tbx6 is a weak activator, these studies would require the use of an activated Tbx6 (Tbx6-VP16) to judge how mutation of each of the four sites would affect transcriptional activity.

Both *Dll1* and *Mesp2* are members of the Notch signaling pathway and both are targets of Tbx6. The next major step will be to identify other Tbx6 targets. The large amount of sequence information found in public databases may allow us to identify putative downstream genes more rapidly by using computer prediction and genomic alignments of orthologous gene enhancer regions for genome wide searches of conserved Tbx6 binding sequences. Rather than searching for binding sites alone, the work described here now suggests that Tbx6 binding sites should be flanked by either miscognate sequences or transcription factor binding sites for transducers of signaling pathways (such as Wnt or Notch); these searches would provide more accurate predictions of target genes rather than narrowly searching for the T consensus and half-site sequences alone.

5.4 DOES TBX6 INTERACT, DIRECTLY OR INDIRECTLY, WITH OTHER T-BOX PROTEINS DURING DEVELOPMENT?

Tbx5 and Tbx20, two T-box proteins grouped into different T-box subfamilies, could physically interact and acted synergistically to affect heart development in *Xenopus* and zebrafish (Brown et al., 2005). Because the T-domain was conserved and many T-box proteins bound to the T consensus sequence, it is fair to hypothesize that T-box proteins with overlapping domains of expression might bind to the same target sequences, thus setting up a direct competition between T-box proteins, like that in Tbx5/Tbx20 above and zebrafish *no-tail*, *spadetail*, and zf*tbx6* (Goering et al., 2003). A competition between T-box proteins could be skewed in the cell simply by overexpressing one of the T-box genes relative to the other(s). This could explain why such tight controls are placed on T-box gene expression (Brown et al., 2005; Goering et al., 2003).

One remaining possibility revolves around the T-box protein, Tbx18. During mouse and zebrafish development, Tbx18 expression was prominent in the heart, but it was also expressed in other sites, including the cranial PAM, genital ridge, limb buds, and PSM, becoming localized to the anterior halves of somites (Begemann et al., 2002; Kraus et al., 2001). Since Tbx6 is expressed in the PSM and tailbud, an intriguing possibility is that Tbx6 interacts with Tbx18 within the PSM. In addition, although the luciferase transcriptional assays described in Chapter 3.7 did not detect any effect on transcriptional activity when Tbx6 and T were co-expressed, it did not rule out the possibility that T and Tbx6 interacted in the PS because the signal to noise ratio was so low in these experiments. It should also be noted that Tbx6 and T may require activated important signaling pathways, such as the Wnt or Notch signaling pathway, to observe such a synergistic effect.

Does the function of T and/or Tbx18 depend upon the function of Tbx6? Ultimately this work needs to return to functional analysis of these proteins in the animal. Like most T-box proteins, Tbx6 and T play profound roles in development by controlling cell fate decisions that ultimately control the behavior of the cells in the organism. Results from the studies described in this thesis may be important for studies with cell fate decisions, such as coaxing stem cells down specific developmental paths or in gaining a better understanding of the myriad of diseases that occur because of breakdowns in these processes.

6.0 METHODS

6.1 BASIC TECHNIQUE FOR PCR AMPLIFICATION FROM VARIOUS PLASMID DNA SOURCES

All amplification of plasmid DNA for cloning of the various constructs was performed in a similar manner using the following PCR reaction conditions: 25 ng of plasmid source DNA, 1X PCR buffer [20 mM Tris, pH 8.4, and 50 mM KCl] (Invitrogen), 3 mM $MgCl_2$, 0.2 mM dNTPs, 10 pmol of each primer, and 1.25 U of *Taq* DNA polymerase (GeneChoice) in a 50 μL volume. The cycling conditions were as follows: 95°C for 5 minutes; 35 cycles of 94°C for 45 seconds, a varied annealing temperature depending on primers used for 45 seconds, and 72°C for a varied time depending on size; and a final extension at 72°C for 5 minutes. The annealing temperature, extension time, PCR primers, and DNA template were listed in appropriate sections. All primers were generated by Integrated DNA Technologies, unless otherwise stated.

6.2 CLONING AND EXPRESSION OF GST FUSION PROTEINS FOR ANTIBODY PRODUCTION

6.2.1 Design and cloning of GST-Tbx6 fusion constructs

Glutathione-S-transferase (GST) was utilized to make GST-Tbx6 fusion proteins as antigens from the following regions of the Tbx6 protein: amino acids (AA) 2-78 [N-terminal], AA 311-408 [Internal], and AA 467-537 [C-terminal]. These three regions were amplified from the *Tbx6* cDNA by polymerase chain reaction (PCR) using the primers listed below, an annealing

temperature of 55°C, and an extension time of 55 seconds. The *Tbx6*-specific primers included BamHI and EcoRI sites for cloning purposes. The sizes of the PCR products were verified on agarose gels according to standard procedures, and the products of 240 bp, 300 bp, and 250 bp, respectively, were purified using MinElute PCR purification columns (Qiagen, Valencia, CA). Purified PCR fragments were digested with BamHI and EcoRI, and then cloned in-frame with the C-terminus of GST using the BamHI and EcoRI sites within the pGEX-3X vector (Amersham Biosciences). The cloned expression plasmids were sequenced using the primers listed below, and then expressed in bacteria (see Section 6.2.2).

Primers

N-terminal region of Tbx6

pGEXN forward: 5'- ACAATGGATCCTAGTGTACCATCCACG -3'

pGEXN reverse: 5'- ACATATGAATTCTCTGCTGTCTCGGGG -3'

Internal region of Tbx6

intant forward: 5'- ACAATGGATCCTATGTGACTCCACC -3'

intant reverse: 5'- ACATATGAATTCTCAAGTGTGGGGC -3'

C-terminal region of Tbx6

pGEXC forward: 5'- ACAATGGATCCTAACCAAATGGCTGGCT -3'

pGEXC reverse: 5'- ACATATGAATTCTCGTAACAAGGCCC -3'

pGEX sequencing primers

pGEX seq forward: 5'- AATGTGCCTGGATGCGTTC -3'

pGEX seq reverse: 5'- TACACTCCGCTATCGCTACG -3'

6.2.2 Small scale expression and purification of GST-Tbx6 fusion proteins for solubility testing

GST-Tbx6 fusion proteins were expressed on a small scale in BL21 cells to determine solubility as described in Molecular Cloning (Sambrook and Russell, 2001). The GST-Tbx6 fusion

proteins were transformed into BL21 cells using standard procedures and plated onto NZCYM media supplemented with 2% glucose and 60 μg/mL ampicillin. A single colony was grown in 5 mL NZCYM media supplemented with 60 μg/mL ampicillin for approximately 2 hours at 30°C until the OD was approximately 1. At this time, a portion of the culture was retained as uninduced cells. Bacteria were pelleted and stored for later analysis at -20°C.

The remaining cultures were induced with 1 mM IPTG and grown for 2.5 hours at 30°C. Bacteria were chilled on ice, pelleted by centrifugation at room temperature (RT), and then either stored at -20°C as the induced sample or used for examining the solubility of the fusion proteins. To test for fusion protein solubility, bacteria were resuspended in 100 μL cold 1X lysis buffer plus protease inhibitors (Appendix A) and 0.2 mg/mL lysozyme (USB). Lysates were incubated on ice 10 minute, and then sonicated briefly on ice for bursts of five seconds each until they no longer appeared viscous. Lysates were centrifuged at 4°C for 10 minutes at 18,000 g to precipitate insoluble debris, and the supernatant (containing soluble proteins) was moved to a new tube.

The precipitated bacteria representing transferred uninduced and induced samples, along with the insoluble fraction, were resuspended in 200 μL of 1X SDS Sample buffer (Sambrook and Russell, 2001). The soluble lysates were mixed with an equal volume of 2X SDS Sample buffer. All samples were denatured in a boiling water bath for 4 minutes, centrifuged for 1 minute at RT to remove debris, and 10 μL of each sample was separated on 12.5% SDS-polyacrylamide gels using standard procedures (Sambrook and Russell, 2001). The gels were stained with Coomassie blue stain following standard procedures (Sambrook and Russell, 2001).

6.2.3 Large scale purification of all soluble GST fusion proteins

GST, the GST-Tbx6 soluble fusion proteins, and GST-Su[H] were all expressed on a large scale essentially as described in Molecular Cloning (Sambrook and Russell, 2001). The GST-Su[H] plasmid was a gift from C. Xu and R. Carthew (Northwestern University). Plasmids were transformed into BL21 cells using standard procedures and grown on NZCYM plates supplemented with 2% glucose and 60 μg/mL ampicillin. Single colonies were grown in 5 mL NZCYM selection media (NZCYM media which contained 2% glucose and 60 μg/mL

ampicillin) for approximately 8 hours at 37°C; the entire cultures were then diluted 1:10 in NZCYM selection media and grown 16 hours at 37°C. The cultures were diluted again 1:20 in NZCYM media with 60 μg/mL ampicillin only and grown at 37°C for 60 minutes, then at 30°C until the culture OD_{A600} readings were between 0.6 and 0.8. At this time, a portion of each culture (1 mL) was retained as uninduced cells. The fusion proteins were induced with 0.4 mM IPTG and the cultures grown for 2.5 hours at 30°C. Cultures were then chilled on ice, a portion of each culture (1 mL) was retained as induced cells, while the remaining cultures were centrifuged at 5000 g for 10 minutes at 4°C and pellets stored at -80°C.

Bacterial pellets were resuspended in 20 mL total volume cold 1X lysis buffer with protease inhibitors (Appendix A) per 500 mL culture media volume. Lysozyme (USB) was added to a final concentration of 0.2 mg/mL, and the lysates were incubated on ice for 30 minutes. The lysates were sonicated briefly on ice for bursts of five seconds each until no longer viscous, and centrifuged at 18,000 g for 20 minutes at 4°C to remove insoluble debris. Supernatants were centrifuged again at 18,000 g for 20 minutes at 4°C to ensure that no insoluble material remained. Soluble lysates were filtered through 0.22 μm filter discs (Pall Life Sciences) for purification.

Proteins of interest were purified from supernatants using the GST-Bind purification system (Novagen). The GST-Bind resin column was prepared following the manufacturer's protocol, using a 2 mL settled resin volume, and purification was performed at RT. The entire volume of soluble, filtered lysates were allowed to flow through the GST-Bind resin column, and then washed with 10 volumes of 1X GST-bind binding buffer (Appendix A). The purified GST or GST fusion proteins were eluted from the column using 3 column volumes of 1X GST-bind Elute buffer (Appendix A), and fractions were collected at 0.75 mL intervals. Fractions containing the protein of interest were determined by spot testing each fraction for the presence of protein. This was done by spotting 3 μL of each fraction to Whatman filter paper, staining with 1X Spot protein stain [45 % methanol, 9 % acetic acid, 0.05 % Coomassie Blue], followed by destaining using 1X Spot protein destain [10 % methanol, 7.5 % acetic acid]. Fractions containing protein appeared as dark spots after destaining. These protein-containing fractions were pooled and dialyzed against 1X PBS buffer, pH 7.4 (Appendix A). Proteins used as antigens in antibody generation were concentrated using Spectra/Gel absorbent (Spectrum Labs) after completing dialysis. Protein used for coupling antigen to CNBr-Sepharose resin for

antibody purification was subsequently dialyzed against 1X Coupling buffer, pH 8.3 (Appendix A). Protein used for Su[H] EMSAs were dialyzed against 1X Su[H] binding buffer (Appendix A) for 4 hours at 4°C, then dialyzed twice against 1X Su[H] binding buffer overnight at 4°C. Proteins were stored at -80°C in aliquots until used in EMSAs.

