THE UNIVERSITY OF CHICAGO

TBX5: A KEY REGULATOR OF CARDIOPULMONARY DEVELOPMENT AND DISEASE

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

COMMITTEE ON DEVELOPMENT, REGENERATION, AND STEM CELL BIOLOGY

BY

JEFFREY DAVID STEIMLE

CHICAGO, ILLINOIS

AUGUST 2019

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

First, I have to thank and acknowledge the support and mentorship from my PhD advisor, Ivan Moskowitz. Ivan has provided me with innumerable support and guidance throughout my PhD. I chose to attend the University of Chicago, because there was no particular lab I was interested in prior to starting; I wanted to have a fresh slate. When I heard Ivan give all-stars during Fall Quarter of my first year, I knew I had to rotate in his lab. From there, Ivan has allowed me the freedom and opportunity to explore my interests, develop my own questions, and gain new skills. I have enjoyed my 8 years in the lab and graduate school, and look forward to continue working with Ivan in the field as I transition into my postdoc and independent career.

I would like to thank the members of my committee, Robert Ho, Marcelo Nobrega, and Neil Shubin. The three of them have provided me a secondary line of support, and helped keep my PhD on a timely track. Additionally, I feel like I chose an amazing committee, how many graduate students can say a committee member brought data to their committee meeting?

I would like to acknowledge the members of the Moskowitz Lab who have not only provided me with help and support over the years, but who have also become my friends. First, I would like to thank Andrew Hoffmann, who worked closely with me during my rotation and helped me set a strong research foundation in the lab. He instilled in me the importance of being able to do a good *in situ* hybridization, because when nothing else is working, it is reassuring to know that I could always do an *in situ* and have it work. I would like to next like to thank Ozanna Burnicka-Turek, who welcomed me my first day, and has always been a constant in the lab. For good or bad, I knew early on that she would be in the lab during my entire PhD; I just don’t think either of us planned to be here this many years. I would like to thank Chul Kim, who I have also had the pleasure of working with the entirety of my PhD. I have enjoyed the times

Chul comes to my desk looking to discuss his interesting ideas and plans for the next experiment forever on his quest for the perfect differentiation, which is what I have always considered to be "Chul's Dream." I would next like to thank Junghun Kweon, first for performing the RNA-seq experiments alongside me that would set the groundwork for my entire PhD, and second for giving up the corner desk to take David's bench; he helped set unrealistic expectations on career advancement and the coveted "corner office." I would like to thank Holly Yang, who, along with Junghun, joined the group around the same time as me. Holly has helped provide me the support and encouragement to learn bioinformatics and computer science early on, and helped me become an independent investigator in these regards. I have also valued assisting her with grant proposals over the years, because they have helped me gain practice and understanding into the writing process. I would next like to thank my classmate Erika Hanson. Although Erika's time in the lab was short, it was nice having another early-career graduate student in the lab to talk and discuss our plans and goals. I would like to thank Rajiv Nadadur. I will forever cherish our morning coffees together, throwing all sorts of crazy ideas and experiments at the wall and seeing what sticks. I would say 99% of the ideas we came up with went nowhere, but in the remaining 1%, we came up some of the best experiments that opened up entire new avenues of study in the lab. I would like to thank Alex Guzzetta, who has not only kept me abreast of the latest and greatest advances in science technology, but also in the best new fantasy/sci-fi books and the latest, binge-worthy TV. I would like to thank Jenna Bekeny for her time and ability to make RNA appear out of the aether. I would next like to thank Kohta Ikegami. With his outside experience, Kohta has shown me new avenues of investigation and ways of thinking. I appreciate the mentorship Kohta has provided over the last year or so. I would like to thank Ariel Rydeen who has filled in for Rajiv in these later years providing an avenue to talk both about science and

life. I would like to thank everyone else in the lab who I have interacted with over the years, you have provided me interesting conversations both scientific and not.

I would like to thank my DRSB classmates, Steve Briscoe, Phil Lee, Danielle Howell, and Jiajie Xu, for the support and camaraderie during the early years of graduate school. I would also like to thank everyone else in the DRSB program, faculty, administration, and students alike. I had originally applied to the Human Genetics program, but I feel the DRSB program has provided me the right mixture of everything.

I would like to acknowledge and thank the collaborators with whom I have worked with over the years. I would like to first thank Aaron Zorn and Scott Rankin at Cincinnati Children's. I have greatly appreciated the collaboration with the Zorn lab, and I have enjoyed receiving what I call "Scott emails" with the freshest data and new ideas. I would next like to thank Frank Conlon and Lauren Waldron at UNC-Chapel Hill for our on-going work examining the roles and interactions of the cardiogenic transcription factors. I would like to thank Bjarke Jensen at AMC, with whom I have enjoyed discussing the evolution of the heart. I would like to thank both Steve Vokes of UT-Austin and Roger Reeves of Johns Hopkins for inviting me to and opening up their labs to me. I would like to thank Michael Broman at the University of Chicago; I have valued our interactions starting all the way back to my rotation in the Svensson lab. I would like to thank Lindsey Montefiori of the Nobrega lab for our discussions over the years; I too want know what *PITX3* is doing in the heart. I would like to thank Erin Boyle-Anderson of the Ho lab for providing me the experience and expertise on *tbx5a*/*b* in the zebrafish.

I would like to thank all of the core facilities on campus, which provided the resources necessary to complete many aspects of the work. I would like to specifically thank the HTRC,

especially Xin Jiang, Christy Schmehl, Terri Li, and Can Gong; Bill Buikema and the whole gang in the DNA core; and lastly Pieter Faber and the Genomics Facility.

I would like to thank the members of the 5th floor KCBD. I have spent nearly 8 full years working, eating, complaining, and celebrating alongside all of them. Although we were not in the same lab, we were all in this together. I would also like to thank the building crew, including the early morning cleaning staff and the dock guys.

I would like to acknowledge and thank two amazing sources of funding during my PhD. First, I would like to thank Ben Glick, Phoebe Rice, and Catherine Will for allowing me three years of funding on the Molecular and Cellular Biology training grant. I would like to next thank Beth McNally, Jim Liao, and Pat Ongusaha for providing me an additional three years of support through the Cardiovascular Research Training Grant.

I would like to thank Hope Hollocher and the rest of her lab during my time as an undergraduate at the University of Notre Dame. Hope provided me a strong understanding in the philosophy of science and thinking about the big questions. I will forever value the time and effort she spent providing assistance to get to this level.

Finally, I would like to thank my family. I would like to thank my parents, Dave and Sue Steimle, for providing me the foundation for all of this. None of this would exist without their love and support these past 29 years of my life. I would like to thank my wife, Allison Nipper, my biggest supporter throughout my PhD, even as she worked on her own. Allison has read, listened to, or edited almost every email I've ever sent, every paper or abstract I've ever written (including this!), and every talk I've ever given. Allison has had to listen to me talk about my science, and hers, every day on the way home for almost 4 years straight, even when she doesn't want to think about work anymore. Not to be outdone by Allison's acknowledgements, I too

would like to thank our cat, Mr. Pickles, a constant distraction and a “friendly face” to greet us every day when we get home.

CHAPTER I: *TBX5*: A KEY REGULATOR OF HEART DEVELOPMENT

Abstract

TBX5 is a member of the T-box transcription factor family and is primarily known for its role in cardiac and forelimb development. Human patients with dominant mutations in *TBX5* are characterized by Holt-Oram syndrome, and show defects of the cardiac septa, cardiac conduction system, and the anterior forelimb. The range of cardiac defects associated with *TBX5* mutations in humans suggests multiple roles for the transcription factor in cardiac development and function. Animal models demonstrate similar defects and have provided a useful platform for investigating the roles of *TBX5* during embryonic development. During early cardiac development, *TBX5* appears to act primarily as a transcriptional activator of genes associated with cardiomyocyte maturation and upstream of morphological signals for septation. During later cardiac development, *TBX5* is required for patterning of the cardiac conduction system and maintenance of mature cardiomyocyte function. A comprehensive understanding of the integral roles of *TBX5* throughout cardiac development and adult life will be critical for understanding human cardiac morphology and function.