6.3 ANTI-TBX6 ANTIBODY GENERATION AND PURIFICATION

6.3.1 Generation of Tbx6 antibodies in rabbits

The Tbx6 N-terminal, Internal, and C-terminal GST fusion proteins were used as antigens for antibody production in rabbits by Cocalico Biologicals (Reamstown, PA). The soluble antigens (N-terminal and Internal fusion proteins) were previously concentrated to approximately 4 mg/mL in 1X PBS, pH 7.4 (Section 6.2.3). The insoluble C-terminal GST fusion protein was expressed in bacteria; bacteria were then lysed and sonicated on a large scale, as described above for soluble GST fusion proteins, however the pellet contained the GST-C-terminal fusion protein. The pellet was resuspended in 500 μL of 1X SDS Sample buffer (Standard, Sambrook and Russell, 2001), heated in a boiling water bath for 4 minutes, and centrifuged for 1 minute at RT to remove debris. 50 μL of the supernatant was electrophoresed on a 12.5% SDS-polyacrylamide gel using standard procedures and prestained markers (Sambrook and Russell, 2001). The gel was stained with a modified Coomassie blue stain (Appendix A) and the major band of approximately 33 kDa was excised. This gel slice was sent intact to Cocalico Biologicals for injection into rabbits.

Each fusion protein was injected into two different rabbits to control for animal variability. The six rabbits were boosted after 14, 21, and 49 days and bled following the company's standard protocol, with two test bleeds at 35 and 56 days; after 56 days, the animals were boosted every 3 weeks. A larger production bleed to purify antibodies from the antisera occurred after approximately day 87, and a terminal bleed was performed at approximately day 126. The antisera were given the designations anti-T6N143, N144, I145, I146, C147, and C148, where the N, I, and C denote the N-terminal, Internal, and C-terminal antigens used, resp.

6.3.2 Purification of Tbx6 antibodies from terminal bleeds

Purification of Tbx6 antibodies was performed essentially as described (Harlow and Lane, 1999). However the full details, including any changes, are provided below.

6.3.2.1 Preparation of affinity purification columns containing GST and GST-Tbx6 fusion proteins for antibody purification

GST, GST-N-terminal antigen, and GST-Internal proteins were coupled to CNBr-Sepharose 4B (cat# 17-0430-01, Amersham) essentially as described by the manufacturer. Approximately 10 mg (4 mL) of the protein, dialyzed in 1X Coupling buffer (Appendix A), was mixed with the prepared resin and incubated overnight with rocking at 4°C. The mixture was centrifuged at 1500 g for 4 minutes, and the supernatant was discarded. The CNBr resin was washed twice in 10 bed volumes of 1X Coupling buffer, then the active groups on the resin were blocked with 10 bed volumes 1X Active group blocking buffer [0.1 M Tris-HCl, pH 8] at RT for 3 hours with rocking. The resin was then washed twice in 10 bed volumes of 1X Coupling buffer, and transferred to a gravity-flow column (Novagen). The column drained completely to pack with a final bed volume of 2.5 mL and contained approximately 10 mg of antigen.

To remove all loosely bound antigen, the column was washed through the following series of buffers, checking that the pH of the last few drops of eluate reached the pH values of the respective buffers (all listed in Appendix A): (1) 10 bed volumes of 1X Coupling buffer (pH 8.3), (2) 10 bed volumes of Low pH elute buffer (pH 2.5), (3) 10 bed volumes of 1X Coupling buffer (pH 8.3), (4) 10 bed volumes of High pH elute buffer (pH 11.5), (5) 10 bed volumes of 1X Coupling buffer (pH 8.3). After completing all wash steps, the five washes were then repeated to ensure that all loosely bound proteins were removed. These CNBr columns containing GST, GST-N-terminal, or GST-Internal were then ready for antibody purification.

6.3.2.2 Purification of antibodies from antisera

Approximately 10 mL of the various antisera were diluted 1:3 in 0.1 M Tris-HCl, pH 8, and filtered through 0.22 μm filter discs before loading onto the affinity purification CNBr-coupled columns. To eliminate GST-specific antibodies from the various rabbit antisera, each antisera was first loaded onto a CNBr-GST column. The entire eluate from this column was collected and

loaded onto its GST-antigen specific column. (This CNBr-GST column had GST-specific antibodies removed by subsequently following the steps listed below. All wash solutions given in Appendix A.) CNBr-antigen columns were washed with 20 bed volumes of 1X Coupling buffer, checking that the last few drops of the eluting wash buffer reached pH 8.3. Tbx6-specific antibodies were eluted from columns using acid/base elution of antibodies as described below. The acid-sensitive antibody-antigen interactions were eluted with 10 volumes Low pH elute buffer, pH 2.5. The entire eluate was collected in a tube containing 1 bed volume of 1 M Tris, pH 8. Following collection, the pH of the eluate was adjusted to approximately 7.5 with additional 1 M Tris, pH 8 and the eluate was stored at 4°C until further use. The column was washed with 20 bed volumes of 1X Coupling buffer, checking that the pH of the last few drops of the eluting wash buffer reached pH 8.3. The base-sensitive antibody-antigen interactions were eluted with 10 bed volumes High pH elute buffer, pH 11.5. The entire eluate was collected in a tube containing 1 bed volume of 0.5 M Tris pH 6.8. Following collection, the pH of the eluate was adjusted to approximately 7.5 with additional 2 M Tris, pH 7.5 and the eluate was stored at 4°C until further use. Columns were washed with 20 bed volumes of 1X Coupling buffer, checking that the last few drops of the eluting wash buffer reached pH 8.3; the columns were stored at 4°C in 1X Coupling buffer until further use. Each antisera resulted in two eluates, one at low pH and one at high pH; these two stored eluates, containing purified antibodies, were combined and used for final antibody purification and concentration as described below.

6.3.2.3 Final Antibody Purification and Concentration

For final antibody purification, a Protein A Sepharose column was prepared following manufacturer directions (17-0780-01, Amersham, Piscataway, NJ). Protein A Sepharose binds antibodies, therefore it will allow for concentration of the eluted antibody obtained in Section 6.3.2.2. Briefly, the Protein A Sepharose was expanded in deionized water in a 15 mL conical centrifuge tube, washed for 3 washes of 10 mL 1 mM HCl, with 10 minutes rocking followed by centrifugation at 5000 g for 5 minutes. The Protein A Sepharose resin was loaded onto a gravity-flow column (Novagen) and allowed to drain completely, packing the column with a final bed volume of 1.5 mL Protein A Sepharose resin; the column was washed with 10 bed volumes 1X PBS, pH 7.4 (Appendix A) to equilibrate the column.

To concentrate the antibody, the mixed eluates obtained above containing the purified Tbx6-specific antibodies were loaded onto the Protein A Sepharose column. The column was washed with 10 bed volumes of 1X PBS buffer, checking that the pH of the last few drops of the wash buffer reached pH 7.4. The antibodies were eluted by washing with 5 volumes low pH elute buffer, pH 2.5. The eluate was collected in fractions of 0.75 mL into microfuge tubes containing 100 μL of 1 M Tris, pH 8. Upon completion, the pH of the eluate was adjusted to approximately 7.5.

Fractions containing the antibodies were determined by spot testing each fraction for the presence of protein as described in Section 6.2.3. The fractions containing protein were pooled (approx 3 mL total volume); to remove any remaining GST-specific antibodies, the antibody was mixed with 50 μL of GST coupled to CNBr-Sepharose and incubated overnight at 4°C with rocking. The antibodies were centrifuged at 1500 g for 5 minutes and the supernatant was dialyzed against 1X PBS, pH 7.4, for 4 hours at 4°C, then dialyzed twice overnight against 1X PBS, pH 7.4 at 4°C. Upon completing dialysis, the dialyzed antibody was concentrated using Spectra/Gel absorbent (Spectrum Labs) and then stored in 1X PBS, pH 7.4 with 1% BSA added as a stabilizer. Antibodies were sterile filtered through a 0.22 μm filter disc, aliquoted, and the purified antibodies were stored at 4°C; they were stable for approximately 18 months.

6.4 PROTEIN PRODUCTION BY *IN VITRO* TRANSLATION USING RABBIT RETICULOCYTE LYSATES

6.4.1 Generation of T-box expression plasmids and *in vitro* translation of T-box proteins

The expression plasmid for the T7-tagged Tbx6 fusion protein was previously constructed by D. L. Chapman and contained the full-length *Tbx6* cDNA (p20.1) cloned into the pET21a expression vector (Novagen) in-frame with a T7-tag at its N-terminus. The *Tbx6* cDNA was digested with AflIII, filled in with Klenow, and then digested with HindIII to release the full length coding sequence. This product was then cloned into the pET21a vector, which was digested with BamHI, blunt ended with Klenow, and then digested with HindIII. The plasmid

for the production of Xbra protein with a C-terminal V5 epitope tag and a His tag in the pcDNA3.1/V5-His vector was a kind gift by F. Conlon (UNC, Chapel Hill). The coupled TNT *in vitro* transcription/translation rabbit reticulocyte lysate system (Promega) was used to produce the full-length fusion proteins following the manufacturer's protocol using 1 μg of purified plasmid in a 50 μL final volume. These proteins were used for all Tbx6 DNA binding studies using EMSAs and for all radiolabelled Tbx6 immunoprecipitations. *In vitro* coupled transcription and translation of proteins was performed for unprogrammed, T7-Tbx6-programmed, and Xbra programmed lysates in EMSAs. In addition, *in vitro* coupled transcription and translation of proteins was performed in the presence of ^{35}S-methionine and included unprogrammed, T7-Tbx6-programmed, Xbra, or luciferase-programmed lysate.

6.4.2 Immunoprecipitation of radiolabelled proteins

Samples were translated using the coupled TNT *in vitro* transcription/translation rabbit reticulocyte lysate system as described above using EasyTag ^{35}S-methionine (Perkin Elmer). The radiolabelled T7-Tbx6, Xbra, luciferase, and unprogrammed lysates were used for immunoprecipitation experiments. 20 μL of each translation reaction was added to 250 μL of 1X Lysis buffer (Appendix A). The samples were precleared by the addition of 1 μL of pre-immune antisera for 1 hour on ice followed by the addition of 10 μL of washed Pansorbin cells (Calbiochem, EMD Biosci) for 30 minutes on ice. Samples were centrifuged for 2 minutes at 18,000 g at 4°C, and the supernatants were transferred to new tubes for immunoprecipitation. Samples were preincubated for 2 hours with either pre-immune sera, purified anti-Tbx6 N144 (N-terminal antibody) or purified anti-Tbx6 I146 (Internal antibody). The antibody-protein complexes were precipitated with 20 μL washed Protein A Sepharose 4B (17-0780-01, Amersham, Piscataway, NJ). Complexes were washed four times using 500 μL 1X Lysis buffer, then the precipitated complexes were denatured by adding 100 μL of 1X SDS-PAGE sample buffer followed by boiling in a water bath for 5 minutes. Samples were centrifuged for 1 minute at 18,000 g, and the supernatants were separated on 10% SDS-polyacrylamide gels using standard procedures (Sambrook and Russell, 2001). The gels were dried and exposed to autoradiography film.

A 50:50 Protein A Sepharose slurry was prepared as follows: 0.1g of Protein A Sepharose was added to a microfuge tube, and then expanded by adding 1 mL of ddH_2O. The resin was washed twice with deionized water by centrifugation for 1 minute at 18,000 g, and then incubated with rocking for 15 minutes in deionized water. The resin was washed with the buffer used in experiments (1X Lysis buffer here), and then resuspended using a volume of the wash buffer that equals the volume of the resin itself, a 1:1 ratio. Pansorbin cells were washed with the buffer used in experiments (1X Lysis buffer here), and then resuspended using a volume of the wash buffer that equals the volume of the cells themselves, a 1:1 ratio.

6.5 GENERAL METHODS FOR WESTERN BLOT ANALYSIS

All Western blot procedures were performed in a similar manner using standard procedures (Sambrook and Russell, 2001). Proteins were separated on SDS-polyacrylamide gels along with Western Kaleidoscope broad pre-stained standard markers (Bio-Rad), and transferred to Immobilon-P PDVF membrane (Millipore) for immunoblotting. Immobilon-P PDVF membranes were pre-wet in methanol and soaked in 1X Transfer buffer [25 mM Tris, 0.25 M glycine, 15% methanol] according to the manufacturer's protocol. Proteins were transferred to the membrane using an Owl Scientific semi-dry transfer apparatus at constant amperage for 90 minutes following standard procedure (Sambrook and Russell, 2001). Membrane were blocked in 1X PBS, pH 7.4, with 5% non-fat dry milk overnight at 4°C or 1 hour at room temperature with rocking.

Detection of immune-complexes was completed essentially as described in the manufacturer's protocol for the ECL Western detection reagent (Amersham). Unless otherwise stated, blots were incubated in 1X TBST (Appendix A), 1% BSA, and primary antibody for 1 hour at RT with rocking, followed by three five minute rinses using 1X TBST. Blots were then incubated in 1X TBST, 1% BSA, and secondary antibody for 1 hour at RT with rocking, followed by three five minute rinses using 1X TBST. After washing, immunocomplexes were detected using the ECL Western blotting kit (Amersham). Blots were exposed to film for 1, 5, 15, and 30 minutes or subjected to imaging using a Fuji BAS-3000 imager with chemiluminesence options. Following imaging, if desired, blots were striped using 1X Western

strip buffer (Appendix A) for 1 hour at RT, then rinsed at least five times with 1X TBST, and blocked as described above. Membranes were then probed with another primary antibody.

6.6 DETECTION OF ENDOGENOUS TBX6

6.6.1 Preparation of embryonic lysates to detect endogenous mouse Tbx6

To generate an embryonic protein lysate enriched in Tbx6 protein, Swiss Webster embryos were dissected at e10.5 and the tailbud regions were isolated. Lysates were prepared from the tailbud region and from somitic regions (negative control) of approximately 90 e10.5 embryos. These samples were isolated and lysed in 180 μL 1X Lysis buffer containing protease inhibitors (Appendix A). The samples were lysed by freezing in liquid nitrogen, thawing on ice, and briefly sonicating. Lysates were centrifuged at 18,000 g for 10 minutes at 4°C. Supernatants were transferred to new tubes, and the lysate volumes representing approximately 5 somites and 5 tailbuds were run on a 10% SDS-polyacrylamide gel along with Western Kaleidoscope broad pre-stained standard markers (Bio-Rad) using standard procedures (Sambrook and Russell, 2001).

6.6.2 Western blot analysis of Tbx6 using purified Tbx6 antibodies

The above gels (Section 6.6.1) were blotted using either the N-terminal or Internal primary antibody at a 1:10,000 dilution (approximately 30 ng/mL, 20 mL final volume) as detailed in Section 6.5. The rabbit primary antibodies were detected with goat-anti-rabbit secondary antibody (1:5,000, Amersham ECL Western kit). Following imaging, these blots were striped, and then probed with rabbit anti-actin primary antibody (1:2000, AAN-01, Cytoskeleton), and immune complexes were detected with goat-anti-rabbit secondary antibody (1:5,000, Amersham ECL Western kit) using the ECL Western blot kit (Amersham).

6.7 IMMUNOLOCALIZATION OF TBX6 IN MOUSE EMBRYOS

Swiss Webster (SW) mice were mated and checked for the presence of a copulation plug. Embryos were dissected at e7.5 through e13.5 and immunocytochemistry was performed essentially according to standard protocols (Nagy, 2003). Briefly, embryos were fixed in methanol/DMSO (4:1) overnight at 4°C, transferred to methanol/DMSO/H_2O_2 (4:1:1) for 5 hours at RT to block endogenous peroxidase activity, and then stored at -20°C in 100% methanol. Embryos were rehydrated with rocking at RT in 50% methanol for 30 minutes, 1X PBT [1X PBS (Appendix A), 0.5 % Triton-X100, 0.2 % BSA] for 30 minutes, and two one-hour washes in 1X PBT. Embryos were incubated with either N-terminal or Internal affinity-purified antibodies at a 1:500 dilution (or approximately 1 μg/mL) overnight at 4°C with rocking. Embryos were washed five times total, once in 1X PBT, twice with 1X PBT for 30 minutes each wash, and twice with 1X PBT for 1 hour each wash. Embryos were then incubated with goat anti-rabbit:HRP-conjugated secondary antibody (Jackson ImmunoResearch) at a 1:500 dilution in 1X PBT at 4°C with rocking. Embryos were washed as previously described, and then incubated with 1X PBT/DAB/$NiCl_2$ [1X PBT plus 0.6 mg/mL DAB and 0.6 mg/mL $NiCl_2$] at RT for 30 minutes with shaking before adding hydrogen peroxide (0.0003%) until desired color (usually 10 minutes). Embryos were post fixed one hour in 4% paraformaldehyde in PBS, washed, and photographed following standard protocols (Nagy, 2003).

6.8 CHARACTERIZATION OF THE TBX6 BINDING SITE

6.8.1 Electrophoretic mobility shift assay of various T-box binding sites

The oligonucleotides along with their complements listed below were synthesized and gel purified for use in EMSAs. The T^{bind} sequence matched the *in vitro* selected T palindromic site and the T^{half} sequence represented the T half-site (Kispert and Herrmann, 1993). Oligonucleotides were end-labeled with γ-^{32}P-ATP using T4 polynucleotide kinase, annealed and purified using Micro Bio-Spin P-30 Tris purification columns (Bio-Rad, Hercules, CA), with the

final elution concentration of approximately 0.2 nM. To determine the efficiency of oligo annealing, single stranded oligos were labeled and purified along with annealed double-stranded probes; both sets of annealed oligos were separated on non-denaturing 18% polyacrylamide (37.5:1) – 1X TAE gels. The gels were dried unfixed, exposed to autoradiography film, and imaged using a Fuji BAS-2500 phosphoimager. The double-stranded version of the T^{bind} probe routinely resulted in approximately 80% of the oligos in the annealed, double stranded orientation, while the T^{half} probe routinely resulted in approximately 95% of the oligos in the annealed, double stranded orientation. The amount of double-stranded probe was standardized such that equal amounts of double-stranded probes were used. EMSAs were performed essentially as described in Kispert and Herrmann (1993) except as noted. Briefly, all EMSA binding reactions were prepared in a final reaction volume of 10 µl with 1X BBT buffer [25 mM HEPES, pH 7.4, 75 mM NaCl, 1 mM DTT, 0.25 mM EDTA, 0.1 % NP-40, 1 mM $MgCl_2$, 10 % glycerol, 10 mg/ml BSA] with 0.1 mg/ml Poly (dI-dC) (Amersham, Piscataway, NJ). 10 fmol of ^{32}P-labeled double-stranded probe was added to the programmed lysates in 1X BBT buffer. For antibody supershifts, 0.5 µg of purified N-terminal or Internal Tbx6 antibody was added to the reaction. For competition experiments, increasing concentrations of unlabeled probes from 10 fmol to 1 pmol were added to the reactions. Reactions were separated on non-denaturing 5% polyacrylamide (37.5:1) – 1X TAE gels. The gels were dried unfixed and imaged using a Fuji BAS-2500 phosphoimager.

EMSA probes

T^{bind}: 5'- AATTTCACACCTAGGTGTGACTAG -3'

T^{half}: 5' - ATCGAATTCAGGTGTGAAATTGGATCCACT - 3'

D1T6BS1: 5'- ACAGGACAGCAACACCTACAGTGA - 3'

D1T6BS2: 5'- TCCAAGAATCACACCTCGGGA -3'

D1T6BS3: 5'- GCAGCCCAGTGAGGACACCTGGATCCA -3'

D1T6BS4: 5'- GAGAGGGTGTGTACACCGGGAAAG -3'

6.8.2 Binding site selection to identify Tbx6-specific selected sequences

A PCR-based binding site selection procedure was performed essentially as described by Kispert and Herrmann (1993). The 76 base oligonucleotide (R76) and oligonucleotides used for PCR amplification of R76 were synthesized by IDT DNA Technology as listed below. The R76 oligonucleotide was rendered double-stranded using the Klenow fragment and the reverse primer, Primer F. This double-stranded oligonucleotide was used to identify Tbx6 binding sites *in vitro*. *In vitro* translated T7-Tbx6 protein was used as the protein source for the binding site selection procedure (Section 6.4.1).

The PCR-based binding site selection scheme was performed as described below. Protein lysates were mixed together with R76 to allow Tbx6 to bind Tbx6-specific sequences. This bound DNA was then purified using immunoprecipitation to enrich for Tbx6-specific sequences; bound DNA was amplified and used for subsequent binding reactions with Tbx6 protein. Six rounds of selection were performed according to the procedure described to identify a T binding site (Kispert and Herrmann, 1993). This experiment was performed three separate times with three different antibodies against the T7-tagged Tbx6 protein: (1) a monoclonal anti-T7 antibody directed against the T7-epitope tag (Novagen, EMD), (2) N-terminal Tbx6 antibody, and (3) Internal Tbx6 antibody. Reactions were prepared by mixing 2 μL of T7-Tbx6 programmed lysate with 1 μL (approximately 1 ng) R76, 2 μg Poly d(I-C), and Tbx6-specific antibody (see below) in 250 μL 1X BBE buffer [20 mM HEPES, pH 7.4, 100 mM KCl, 1 mM DTT, 0.25 mM EDTA, 0.1 % NP-40, 20 % glycerol, 1 μg/ml BSA, 1 μg/ml Leupeptin, 0.7 μg/ml Pepstatin A]. When using the monoclonal anti-T7 antibody, 1 μg of the antibody was used along with 1 μg of rabbit anti-mouse antibody to bridge the monoclonal mouse antibody for immunoprecipitation. When using our Tbx6 antibodies, 1 μL of the terminal bleed from anti-T6N143 (N-terminal) or anti-T6I145 (Internal) was used. After incubating the sample overnight at 4°C with rocking, 30 μL of a 50:50 Protein A Sepharose (17-0780-01, Amersham) slurry resuspended in 1X BBE was added to the reaction, and the reaction was further incubated for 1 hour at 4°C with shaking. The protein/DNA complexes were immunoprecipitated and then washed four times with 500 μL 1X BBE for each wash. After the last wash, the DNA was eluted by adding 200 μL 1X DNA elution buffer [50 mM Tris, pH 8, 5 mM EDTA, pH 8, 0.5 % SDS,

and 100 mM sodium acetate] for 15 minutes at 37°C. The eluted DNA was phenol extracted and ethanol precipitated in the presence of 5 μg of glycogen carrier and approximately 6 mM $MgCl_2$. The DNA was resuspended in 20 μL of 50 mM Tris, pH 8, and stored at -20°C until further use. PCR of selected Tbx6-specific sequences was performed as follows: The PCR reaction contained: 2 μL of selected sequence DNA, 1X PCR buffer [20 mM Tris, pH 8.4, and 50 mM KCl] (Invitrogen), 2 mM $MgCl_2$, 0.5 mM dNTPs, 20 pmol of each primer from the set (Primer F and Primer R below), and 2 U of Taq DNA polymerase (Invitrogen) in a 20 μL volume. The cycling conditions were as follows: 15 cycles of 94°C for 45 seconds, 62°C for 45 seconds, and 72°C for 30 seconds; and a final hold at 4°C. Protein and primers were removed from the PCR products using MinElute PCR purification columns (Qiagen), and approximately 20% of the resulting sequences was subjected to the next round of selection.