TBX5, a member of the T-box family of transcription factors

Ever since the association of *TBX5* mutations with Holt-Oram syndrome [1, 2], the *TBX5* gene has been a source of study, particularly with respect to cardiac and limb development. This review covers progress since those initial reports, published two decades ago.

The T-box family of transcription factors share a common T-box DNA binding domain [3], and are named after the founding member, *T*, which encodes the transcription factor Brachyury [4, 5]. The T-box domain is approximately 170-200 amino acids in length [3, 6-8],

binds DNA directly [5], and is required for transcriptional activity [9]. All members of the T-box gene family appear to bind the DNA consensus motif AGGTGHBA [9-15]. While some members, such as *T* and *EOMES*, are only able to bind palindromic sequences as dimers [9, 10, 16, 17], others, including *TBX2*, *TBX3*, and *TBX5* can bind individual motifs as monomers [17-20].

In the human genome, there are 17 coding genes that fall within five subfamilies of the T-box family [7, 21], most of which show sequence similarity between vertebrates and invertebrates [5-7, 11, 21, 22]. The human genes *TBX2*, *TBX3*, *TBX4*, and *TBX5* belong to one subfamily, and are homologous to the Drosophila gene *omb* [7]. These four genes likely originated through tandem duplication of the ancestral gene through unequal crossover to form the ancestral *TBX2/3* and *TBX4/5* [7, 22] followed by cluster duplication prior to the divergence of bony fish and tetrapods approximately 400 million years ago [7, 23]. This second duplication event generated one cluster containing *TBX2* and *TBX4* and a second containing *TBX3* and *TBX5*, which are located on human chromosome 17 and 12, respectively [7, 23].

The T-box family gene *TBX5* encodes a 518 amino acid protein with a 180 amino acid T-box domain located between amino acid residue 56 and 236 [1, 2]. TBX5 contains two nuclear localization sequences (NLS): NLS1 located within the T-box domain (amino acid 78-90), and NLS2 located outside the T-box domain on the C-terminal end (325-340) [24, 25]. While each NLS is sufficient to drive nuclear localization, they appear to work cooperatively [24]. As a transcription factor, TBX5 also contains a transactivation domain located from amino acid 339-379 with functional requirement of amino acids 349-351 [25]. The sequence of amino acids 152-160 have been proposed to act as a nuclear export signal through the CRM1 export pathway by which TBX5 subcellular localization can be regulated through binding with the PDLIM7 protein

[26-28]; however, this theory remains controversial as the crystal structure of TBX5 suggests this domain would be located on the inside of the protein and inaccessible without major protein rearrangements [29]. Aside from these domains, TBX5 also contains several other protein-protein interaction domains, discussed in detail below.

In addition to the best described isoform, sometimes referred to as TBX5a [30], there have been four additional isoforms described in the literature [30, 31]. These isoforms are derived from alternative splicing within the *TBX5* locus and result in proteins of varying lengths, including either an N-terminal or a C-terminal truncated form as well as forms with varying C-terminal modifications [30, 31]. Interestingly, all described isoforms retain the T-box domain [31]. The alternative isoforms show differential expression and activity, including antagonism of TBX5a, and further investigation into these isoforms will be important for understanding the roles of *TBX5* in human development and health [31].

TBX5 Expression

The broad spatiotemporal expression domains of *TBX5* during development appear to be generally conserved throughout vertebrate evolution and consist of the heart, forelimb, and retina [32-38]; however, some tissue specific expression differences occur between species. First, we will examine the expression domains of *TBX5* in the common tetrapod models as well as humans, with the greatest emphasis on sub-cardiac domains, and then we will discuss the regulation of *TBX5* gene expression based on evidence from mouse and human studies.

TBX5 *Expression Domains in the Embryonic and Adult Heart*

The cardiac expression patterns of *TBX5* in human, mouse, chick, and frog are very similar. In human hearts, *TBX5* is expressed in the epicardium, myocardium, and endocardium of embryonic and adult hearts [39]. Human *TBX5* is expressed in the free walls and septa of all four chambers during development; however, atrial expression is much greater than ventricular, as seen in mouse and chicken [39]. *TBX5* is expressed in the embryonic atrioventricular (AV) node, and is excluded from the AV valves [39]. *TBX5* is expressed throughout the epicardium, but not in the endocardium of the left ventricle [39]. Much like in animal models, *TBX5* expression is absent from the developing outflow tract of the heart [39]. In human adults, *TBX5* expression is highest in the atrial appendages, followed by the lungs, left ventricle, and esophagus [40, 41].

In mice, *Tbx5* becomes abundantly expressed around E8.0 throughout the cardiac crescent [35], and this expression becomes restricted to the posterior portion of the forming heart tube, corresponding to the sinus venosa and future atria, between E8.25 and E8.5 [32, 35]. At E9.0, *Tbx5* expression expands throughout the future left ventricle [35]. Additionally, atrial expression of *Tbx5* is stronger than in the left ventricle, and in the ventricular free wall it is higher than in the trabeculae [35]. *Tbx5* is also expressed in the right ventricular trabeculae, but not free wall [35]. Expression of *Tbx5* in the left ventricle and atria is maintained throughout embryonic development [32]. During maturation of the mouse heart, like humans, *Tbx5* is expressed in and co-localizes with markers of the cardiac conduction system, including the atrioventricular bundle and bundle branch [42]. Genetic inducible fate-mapping demonstrated that left ventricular *Tbx5* expression arises from the first heart field, specified prior to morphogenesis of the heart, whereas atrial and atrial septum *Tbx5* expression arises from

Mef2cAHF+ second heart field domain, suggesting potential independent roles of *Tbx5* in the first and second heart fields during development [43].

In cardiac development of the chick, *Tbx5* expression is first detected throughout the entire bilateral cardiac primordia [35]. This expression is maintained following fusion of the heart tube along the entire rostrocaudal length [34], but adopts an anterior-to-posterior gradient shortly after [35]. Although there appears to be a gradient to the expression, *Tbx5* is expressed throughout the whole heart during cardiac looping [34]. After looping is complete, *Tbx5* expression remains in the entire heart except the outflow tract, and this is the only major difference between mouse and chick heart expression. However, as cardiac development and maturation proceed, expression of *Tbx5* is restricted from the right ventricle, similar to the expression pattern in the mouse embryo [35].

In *Xenopus*, the earliest expression domains of *tbx5* are in two lateral stripes, corresponding to the cardiac primordia, on either side of the embryo and continue to be expressed in the migrating precardiac mesoderm [36, 37]. Similar to chick cardiogenesis, after fusion at the midline and formation of the early heart tube, *tbx5* is expressed throughout most of the cardiac tissue including the sinus venosus/inflow tract of the heart [36, 37]. As development continues, *tbx5* expression is lost from the most anterior structure, the bulbus cordis, and is strongly detected in the posterior regions of the heart, while it is also maintained in the ventricle [36, 37]. *tbx5* is expressed robustly in both the endocardium and myocardium and is detected in the epicardium [36].

Extra-cardiac TBX5 *Expression*

In addition to the cardiac expression domains, *TBX5* is expressed in many non-cardiac tissues. Perhaps the best studied expression domain outside of the heart is that of the developing forelimb. *Tbx5* is expressed in the lateral plate mesoderm giving rise to the forelimb starting at E8.8 of mouse embryonic development [33] and is robustly expressed in the forelimb bud at E9.5 [32, 33]. Expression throughout the developing limb is maintained until E11.5, when it then becomes restricted to the proximal portion of the forelimb [33]. *Tbx5* is also expressed in the periochondrium of the forelimb at E13.5 [33].