This procedure enriched for Tbx6-specific sequences, however the number of times that the pool of selected DNA sequences must be subjected to additional rounds of selection varied. Therefore, the efficiency of the selection scheme was tested using EMSAs to ensure that the pool of selected sequences truly represented sequences that Tbx6 bound *in vitro*. An aliquot of double-stranded R76 and selected DNA sequences at steps 2, 4, and 6 were radiolabelled with γ-^{32}P-ATP using T4 polynucleotide kinase and purified using Micro Bio-Spin P-30 Tris purification columns (Bio Rad) following previously described methods (Section 6.8.1). EMSA reactions were prepared as described in Section 6.8.1, with the radioactive probes being incubated with T7-tagged Tbx6 or control lysate. The reactions were separated on non-denaturing 4% polyacrylamide (37.5:1) – 1X TAE gels following standard protocols (Sambrook and Russell, 2001). The gels were dried unfixed and exposed to autoradiographic film. Following the fourth (N-terminal and Internal antibodies) or sixth (T7 antibody) round, selected sequences were PCR amplified as described above, digested with BamHI and EcoRI, and cloned into pBluescript II SK-. Approximately equal numbers of inserts from each antibody selection procedure were sequenced (Genewiz).

R76 oligonucleotide and PCR primers

5’-CAGGTCAGTTCAGCGGATCCTGTCG$(N)_{26}$GAGGCGAATTCAGTGCAACTGCAGC-3’

Primer F: 5'- GCT GCA GTT GCA CTG AAT TCG CCT C -3'

Primer R: 5'- CAG GTC AGT TCA GCG GAT CCT GTC G -3'

6.9 BASIC MAINTAINANCE / TRANSFECTION OF TISSUE CULTURE CELL LINES

6.9.1 Basic Cell Maintenance for COS-7 and 293T/17 cell lines

The following cell lines were obtained from the American Type Culture Collection and used in these experiments: COS-7 (ATCC CRL-1651) and 293T/17 (ATCC CRL-11268). Cells were maintained in media at 37°C and 5% CO_2. Media was changed every two days, and cells were passaged at 1:5 approximately every three to four days, when they reached approximately 95-100% confluency. Cells were maintained in the following 1X growth media [high glucose DMEM (12-604f, BioWhittaker) supplemented with 10% fetal bovine serum (35-010-CV, CellGro), 2 mM L-glutamine (25030-081, Gibco/Invitrogen), and 1X penicillin/streptomycin (15140-122, Gibco/Invitrogen) antibiotics]. All cell lines were frozen in 1X growth media plus 5% DMSO.

6.9.2 Transfections of tissue culture cell lines

Transfections were performed using Lipofectamine 2000 (Invitrogen) essentially as described by the manufacturer. For multiple vector transfections, the DNA concentration was kept constant. The 293T/17 cell line was grown overnight on tissue culture plates treated with 25 mg/mL polyethyleneimine (PEI) in 1X PBS, pH 7.4, a positively charged polymer that improves cell adhesion and transfection efficiency for various cell lines (Vancha et al., 2004). The COS-7 cell line was grown on tissue culture plates without any treatment. When growing all cell lines on coverslips for immunofluorescence imaging, coverslips were treated with 0.1% gelatin following standard procedure before transfections (Spector and Goldman, 2003).

6.10 CLONING VARIOUS EXPRESSION PLASMIDS FOR TISSUE CULTURE LUCIFERASE ASSAYS

To generate a Tbx6 expression vector, the T7-tagged Tbx6 fusion protein (Section 6.3.1) was used. To generate a tagged Brachyury protein, the full length *Brachyury* cDNA was PCR amplified using an annealed temperature of 56°C, a two minute extension time, and the primers (shown below) that introduced a BamHI site in the primer before the start codon of *Brachyury* and a NotI site replacing the stop codon. The size of the PCR product was verified on agarose gels, and the product of approximately 1300 bp was purified using a MinElute PCR purification columns (Qiagen). Purified PCR was digested with BamHI and NotI, cloned into pBluescript II SK- with Blue/White selection using BamHI and NotI sites within the vector, and subsequently cloned in-frame with the T7 tag of pET21a using the BamHI and NotI sites within the pET21a vector. The cloned expression plasmid was sequenced (Genewiz), and then used for further cloning. For both Tbx6 and Brachyury, the T7-tagged full-length fusion proteins were digested with XbaI and NotI, excising the coding region for the T7-tag and entire protein coding region, and subcloned into pBluescript II SK- using XbaI and NotI as an intermediate step in cloning. These vectors were then digested with Asp718/NotI (Tbx6) or HindIII/NotI (Brachyury) and cloned into pcDNA3.1/V5-His to create expression vectors for tissue culture. Both expression vectors expressed a full-length protein with a T7 tag; however the Brachyury construct also contained a V5 epitope and a His tag on the C-terminus. To generate a control vector for use in transfections, the MCS region from pBluescript II SK- (Asp718/NotI) was cloned into the Asp718 and NotI sites within the pcDNA3.1/V5-His vector, creating the control vector for tissue culture, pcDNA3.1/V5-His-MCS or MCS.

Primers

Tclone for: 5'- aatctggatccATGAGCTCGCCGGGCAC -3'

Tclone rev: 5'- aatgtgcggccgccCATAGATGGGGGTGACACAGG -3'

6.11 IMMUNOFLUORESCENCE OF TISSUE CULTURE CELL LINES

Coverslips in 6 well plates were treated with 0.1% gelatin. Cells were grown as previously described for 24 hours prior to transfection. Cells were transfected using the Lipofectamine 2000 reagent (Invitrogen), with 4 μg per well T7-tagged Tbx6 in pcDNA3.1, following manufacturer protocol and grown for 24 hours. Immunofluorescence of fixed cells was performed essentially as described in Chapter 64: Immunofluorescence Localization of Nuclear Proteins (Spector and Goldman, 2003). Cells were washed twice with warm PBS (37°C), and then fixed in warm 2% paraformaldehyde in 1X PBS, pH 7.4, for 15 minutes at RT. Fixed cells were washed in 1X PBS, pH 7.4 (Appendix A) twice for 5 minutes each, and then permeabilized using 0.2% Triton X-100 plus 1% normal goat serum (NGS) in 1X PBS for 5 minutes on ice. Cells were washed in 1X PBS plus 1% NGS three times for 10 minutes each, then cells were incubated with purified anti-Tbx6 (N-terminal or Internal antibody, approximately 25 ng/mL) or pre-bleed antisera (1 μL) in 1X PBS plus 1% NGS for 1 hour at RT in a humidified chamber. Cells were washed in 1X PBS plus 1% NGS three times for 10 minutes each. All cells were incubated with Alexa goat anti-rabbit (1:500, 488nm, Molecular Probes), TRITC-Phalloidin (1:4000, Sigma), and TO-PRO-3 (1:1500, Molecular Probes) in 1X PBS plus 1% NGS for 1 hour at RT in a humidified chamber. Cells were washed in 1X PBS four times for 10 minutes each. The coverslips were mounted onto slides using VectaShield mounting media (H-1000, Vector Labs), and optical sections were visualized on a BioRad scanning laser confocal microscope. NIH Object Image was used to select section stacks, and Adobe Photoshop was used to merge the different channel images.

For immunofluorescence of the T fusion protein, the procedure was essentially as described above, except the block was 1% Normal Donkey Serum, the primary antibody was a goat polyclonal anti-Brachyury N1-19 (sc-17743, 1:100, Santa Cruz Biotech), and changes to the secondary antibodies were a donkey anti-goat Rhodamine secondary antibody (1:300, Santa Cruz Biotech) and FITC-Phalloidin (1:4000, Sigma). When using Adobe Photoshop to merge the different channel images, the anti-Brachyury channel was moved to the green channel and the Phalloidin was moved to the red channel.

6.12 WESTERN BLOT ANALYSIS OF BRACHYURY USING ANTI-BRACHYURY

Western blot procedures were performed as described in Section 6.5, except as noted below. Cell culture lysates enriched in Brachyury were blotted using the anti-Brachyury primary antibody (sc-17743, 1:200 dilution, Santa Cruz Biotech) in 1X TBST (Appendix A) plus 3% BSA for 1 hour at RT. The goat primary antibody was detected with donkey-anti-goat secondary antibody (sc-2056, 1:5000 dilution, Santa Cruz Biotech) in 1X TBST plus 1% BSA for 1 hour at RT. Following imaging, anti-Brachyury blots were striped and then probed with rabbit anti-actin primary antibody (1:2000, AAN-01, Cytoskeleton). The rabbit primary antibody was detected with goat-anti-rabbit secondary antibody (1:5,000, Amersham ECL Western kit).

6.13 CLONING OF LUCIFERASE EXPRESSION PLASMIDS FOR TISSUE CULTURE ASSAYS

6.13.1 Luciferase Source Vectors

There were two different firefly luciferase source vectors used in these assays. The first used were the pGL3-based vectors (Promega), consisting of the pGL3-Basic [pGL3B] vector, a promoterless vector that expresses the *luc* cDNA (luciferase), and the pGL3-Promoter [pGL3P] vector, containing the pGL3 vector under the control of a SV40 minimal promoter. The second firefly luciferase vector was based on the pGL4.10 (Promega) promoterless plasmid. The constitutively active SV40 early promoter and enhancer [pGL4.SV40] was made to provide a luciferase positive control, ensuring that the transfection occurred properly and that the luciferase transcriptional assay resulted in data within a range readable by a luminometer at the concentrations found within the cell lysates. I modified the pGL4.10 plasmid to remove one potential Tbx6 binding site, at approximately 400 bp in the *luc2* luciferase cDNA (pGL4M). Using the pGL4M plasmids, I developed plasmids that express luciferase under the control of the β-globin [pGL4M(βg)] and hsp70 minimal promoters [pGL4M(hsp)]. The sequences of all these plasmids were confirmed by DNA sequencing (Genewiz). The pSV-β-galactosidase control

vector (Promega) contained the constitutively active SV40 early promoter and enhancer to provide a positive control vector for monitoring transfection efficiencies of mammalian cells; the vector acted as an internal control for transient expression assays.

6.13.2 Creation of T/Tbx6 binding site luciferase reporter vectors

To create the T/Tbx6 luciferase reporter vectors, I modified the luciferase expression plasmids (described in Section 6.13.1) to place luciferase under the control of the SV-40 [pGL3P], β-globin [pGL4M(βg)], and hsp70 minimal promoters [pGL4M(hsp)]. The T palindromic sequence (Tbind), T half-site sequence (Thalf), and a mutated form of the T half-site (Tmut) sequence (listed below) were cloned at the KpnI and SacI restriction sites within the multiple cloning region of these vectors. The T half site sequence is underlined in all oligonucleotide sequences, and the mutated form of the T half site changes the 'TG' to 'GT' as bolded below. Annealed oligonucleotides ligated into digested vectors in the direction listed below approximately 30 bp upstream of the minimal promoter. The sequences of all plasmids were confirmed by sequencing (Genewiz).

Primers

Tbind forward: 5'- AATTCAATTTCACACCTAGGTGTGAAATTG -3'
Tbind reverse: 5'- GATCCAATTTCACACCTAGGTGTGAAATTG -3'

Thalf forward: 5'- AATTCAGGTGTGAAATTG -3'
Thalf reverse: 5'- GATCCAATTTCACACCTG -3'

Tmut forward: 5'- AATTCAGG**GT**TGAAATTG -3'
Tmut reverse: 5'- GATCCAATTTCA**AC**CCTG -3'

6.13.3 Creation of *Dll1* enhancer luciferase reporter vectors

A portion of the previously published *Dll1* 'msd' enhancer (Beckers et al., 2000a) was PCR amplified from the genomic DNA from a wild type FVB/N mouse with the PCR primers listed below, an annealing temperature of 50°C, and an extension time of 45 seconds. The PCR primers contained the Asp718 and XhoI restriction sites for cloning purposes. The size of the PCR product was verified on an agarose gel, and the 187 bp region of the *Dll1* enhancer was purified using a MinElute PCR purification columns (Qiagen). Purified PCR products were digested with Asp718 and XhoI and then cloned into the pBluescript II SK- vector. The sequence of the plasmid above was confirmed by sequencing (Genewiz).

Four putative Tbx6 binding sites found in the *Dll1* 'msd' enhancer were mutated by site directed mutagenesis using the QuikChange kit (Stratagene) following manufacturer's directions. Oligonucleotides used to mutate these putative Tbx6 binding sites and their complements were synthesized (IDT DNA Technology); the underlined region is the putative Tbx6 binding site except that it contains a 'TG' to 'AA' mutation (in bold). The resulting plasmids were sequenced to confirm mutation of the original putative Tbx6 binding sites (Genewiz).

Since the cloned *Dll1* 'msd' enhancer and the 'msd' enhancer containing single site directed mutants were cloned into pBluescript II SK-, the inserts were PCR amplified using the standard M13 forward and M13 reverse primers for subsequent cloning into luciferase reporter constructs. The size of the PCR products were verified on an agarose gel, purified using MinElute PCR purification columns (Qiagen), and digested with Asp718 and Xho1 for cloning into the pGL4M(βg) and pGL3P vectors, the luciferase expression vectors described in Section 6.13.1. Plasmids were sequenced to verify proper cloning (Genewiz).

Primers

Cloning of *Dll1* enhancer region

D1asp forward: 5'- aatggtacc CCTAAATGATCAAGC-3'

D1xho reverse: 5'- attactcgag CCTCTTTACTCCACC -3'

Site directed mutagenesis of *Dll1* enhancer region

D1BS1 SDM for: 5'- GTCCTCACTGTAGG**AA**TTGCTGTCCTGTTGG -3'

D1BS2 SDM for: 5'- ATTAGAATCCCGAGG**AA**TGATTCTTGGAGCTAG -3'

D1BS3 SDM for: 5'- GTGTACCTGACAGG**AA**TCCTCACTGTAGGTGTT -3'

D1BS4 SDM for: 5'- CACAGCCCCTAGGG**AA**TACCTGACAGGTGTCCT -3'

6.14 THE EXPERIMENTAL PROCEDURES FOR CONDUCTING VARIOUS TBX6 LUCIFERASE ASSAYS

Cells were grown on 24 well tissue culture plates as described for 24 hours prior to transfection (Section 6.11.1). Cells were transfected using Lipofectamine 2000 reagent (Invitrogen) with 0.8 μg per well total DNA following the manufacturer's protocol; the specific concentrations of the various DNA constructs are described below. 24 hours post-transfection, cells were washed twice with warm 1X PBS, pH 7.4 (37°C), incubated in 300 μL of 1X Reporter lysis buffer (Promega) for 15 minutes, then scraped off the plates using pipet tips and placed into microfuge tubes. Cells lysates were centrifuged at 18,000 g for 30 seconds at RT to remove insoluble debris and nuclear material, and the supernatants were used immediately or stored at -80°C until use.

To detect the expression of the β-galactosidase transfection control, the β-galactosidase Assay System (E2000, Promega) was used, following the protocol provided by the manufacturer. Briefly, 50 μL of cell lysate prepared above was mixed with 50 μL of 2X assay buffer (Promega). Samples were incubated approximately 3 hours at 37°C, or until a faint yellow color appeared. 200 μL of 1 M sodium acetate was added to quench the reaction, and reactions were read using 96 well plates for processing through a Bio-Rad 3550 microplate reader. This assay was performed as triplicate reads of each sample to calculate confidence intervals, and included a β-galactosidase enzyme positive control diluted from 0.1 to 0.5 milliunits.

To detect the expression of the luciferase reporter, the Luciferase Assay System (E4030, Promega) was used, following the protocol provided by the manufacturer. Briefly, 20 μL of cell lysate was added to a luminometer cuvette for reading in a BD Biosciences Monolight single

tube 2010 luminometer (BD Biosciences). The luminometer injected 100 μL of 1X luciferase assay buffer (Promega) for mixing the enzyme-substrate complex, and was programmed to pause 2 seconds before recording luminescence over 10 seconds. This assay was performed as a triplicate read of each sample to calculate the confidence interval for a mean value representing luminescence for each sample. The pGL4.SV40 construct was included in transfections for use as a luciferase positive control.

6.14.1 Expression studies of Tbx6 on luciferase activity

To study the effect of Tbx6 on luciferase gene expression, I used the nine reporter constructs described in Section 6.13.3 and summarized in Table 3-2; 500 ng of the 800 ng total DNA consisted of the luciferase reporter DNA. The β-galactosidase control vector was used at 150 ng of the 800 ng total DNA for normalizing transfection efficiency as previously described (Section 6.13.2). The Tbx6 fusion protein was added at a constant level by using 150 ng of the expression construct. Results from the luciferase assays were used to calculate the average luciferase units as determined by the following calculation: (Average raw, zeroed luciferase activity) divided by Average calculated β-galactosidase value (milliunits). This average luciferase units was used to compare the level of luciferase activity across the processed samples. The relative fold change in the luciferase activity was presented in reference to the identical sample, except transfected with empty protein vector (pcDNA3.1/V5-His-MCS) in place of the Tbx6 expression construct. This value, the Relative Luciferase Units (RLU), was determined as follows: Average Luciferase Units (sample containing Tbx6) divided by Average Luciferase Units (sample containing control vector).

6.14.2 Co-expression studies of Tbx6 and T on luciferase activity

To study the effect of Tbx6 and T co-expression on target gene expression, I used the pGL3P.Tbind plasmid at 450 ng of the 800 ng total DNA (Section 6.13.2). The β-galactosidase control vector was used at 125 ng of the 800 ng total DNA for normalizing transfection efficiency as previously described (Section 6.13.2). Either T or Tbx6 was added at a constant level by using 100 ng of the fusion protein expression construct. The competing T-box fusion protein expression construct was varied by addition of increasing levels of construct DNA (from 25 ng to 125 ng of 800 ng total DNA) while keeping the total DNA used for transfection constant with empty vector (pcDNA3.1/V5-His-MCS) to measure the affect of both proteins on the reporter. In addition to the controls described in Section 6.13.2, I included the pGL3P.Tbind without T or Tbx6 to establish the baseline for these luciferase expressions. Results were expressed as in Section 6.14.1 in terms of a ratio of the normalized luciferase units of the experimental T/Tbx6 samples over the pGL3P.Tbind without T or Tbx6 (an empty vector sequence detecting only background activity).

6.15 RT-PCR ANAYLSIS OF TBX6 TRANSCRIPTS FROM *TBX6* MUTANT EMBRYOS

6.15.1 mRNA isolation from *Tbx6*$^{rv/rv}$ and *Tbx6*^{Tg46R} embryos

Embryos from *Tbx6*$^{rv/rv}$x *Tbx6*$^{rv/+}$ or *Tbx6*$^{+/-}$*Tbx6*$^{TG46/+}$ x *Tbx6*$^{+/-}$ crosses were dissected at e10.5; yolk sacs were isolated to genotype the embryos and embryos isolated were stored at -70°C until use for RNA isolation. The Micro FastTrack mRNA isolation kit (Invitrogen) was used to purify total mRNA from one embryo of each genotype according to the manufacturer protocol. Isolated mRNA was resuspended in 8 μL of 50 mM Tris, pH 8, and 1 μL of this was treated with RQ1 DNase (M6101, Promega, Madison, WI) to remove contaminating DNA from the mRNA sample.

6.15.2 RT-PCR of *Tbx6* from *Tbx6*$^{rv/rv}$ and *Tbx6*^{Tg46R} embryos

Tbx6 cDNA was synthesized using the Thermoscript RT enzyme (12236-014, Invitrogen), essentially following the manufacturer's protocol. DNase-treated, purified mRNA, the T6RT reverse primer (20 pmol), and 0.4 mM dNTPs were heated to 65°C for 5 minutes, and then cooled to RT to anneal the synthesis primer with the mRNA. After cooling, the reaction buffer, DTT and water were added to a final volume of 24 μL. Half of the reaction (12 μL) was stored at -70°C for use as a negative control to ensure that the DNase treatment was successful. 1 μL of Thermoscript RT enzyme was added to the other half, heated to 55°C for 45 minutes for cDNA synthesis, followed by heating to 85°C for 5 minutes to denature the enzyme.

These cDNAs were then used for PCR using PCR primers (listed below) designed to amplify the *Tbx6* cDNA. Amplification of the *Tbx6* cDNA was performed as follows: The PCR reaction contained: 2 μL of sources DNA, 1X PCR buffer [20 mM Tris, pH 8.4, and 50 mM KCl] (Invitrogen), 3.5 mM $MgCl_2$, 0.4 mM dNTPs, 10 pmol of each primer, 5% DMSO, and 1.25 U of Taq DNA polymerase (GeneChoice) in a 50 μL volume. The cycling conditions were as follows: 95°C for 5 minutes; 46 cycles of 94°C for 45 seconds, 56°C for 30 seconds, and 72°C for 2 minutes; and a final extension at 72°C for 8 minutes. A total of 8 samples were run including: no DNA (negative control), 15 pg previously cloned *Tbx6* cDNA (positive control), and for each embryo, 2 μL mRNA (before DNase treatment, test for initial DNA contamination), 2 μL cDNA reaction (with no Thermoscript RT enzyme added), and 2 μL cDNA reaction (after Thermoscript RT enzyme added).

The PCR fragments were resolved on 0.8% agarose gels to confirm amplification of the cDNA, gel purified, cloned into the pCR4 vector (TOPO-TA kit, Invitrogen), and transformed into bacteria. Plasmids from five independent colonies from the *rv/rv* sample were isolated using the Qiagen Spin mini prep kit (Qiagen), and sequenced using the primers listed below.

PCR Primers

T6RT forward primer: 5'- GCCAGAAGAAACTACAAC -3'

T6RT reverse primer: 5'- GATCCTGAGAACAAATTG -3'

Sequencing Primers

T6 seq for: 5'- GAAACTGTAAGAGGGAGCG -3'

T6 seq rev: 5'- AAAGCGGCAGGGTGTAGAAGGTAG -3'

T7 for: 5'- TAATACGACTCACTATAGGG -3'

M13 rev: 5'- CAGGAAACAGCTATGAC -3'

6.16 PCR OF GENOMIC DNA FOR CLONING OF THE *RIB-VERTEBRAE* MUTATION

Genomic DNA from *Tbx6*$^{rv/rv}$ and C57B16/J wild type mice was used in PCR using the primers flanking the affected region and designed from sequences obtained at the Ensembl Mouse Genome Server. The PCR reaction contained: 1 μg (3 μL) of genomic DNA isolated from a *Tbx6*$^{rv/rv}$ or wild type animal, 1X PCR buffer [20 mM Tris, pH 8.4, and 50 mM KCl] (Invitrogen), 4.25 mM $MgCl_2$, 1 mM dNTPs, 10 pmol of each primer, 5% DMSO, and 1.25 U of Taq DNA polymerase (Invitrogen) in a 50 μL volume; a PCR reaction without DNA was used as a negative control and wild type DNA (C57B16/J) was used as a positive control. The cycling conditions were as follows: 95°C for 5 minutes; 40 cycles of 94°C for 1 minute, 59°C for 30 seconds, and 72°C for 3.25 minutes; and a final extension at 72°C for 10 minutes. The PCR fragments were resolved on 0.8% agarose gels to confirm amplification, gel purified, cloned into the pCR4 vector (TOPO-TA kit, Invitrogen), and transformed into bacteria. Plasmids from three clones for the *Tbx6*$^{rv/rv}$ sample were isolated using the Qiagen Spin mini prep kit (Qiagen), and sequenced using the primers listed below.

PCR primers

T6reg forward primer: 5'- CATTCCCAAAACCCCATTGC -3'

T6reg reverse primer: 5'- TGCCCCTTCACTCTCTCCATC -3'

Sequencing primers

T6regfor: 5'- CATTCCCAAAACCCCATTGC -3'

T6regrev: 5'- TGCCCCTTCACTCTCTCCATC -3'

T6seq2 rev: 5'- TAGCCTTGCGAAGTGAACG -3'

T6seq3 for: 5'- TTAGTCCTGGCTTTCGTTC -3'

T7 for: 5'- TAATACGACTCACTATAGGG -3'

M13 rev: 5'- CAGGAAACAGCTATGAC -3'

6.17 CONDITIONS FOR PCR GENOTYPING OF *TBX6, DLL1* AND *WNT3A* MICE AND EMBRYOS

Genomic DNA was prepared from tail or embryonic tissues and used for PCR. The primers, designed from sequences obtained at Ensembl Mouse Genome Server, are listed below. The PCR reaction contained approximately 1 μg of genomic DNA, 1X PCR buffer [20 mM Tris, pH 8.4, and 50 mM KCl] (Invitrogen), 2 mM $MgCl_2$, 0.2 mM dNTPs, 10 pmol of each primer, 5% DMSO, and 1.25 U of *Taq* DNA polymerase (GeneChoice) in a 50 μL volume. The cycling conditions for *Dll1* and *Tbx6* genotyping were as follows: 95°C for 10 minutes; 42 cycles of 94°C for 45 seconds, a varying annealing temperature for 55 seconds, and 72°C for 55 seconds; and a final extension at 72°C for 4 minutes. The annealing temperature started at 67°C for touchdown PCR for 18 cycles, decreased 0.5°C per cycle, and then remained at 58°C for 24 additional cycles. The cycling conditions for genotyping *Wnt3a* (and *Tbx6* that also works under these conditions) were as follows: 95°C for 10 minutes; 43 cycles of 94°C for 45 seconds, a varying annealing temperature for 45 seconds, and 72°C for 55 seconds; and a final extension at 72°C for 4 minutes. The annealing temperature started at 67°C for touchdown PCR for 18 cycles, decreased 0.5°C per cycle, and then remained at 59.5°C for 25 additional cycles. All PCR reactions were resolved on a 1.5% agarose gel and visualized with ethidium bromide staining.

Primers:

Tbx6 genotyping

T6KO forward: 5'- CAAACTGCGTCCCTGTCTTAGC -3'

T6KO reverse: 5'- CCTTACCCAGAGCCAATCCAAC -3'

T6KO mgeno reverse: 5'- TGCTCCTGCCGAGAAAGTATCC -3'

wild type – 450 bp band, *Tbx6* mutant – 620 bp

Dll1 genotyping

Dll1/38 forward: 5'- ATCCCTGGGTCTTTGAAGAAG-3' (Hrabe de Angelis et al., 1997)

Dll1/39 reverse: 5'- ATACGCGAAAGAAGGTCCTG-3' (Hrabe de Angelis et al., 1997)

D1 mGeno3 reverse: 5'- CGCTGATTTGTGTAGTCGGTTTATG-3'

wild type – 550 bp band, *Dll1* mutant – 750 bp

Wnt3a genotyping

W3a geno2 forward: 5'- CCAACTCCGTGTAAGACCTGAAAC -3'

W3a geno4 reverse: 5'- GGTGAAAAAGCAGCCCTTGC -3'

W3a mgeno3 forward: 5'- GGTGGATGTGGAATGTGTGCG -3'

wild type – 732 bp band, *Wnt3a* mutant – 500 bp

6.18 EMSA OF THE RBP-Jκ RECOGNITION SITE FOUND IN THE TBX6 PRESOMITIC MESODERM ENHANCER FOR BINDING BY RBP-Jκ

Oligonucleotides for the Su[H] binding site ($Su[H]^{bind}$) and a mutated site ($Su[H]^{mut}$), along with their complements, were synthesized and gel purified (Integrated DNA Tech). The $Su[H]^{bind}$ sequence (see below) matched a 36 base pair sequence located at -323 from the *Tbx6* start of transcription with the underlined region matching the conserved Su[H] binding site (Bailey and Posakony, 1995; Brou et al., 1994; Singson et al., 1994). The $Su[H]^{mut}$ sequence contained a substitution of guanine for cytosine at position 18. This G to C mutation had previously been shown to inhibit binding of Su[H] *in vitro* (Bailey and Posakony, 1995).

The oligonucleotides were mixed with their complementary strands (10 pmol each) and were end-labeled with γ–^{32}P-ATP using T4 polynucleotide kinase. The oligonucleotides were then heated to 90°C and slowly cooled to 37°C to create annealed 36 bp probes. Probes were purified using Micro Bio-Spin P-30 Tris purification columns (Bio-Rad, Hercules, CA), with final elution concentrations of 0.1 pmol/μl of annealed probe for Su[H]^{bind} and 0.15 pmol/μl of annealed probe for Su[H]^{mut}.

To determine the efficiency of oligonucleotide annealing, single stranded oligonucleotides (20 pmol) were labeled and purified along with the labeled, annealed double-stranded probes. Both the double- and single-stranded labeled probes were separated on non-denaturing 18% polyacrylamide gels. The gels were dried unfixed and exposed to autoradiographic film. After factoring in loss from purification and annealing, it was determined that the recovery of annealed probe was approximately 55% for the Su[H]^{bind} and 72% for the Su[H]^{mut} annealing reactions, with the remaining DNA present as single-stranded oligos.

Radiolabelled probes were used for EMSAs according to the previously published protocols (Brou et al., 1994; Singson et al., 1994). Briefly, all EMSAs were performed by initially incubating 1X Su[H] binding buffer [10 mM Tris, pH 7.5, 50 mM NaCl, 1 mM DTT, 1 mM EDTA, pH 8, 10 % glycerol], 3 mg BSA, 1 μg Poly d(A-T) (Sigma-Alrich), and 1 μg of purified GST-Su[H] or GST (protein negative control) at RT for 10 minutes. Competitor DNA (if used) and 50 fmol of ^{32}P-labeled double-stranded probe were then added and samples incubated an additional 20 minutes in a final volume of 10 μl of 1X Su[H] binding buffer. For competition experiments, increasing concentrations of unlabeled probes from 50 fmol to 2 pmol were added to the reactions. The entire reaction volume was separated on non-denaturing 4% polyacrylamide gels following standard protocol (Sambrook and Russell, 2001). The gels were dried unfixed and exposed to autoradiographic film.

Primers

Su[H]^{bind}: 5' - GTCTGCCCCC<u>ACCGTGG</u>**G**<u>AACGG</u>AGACCAGCTGGCC - 3'

Su[H]^{mut}: 5' - GTCTGCCCCC<u>ACCGTGG</u>**C**<u>AACGG</u>AGACCAGCTGGCC - 3

6.19 SITE-DIRECTED MUTAGENESIS OF THE RBP-Jκ BINDING SITE IN THE *TBX6* ENHANCER :: *LACZ* REPORTERS

Site-directed mutagenesis was used to mutate the RBP-Jκ site found within the PSM enhancer. Oligonucleotides representing the mutated RBP-Jκ (henceforth named Su[H]^{mut}) and its complement were synthesized and gel purified (Integrated DNA Tech). The Su[H]^{mut} oligonucleotide contained the RBP-Jκ binding site, except that it contained a guanine to cytosine substitution at position 8 (bold). The Su[H]^{mut} and its complement were used for site-directed mutagenesis of the RBP-Jκ binding site in the *pTbx6*$^{Tg122/lacZ}$ construct using the QuikChange kit (Stratagene) following the manufacturer's directions. The *pTbx6*$^{Tg122/lacZ}$ construct consisted of the *Tbx6* enhancer region from XbaI to Acc65I (position 2474 to 174 relative to the start of *Tbx6* transcription) driving *lacZ.* The plasmid was sequenced to verify the single base change using the sequencing primer given below.

The *pTbx6*$^{Tg122/lacZ\text{-}Su[H]}$ construct created above was then used for subsequent cloning to create p1229-AB3.8/lacZ-Su[H], consisting of the entire 2.3kb *Tbx6* PSM enhancer region, the entire coding region of *Tbx6*, and approximately 2.4 kb of downstream sequence, but maintaining the Su[H] site directed mutation. All cloning steps were completed by gel purifying the desired DNA band(s) prior to ligation. The *pTbx6*$^{Tg122/lacZ\text{-}Su[H]}$ construct (above) was digested with XbaI and Acc65I, and the 2.3kb band was subcloned into pLIT28; the pAB3.8 8.8 kb *Tbx6* genomic clone was digested with Acc65I and NcoI (~1.8kb). These fragments were ligated with the XbaI/NcoI sites into pLIT28, creating pLIT28-AB3.8X/NcoISu[H]SDM. This construct was then digested with XbaI and NcoI (~4.1 kb) and ligated to a 4.7 kb NcoI/XbaI fragment of the *Tbx6* genomic clone, and cloned into pBluescript II creating pBSII-AB3.8Su[H]SDM. Lastly, the pBSII-AB3.8Su[H]SDM was digested with XbaI, and the 8.8kb band representing the Tbx6 genomic clone containing the Su[H] mutated binding site was cloned into the p1229 vector at the XbaI site (p1229-AB3.8/lacZ-Su[H]). The plasmid was sequenced verifying the base change (primer below), and this construct is awaiting microinjection.

Primers

Su[H]^{mut} oligonucleotide: 5' - GTCTGCCCCC<u>ACCGTGG</u>**C**<u>AACGG</u>AGACCAGCTGGCC - 3'

T6SuH sequencing primer for: 5'- GATACCATAGATGTGTAACTG -3'

APPENDIX A

BUFFER RECIPES

1X Coupling buffer (pH 8.3)

0.1 M $NaHCO_3$

0.5 M NaCl

1X GST binding buffer (pH 7.3)

4.3 mM Na_2HPO_4O

1.47 mM KH_2PO_4O

137 mM NaCl

2.7 mM KCl

(may substitute 1X PBS, pH 7.4 for 1X GST binding buffer (pH 7.3))

1X GST elution buffer (pH 8)

50 mM Tris-HCl, pH 8

100 mM reduced glutathione

High pH Elute buffer (pH 11.5)

20 mM Tris-HCl, pH 11.5

0.5 M NaCl

1 mM EDTA, pH 8

1X Lysis buffer

40 mM Tris-HCl, pH 8

0.1 M NaCl

5 mM EDTA, pH 8

10 mM EGTA, pH 8

0.5 % NP-40

2 μg/mL Leupeptin

1 μg/mL Pepstatin A

Low pH Elute buffer (pH 2.5)

0.1 M glycine, pH 2.5

20 mM NaCl

1X PBS, pH 7.4

10 mM Na_2HPO_4O

2 mM KH_2PO_4O

137 mM NaCl

2.7 mM KCl

1X TBST, pH 7.5

20 mM Tris-HCl, pH 7.5

150 mM NaCl

0.05 % Tween 20

1X Western Strip buffer

62.5 mM Tris-HCl, pH 6.8

100 mM 2-β-mercaptoethanol

2 % SDS

REFERENCES

Agulnik, S. I., Bollag, R. J. and Silver, L. M. (1995). Conservation of the T-box gene family from Mus musculus to Caenorhabditis elegans. *Genomics* **25**, 214-9.

Agulnik, S. I., Garvey, N., Hancock, S., Ruvinsky, I., Chapman, D. L., Agulnik, I., Bollag, R., Papaioannou, V. and Silver, L. M. (1996). Evolution of mouse T-box genes by tandem duplication and cluster dispersion. *Genetics* **144**, 249-54.

Amacher, S. L., Draper, B. W., Summers, B. R. and Kimmel, C. B. (2002). The zebrafish T-box genes no tail and spadetail are required for development of trunk and tail mesoderm and medial floor plate. *Development* **129**, 3311-23.

Arnold, S. J., Stappert, J., Bauer, A., Kispert, A., Herrmann, B. G. and Kemler, R. (2000). Brachyury is a target gene of the Wnt/beta-catenin signaling pathway. *Mech Dev* **91**, 249-58.

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.

Bailey, A. M. and Posakony, J. W. (1995). Suppressor of hairless directly activates transcription of enhancer of split complex genes in response to Notch receptor activity. *Genes Dev* **9**, 2609-22.

Bamshad, M., Le, T., Watkins, W. S., Dixon, M. E., Kramer, B. E., Roeder, A. D., Carey, J. C., Root, S., Schinzel, A., Van Maldergem, L. et al. (1999). The spectrum of mutations in TBX3: Genotype/Phenotype relationship in ulnar-mammary syndrome. *Am J Hum Genet* **64**, 1550-62.

Bamshad, M., Lin, R. C., Law, D. J., Watkins, W. S. and Krakowiak, P. A. (1997). Mutations in human TBX3 alter limb, apocrine and genital development in ulnar-mammary syndrome. *Nat. Genet.* **16**, 311.

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 [corrected] cause limb and cardiac malformation in Holt-Oram syndrome. *Nat Genet* **15**, 30-5.

Basson, C. T., Huang, T., Lin, R. C., Bachinsky, D. R., Weremowicz, S., Vaglio, A., Bruzzone, R., Quadrelli, R., Lerone, M., Romeo, G. et al. (1999). Different TBX5 interactions in heart and limb defined by Holt-Oram syndrome mutations. *Proc Natl Acad Sci U S A* **96**, 2919-24.

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.
Begemann, G., Gibert, Y., Meyer, A. and Ingham, P. W. (2002). Cloning of zebrafish T-box genes tbx15 and tbx18 and their expression during embryonic development. *Mech Dev* **114**, 137-41.
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.
Birkenmeier, C. S., Gifford, E. J. and Barker, J. E. (2003). Normoblastosis, a murine model for ankyrin-deficient hemolytic anemia, is caused by a hypomorphic mutation in the erythroid ankyrin gene Ank1. *Hematol J* **4**, 445-9.
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.
Bray, N., Dubchak, I. and Pachter, L. (2003). AVID: A Global Alignment Program. *Genome Res.* **13**, 97-102.
Bray, S. and Furriols, M. (2001). Notch pathway: making sense of suppressor of hairless. *Curr Biol* **11**, R217-21.
Brou, C., Logeat, F., Lecourtois, M., Vandekerckhove, J., Kourilsky, P., Schweisguth, F. and Israel, A. (1994). Inhibition of the DNA-binding activity of Drosophila suppressor of hairless and of its human homolog, KBF2/RBP-J kappa, by direct protein-protein interaction with Drosophila hairless. *Genes Dev* **8**, 2491-503.
Brown, D. D., Binder, O., Pagratis, M., Parr, B. A. and Conlon, F. L. (2003). Developmental expression of the Xenopus laevis Tbx20 orthologue. *Dev Genes Evol* **212**, 604-7.
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.
Bruneau, B. G., Logan, M., Davis, N., Levi, T. and Tabin, C. J. (1999). Chamber-specific cardiac expression of Tbx5 and heart defects in Holt-Oram syndrome. *Dev. Biol.* **211**, 100.
Bruneau, B. G., Nemer, G., Schmitt, J. P., Charron, F., Robitaille, L., Caron, S., Conner, D. A., Gessler, M., Nemer, M., Seidman, C. E. et al. (2001). A murine model of Holt-Oram syndrome defines roles of the T-box transcription factor Tbx5 in cardiogenesis and disease. *Cell* **106**, 709-21.
Busch, S. J. and Sassone-Corsi, P. (1990). Dimers, leucine zippers and DNA-binding domains. *Trends Genet* **6**, 36-40.
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.
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.
Casey, E. S., Tada, M., Fairclough, L., Wylie, C. C., Heasman, J. and Smith, J. C. (1999). Bix4 is activated directly by VegT and mediates endoderm formation in Xenopus development. *Development* **126**, 4193-200.
Cattanach, B. and Rasberry, C. (1987). A new Curtailed mutation. *Mouse News Lett* **77**, 122.

Chapman, D. L., Agulnik, I., Hancock, S., Silver, L. M. and Papaioannou, V. E. (1996a). 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., Garvey, N., Hancock, S., Alexiou, M., Agulnik, S. I., Gibson-Brown, J. J., Cebra-Thomas, J., Bollag, R. J., Silver, L. M. and Papaioannou, V. E. (1996b). Expression of the T-box family genes, Tbx1-Tbx5, during early mouse development. *Dev Dyn* **206**, 379-90.
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.
Chesley, P. (1935). Development of the short-tailed mutant in the house mouse. *J. Exp. Zool* **70**, 429-459.
Chiba, S. (2006). Notch signaling in stem cell systems. *Stem Cells*.
Christ, B. and Ordahl, C. P. (1995). Early stages of chick somite development. *Anat Embryol (Berl)* **191**, 381-96.
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, F. L., Wright, C. V. and Robertson, E. J. (1995). Effects of the TWis mutation on notochord formation and mesodermal patterning. *Mech Dev* **49**, 201-9.
Couly, G. F., Coltey, P. M. and Le Douarin, N. M. (1992). The developmental fate of the cephalic mesoderm in quail-chick chimeras. *Development* **114**, 1-15.
Cunliffe, V. and Smith, J. C. (1992). Ectopic mesoderm formation in Xenopus embryos caused by widespread expression of a Brachyury homologue. *Nature* **358**, 427-30.
Cunliffe, V. and Smith, J. C. (1994). Specification of mesodermal pattern in Xenopus laevis by interactions between Brachyury, noggin and Xwnt-8. *Embo J* **13**, 349-59.
Dobrovolskaia-Zavadskaia, N. (1927). Sur la mortification spontanee de la queue chez la souris nouveau-nee et sur l'existence d'un caractere (facteur) hereditaire "non viable". *C. R. Acad. Sci. Biol.* **97**, 114.
Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M. and Dubchak, I. (2004). VISTA: computational tools for comparative genomics. *Nucl. Acids Res.* **32**, W273-279.
Fujimoto, H. and Yanagisawa, K. O. (1983). Defects in the archenteron of mouse embryos homozygous for the T-mutation. *Differentiation* **25**, 44-7.
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.
Gluecksohn-Schoenheimer, S. (1944). The Development of Normal and Homozygous Brachy (T/T) Mouse Embryos in the Extraembryonic Coelom of the Chick. *Proc Natl Acad Sci U S A* **30**, 134-40.
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.

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.
Gossler, A. and Hrabe de Angelis, M. (1998). Somitogenesis. *Curr Top Dev Biol* **38**, 225-87.
Greenwald, I. (1998). LIN-12/Notch signaling: lessons from worms and flies. *Genes Dev* **12**, 1751-62.
Griffin, K. J., Amacher, S. L., Kimmel, C. B. and Kimelman, D. (1998). Molecular identification of spadetail: regulation of zebrafish trunk and tail mesoderm formation by T-box genes. *Development* **125**, 3379-88.
Gunther, C. V. and Riddle, D. L. (2004). Alternative polyadenylation results in a truncated daf-4 BMP receptor that antagonizes DAF-7-mediated development in Caenorhabditis elegans. *J Biol Chem* **279**, 39555-64.
Halpern, M. E., Ho, R. K., Walker, C. and Kimmel, C. B. (1993). Induction of muscle pioneers and floor plate is distinguished by the zebrafish no tail mutation. *Cell* **75**, 99-111.
Halpern, M. E., Thisse, C., Ho, R. K., Thisse, B., Riggleman, B., Trevarrow, B., Weinberg, E. S., Postlethwait, J. H. and Kimmel, C. B. (1995). Cell-autonomous shift from axial to paraxial mesodermal development in zebrafish floating head mutants. *Development* **121**, 4257-64.
Hamaguchi, T., Yabe, S., Uchiyama, H. and Murakami, R. (2004). Drosophila Tbx6-related gene, Dorsocross, mediates high levels of Dpp and Scw signal required for the development of amnioserosa and wing disc primordium. *Dev Biol* **265**, 355-68.
Hames, B. D. (1998). Gel electrophoresis of proteins
a practical approach, (ed., pp. xx, 352 p. Oxford ; New York: Oxford University Press.
Harlow, E. and Lane, D. (1999). Using antibodies : a laboratory manual. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.
Hartenstein, A. Y., Rugendorff, A., Tepass, U. and Hartenstein, V. (1992). The function of the neurogenic genes during epithelial development in the Drosophila embryo. *Development* **116**, 1203-20.
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.
Haworth, K., Putt, W., Cattanach, B., Breen, M., Binns, M., Lingaas, F. and Edwards, Y. H. (2001). Canine homolog of the T-box transcription factor T; failure of the protein to bind to its DNA target leads to a short-tail phenotype. *Mamm Genome* **12**, 212-8.
Herrmann, B. G. (1991). Expression pattern of the Brachyury gene in whole-mount TWis/TWis mutant embryos. *Development* **113**, 913-7.
Herrmann, B. G. and Kispert, A. (1994). The T genes in embryogenesis. *Trends Genet* **10**, 280-6.
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.
Hilton, E., Rex, M. and Old, R. (2003). VegT activation of the early zygotic gene Xnr5 requires lifting of Tcf-mediated repression in the Xenopus blastula. *Mech Dev* **120**, 1127-38.
Hiroi, Y., Kudoh, S., Monzen, K., Ikeda, Y., Yazaki, Y., Nagai, R. and Komuro, I. (2001). Tbx5 associates with Nkx2-5 and synergistically promotes cardiomyocyte differentiation. *Nat Genet* **28**, 276-80.
Ho, R. K. and Kane, D. A. (1990). Cell-autonomous action of zebrafish spt-1 mutation in specific mesodermal precursors. *Nature* **348**, 728-30.

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.
Holland, P. W., Koschorz, B., Holland, L. Z. and Herrmann, B. G. (1995). Conservation of Brachyury (T) genes in amphioxus and vertebrates: developmental and evolutionary implications. *Development* **121**, 4283-4291.
Horb, M. E. and Thomsen, G. H. (1999). Tbx5 is essential for heart development. *Development* **126**, 1739-51.
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.
Hug, B., Walter, V. and Grunwald, D. J. (1997). tbx6, a Brachyury-related gene expressed by ventral mesendodermal precursors in the zebrafish embryo. *Dev Biol* **183**, 61-73.
Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., De Castro, E., Langendijk-Genevaux, P. S., Pagni, M. and Sigrist, C. J. A. (2006). The PROSITE database. *Nucl. Acids Res.* **34**, D227-230.
Hurlin, P. J., Steingrimsson, E., Copeland, N. G., Jenkins, N. A. and Eisenman, R. N. (1999). Mga, a dual-specificity transcription factor that interacts with Max and contains a T-domain DNA-binding motif. *Embo J* **18**, 7019-28.
Johnson, D. R. (1974). Hairpin-tail: a case of post-reductional gene action in the mouse egg. *Genetics* **76**, 795-805.
Kao, H. Y., Ordentlich, P., Koyano-Nakagawa, N., Tang, Z., Downes, M., Kintner, C. R., Evans, R. M. and Kadesch, T. (1998). A histone deacetylase corepressor complex regulates the Notch signal transduction pathway. *Genes Dev* **12**, 2269-77.
Kaufman, M. H. (1992). The atlas of mouse development. London ; San Diego: Academic Press.
Kinder, S. J., Tsang, T. E., Quinlan, G. A., Hadjantonakis, A. K., Nagy, A. and Tam, P. P. (1999). The orderly allocation of mesodermal cells to the extraembryonic structures and the anteroposterior axis during gastrulation of the mouse embryo. *Development* **126**, 4691-701.
Kispert, A. and Herrmann, B. G. (1993). The Brachyury gene encodes a novel DNA binding protein. *Embo J* **12**, 3211-20.
Kispert, A. and Herrmann, B. G. (1994). Immunohistochemical analysis of the Brachyury protein in wild-type and mutant mouse embryos. *Dev Biol* **161**, 179-93.
Kispert, A., Koschorz, B. and Herrmann, B. G. (1995a). The T protein encoded by Brachyury is a tissue-specific transcription factor. *Embo J* **14**, 4763-72.
Kispert, A., Ortner, H., Cooke, J. and Herrmann, B. G. (1995b). The chick Brachyury gene: developmental expression pattern and response to axial induction by localized activin. *Dev Biol* **168**, 406-15.
Knezevic, V., De Santo, R. and Mackem, S. (1997). Two novel chick T-box genes related to mouse Brachyury are expressed in different, non-overlapping mesodermal domains during gastrulation. *Development* **124**, 411-9.
Kohler, J. J. and Schepartz, A. (2001). Kinetic studies of Fos.Jun.DNA complex formation: DNA binding prior to dimerization. *Biochemistry* **40**, 130-42.
Koshiba-Takeuchi, K., Takeuchi, J. K., Arruda, E. P., Kathiriya, I. S., Mo, R., Hui, C. C., Srivastava, D. and Bruneau, B. G. (2006). Cooperative and antagonistic interactions between Sall4 and Tbx5 pattern the mouse limb and heart. *Nat Genet* **38**, 175-83.

Kraus, F., Haenig, B. and Kispert, A. (2001). Cloning and expression analysis of the mouse T-box gene Tbx18. *Mech Dev* **100**, 83-6.
Landschulz, W. H., Johnson, P. F. and McKnight, S. L. (1988). The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* **240**, 1759-64.
Li, Q. Y., Newbury-Ecob, R. A., Terrett, J. A., Wilson, D. I., Curtis, A. R., Yi, C. H., Gebuhr, T., Bullen, P. J., Robson, S. C., Strachan, T. et al. (1997). Holt-Oram syndrome is caused by mutations in TBX5, a member of the Brachyury (T) gene family. *Nat Genet* **15**, 21-9.
Lo Conte, L., Chothia, C. and Janin, J. (1999). The atomic structure of protein-protein recognition sites. *J Mol Biol* **285**, 2177-98.
Louvi, A., Arboleda-Velasquez, J. F. and Artavanis-Tsakonas, S. (2006). CADASIL: a critical look at a Notch disease. *Dev Neurosci* **28**, 5-12.
Louvi, A. and Artavanis-Tsakonas, S. (2006). Notch signalling in vertebrate neural development. *Nat Rev Neurosci* **7**, 93-102.
Lyon, M. F., Rastan, S., Brown, S. D. M. and International Committee on Standardized Genetic Nomenclature for Mice. (1995). Genetic variants and strains of the laboratory mouse. New York: Oxford University Press.
Mayor, C., Brudno, M., Schwartz, J. R., Poliakov, A., Rubin, E. M., Frazer, K. A., Pachter, L. S. and Dubchak, I. (2000). VISTA : visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics* **16**, 1046-7.
Miele, L., Miao, H. and Nickoloff, B. J. (2006). NOTCH signaling as a novel cancer therapeutic target. *Curr Cancer Drug Targets* **6**, 313-23.
Mitani, Y., Takahashi, H. and Satoh, N. (1999). An ascidian T-box gene As-T2 is related to the Tbx6 subfamily and is associated with embryonic muscle cell differentiation. *Dev Dyn* **215**, 62-8.
Mohr, O. L. (1919). Character changes caused by mutation of an entire region of a chromosome in *Drosophila*. *Genetics* **4**, 275-282.
Moskowitz, I. P., Pizard, A., Patel, V. V., Bruneau, B. G., Kim, J. B., Kupershmidt, S., Roden, D., Berul, C. I., Seidman, C. E. and Seidman, J. G. (2004). The T-Box transcription factor Tbx5 is required for the patterning and maturation of the murine cardiac conduction system. *Development* **131**, 4107-16.
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.
Nacke, S., Schafer, R., Habre de Angelis, M. and Mundlos, S. (2000). Mouse mutant "rib-vertebrae" (rv): a defect in somite polarity. *Dev Dyn* **219**, 192-200.
Nagy, A. (2003). Manipulating the mouse embryo : a laboratory manual. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.
Naiche, L. A., Harrelson, Z., Kelly, R. G. and Papaioannou, V. E. (2005). T-Box Genes In Vertebrate Development. *Annual Review of Genetics* **39**, 219-239.
O'Reilly, M. A., Smith, J. C. and Cunliffe, V. (1995). Patterning of the mesoderm in Xenopus: dose-dependent and synergistic effects of Brachyury and Pintallavis. *Development* **121**, 1351-9.
Opdecamp, K., Nakayama, A., Nguyen, M. T., Hodgkinson, C. A., Pavan, W. J. and Arnheiter, H. (1997). Melanocyte development in vivo and in neural crest cell cultures: crucial dependence on the Mitf basic-helix-loop-helix-zipper transcription factor. *Development* **124**, 2377-86.
Papaioannou, V. and Goldin, S. (2003). Introduction to the T-box genes and their roles in developmental signaling pathways. In Inborn Errors of Development. The Molecular Basis of

Clinical Disorders of Morphogenesis. Oxford Monogr. Med. Genet. No. 49, ed. CJ Epstein, RP Erickson, A Wynshaw-Boris, pp. 686–98. Oxford: Oxford Univ. Press.
Papaioannou, V. E. (2001). T-box genes in development: from hydra to humans. *Int Rev Cytol* **207**, 1-70.
Papapetrou, C., Putt, W., Fox, M. and Edwards, Y. H. (1999). The human TBX6 gene: cloning and assignment to chromosome 16p11.2. *Genomics* **55**, 238-41.
Quadrini, K. J. and Bieker, J. J. (2006). EKLF/KLF1 is ubiquitinated in vivo and its stability is regulated by activation domain sequences through the 26S proteasome. *FEBS Lett* **580**, 2285-93.
Rashbass, P., Cooke, L. A., Herrmann, B. G. and Beddington, R. S. (1991). A cell autonomous function of Brachyury in T/T embryonic stem cell chimaeras. *Nature* **353**, 348-51.
Rechsteiner, M. and Rogers, S. W. (1996). PEST sequences and regulation by proteolysis. *Trends Biochem Sci* **21**, 267-71.
Reim, I., Lee, H. H. and Frasch, M. (2003). The T-box-encoding Dorsocross genes function in amnioserosa development and the patterning of the dorsolateral germ band downstream of Dpp. *Development* **130**, 3187-204.
Ruvinsky, I., Silver, L. M. and Ho, R. K. (1998). Characterization of the zebrafish tbx16 gene and evolution of the vertebrate T-box family. *Dev Genes Evol* **208**, 94-9.
Saga, Y. and Takeda, H. (2001). The making of the somite: molecular events in vertebrate segmentation. *Nat Rev Genet* **2**, 835-45.
Sambrook, J. and Russell, D. W. (2001). Molecular cloning : a laboratory manual. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.
Schulte-Merker, S., Ho, R. K., Herrmann, B. G. and Nusslein-Volhard, C. (1992). The protein product of the zebrafish homologue of the mouse T gene is expressed in nuclei of the germ ring and the notochord of the early embryo. *Development* **116**, 1021-32.
Schulte-Merker, S., van Eeden, F. J., Halpern, M. E., Kimmel, C. B. and Nusslein-Volhard, C. (1994). no tail (ntl) is the zebrafish homologue of the mouse T (Brachyury) gene. *Development* **120**, 1009-15.
Schweisguth, F. (2004). Notch signaling activity. *Curr Biol* **14**, R129-38.
Searle, A. G. (1966). Curtailed, a new dominant T-allele in the house mouse. *Genet Res* **7**, 86-95.
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. USA* **85**, 180.
Showell, C., Binder, O. and Conlon, F. L. (2004). T-box genes in early embryogenesis. *Dev Dyn* **229**, 201-18.
Sims-Mourtada, J. C., Bruce, S., McKeller, M. R., Rangel, R., Guzman-Rojas, L., Cain, K., Lopez, C., Zimonjic, D. B., Popescu, N. C., Gordon, J. et al. (2005). The human AKNA gene expresses multiple transcripts and protein isoforms as a result of alternative promoter usage, splicing, and polyadenylation. *DNA Cell Biol* **24**, 325-38.
Singson, A., Leviten, M. W., Bang, A. G., Hua, X. H. and Posakony, J. W. (1994). Direct downstream targets of proneural activators in the imaginal disc include genes involved in lateral inhibitory signaling. *Genes Dev* **8**, 2058-71.
Sinha, S., Abraham, S., Gronostajski, R. M. and Campbell, C. E. (2000). Differential DNA binding and transcription modulation by three T-box proteins, T, TBX1 and TBX2. *Gene* **258**, 15-29.

Smith, J. C., Price, B. M., Green, J. B., Weigel, D. and Herrmann, B. G. (1991). Expression of a Xenopus homolog of Brachyury (T) is an immediate-early response to mesoderm induction. *Cell* **67**, 79-87.
Spector, D. L. and Goldman, R. D. (2003). Essentials from cells
a laboratory manual, (ed., pp. 1 CD-ROM. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
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.
Takada, N., Satoh, N. and Swalla, B. J. (2002). Expression of Tbx6, a muscle lineage T-box gene, in the tailless embryo of the ascidian Molgula tectiformis. *Dev Genes Evol* **212**, 354-6.
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.
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. (1988). Vertebral malformations. *Adv Anat Embryol Cell Biol* **112**, 1-99.
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.
Tun, T., Hamaguchi, Y., Matsunami, N., Furukawa, T., Honjo, T. and Kawaichi, M. (1994). Recognition sequence of a highly conserved DNA binding protein RBP-J kappa. *Nucleic Acids Res* **22**, 965-71.
Uchiyama, H., Kobayashi, T., Yamashita, A., Ohno, S. and Yabe, S. (2001). Cloning and characterization of the T-box gene Tbx6 in Xenopus laevis. *Dev Growth Differ* **43**, 657-69.
van Noort, M. and Clevers, H. (2002). TCF transcription factors, mediators of Wnt-signaling in development and cancer. *Dev Biol* **244**, 1-8.
Vancha, A. R., Govindaraju, S., Parsa, K. V., Jasti, M., Gonzalez-Garcia, M. and Ballestero, R. P. (2004). Use of polyethyleneimine polymer in cell culture as attachment factor and lipofection enhancer. *BMC Biotechnol* **4**, 23.
Weinmaster, G. (1997). The Ins and Outs of Notch Signaling. *Molecular and Cellular Neuroscience* **9**, 91-102.
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., 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, R. W., Ballantyne, C. M., Smith, C. W., Montgomery, C., Bradley, A., O'Brien, W. E. and Beaudet, A. L. (1993a). Gene targeting yields a CD18-mutant mouse for study of inflammation. *J Immunol* **151**, 1571-1578.
Wilson, V. and Beddington, R. (1997). Expression of T protein in the primitive streak is necessary and sufficient for posterior mesoderm movement and somite differentiation. *Dev Biol* **192**, 45-58.
Wilson, V. and Conlon, F. L. (2002). The T-box family. *Genome Biol* **3**, REVIEWS3008.

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.
Wilson, V., Rashbass, P. and Beddington, R. S. (1993b). Chimeric analysis of T (Brachyury) gene function. *Development* **117**, 1321-31.
Wolpert, L. (1998). Principles of development. London ; New York; Oxford;: Current Biology ; Oxford University Press.
Xu, W., Gong, L., Haddad, M. M., Bischof, O., Campisi, J., Yeh, E. T. and Medrano, E. E. (2000). Regulation of microphthalmia-associated transcription factor MITF protein levels by association with the ubiquitin-conjugating enzyme hUBC9. *Exp Cell Res* **255**, 135-43.
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.
Yanagisawa, K. O. and Fujimoto, H. (1977). Differences in rotation-mediated aggregation between wild-type and homozygous Brachyury (T) cells. *J Embryol Exp Morphol* **40**, 277-83.
Yanagisawa, K. O., Fujimoto, H. and Urushihara, H. (1981). Effects of the brachyury (T) mutation on morphogenetic movement in the mouse embryo. *Dev Biol* **87**, 242-8.
Yasuhiko, Y., Haraguchi, S., Kitajima, S., Takahashi, Y., Kanno, J. and Saga, Y. (2006). Tbx6-mediated Notch signaling controls somite-specific Mesp2 expression. *PNAS* **103**, 3651-3656.
Zaragoza, M. V., Lewis, L. E., Sun, G., Wang, E. and Li, L. (2004). Identification of the TBX5 transactivating domain and the nuclear localization signal. *Gene* **330**, 9.