Outside of the heart and forelimb, *TBX5* expression has been reported in several notable domains during development. The earliest reported expression domain for *Tbx5* during mouse development is in the allantois at E7.5 where it is transiently co-expressed with *Tbx4* [32]. Expression of *Tbx5* in the allantois has been suggested to be a mammalian-specific trait, as transcription of *Tbx5* is never observed in the allantois of chick embryos [34]. *Tbx5* is also expressed in the optic vesicle and the neural retina of the developing eye in mouse, chick, and *Xenopus* where it is co-expressed with the other members of the *omb* family of T-box genes [32, 34, 36, 37]. Additionally, *Tbx5* is expressed in the mesenchyme of the mandibular arch, the trachea, and the lung, as well as the body wall of the thorax [32, 34]. In both mouse and chicken, expression of *Tbx5* has been reported in the genital papilla [32, 34]. Specific to avian development, *Tbx5* expression is observed in the notochord during mid-embryonic development [34].

Transcriptional Regulation of TBX5

The mechanisms governing spatio-temporal regulation of *TBX5* have begun to be addressed by defining cis-regulatory elements driving *TBX5* expression in the embryo and adult. Preliminary investigations have identified several elements that drive distinct spatial domains during development; however, this area is ripe for future investigations. Tiling experiments have identified three enhancers associated with *in vivo* expression of *Tbx5* in the mammalian heart [44, 45]. The first enhancer, corresponding to hg19 chr12:114,463,712–114,464,080, drives expression of a reporter construct in both ventricular and atrial myocardium of E11.5 mouse hearts [44]. The second enhancer, hg19 chr12:114,701,207-114,704,691, drives expression in the posterior portion of the heart, including the ventricles, interventricular septum, and atrioventricular canal [44]. Additionally, this second enhancer contains a low-frequency SNP that abrogates the enhancer's ability to drive expression [44]. While this SNP was predicted to disrupt a TAL1 binding site, *Tal1* is not expressed in the myocardium, suggesting other members of the basic helix-loop-helix E-box binding transcription factors may be driving expression of this enhancer [44]. The third *Tbx5* enhancer, hg19 chr12:114,853,271-114,858,238, is sufficient to drive expression in the ventricles, interventricular septum and atrioventricular canal as well as the atria [44].

In addition to the identified cardiac enhancers, there have been two additional enhancers identified that regulate limb expression. The first is located within intron 2 of *Tbx5* and drives expression within the lateral plate mesoderm of the forelimb, but not the heart [45]. This forelimb enhancer is regulated in part through *Hox4*/*5* genes, expressed in the region of the lateral plate mesoderm that gives rise to the forelimb, and has been proposed as the mechanism by which forelimb *Tbx5* expression is positionally defined along the anterior-posterior body axis

[45]. The second forelimb enhancer identified is known as CNS12 and is located approximately 120kbp downstream of the *Tbx5* coding region [46]. The CNS12 enhancer drives expression in the lateral plate mesoderm and is sufficient to drive Tbx5 expression for forelimb formation [46]. Taken together, *Tbx5* expression in the developing heart and forelimb appears to be driven by distinct cis-regulatory elements, although the factors and transcriptional complexes that control the expression of these enhancers have yet to be uncovered.

Recent insight into the regulation of *Tbx5* expression has come from investigations of the local three-dimensional architecture of the *Tbx5* locus and its neighboring genes: *Rbm19*, *Tbx3*, and *Med13l* [47, 48]. Circular chromosome conformation capture with sequencing (4C-seq) data from the viewpoints of the *Tbx3* and *Tbx5* promoters, and the CTCF binding site between the two loci, suggest that the loci are in contact and that putative cis-regulatory elements for each gene are located almost exclusively within their own loci with CTCF sites acting as a regulatory barrier (Figure I.1) [47]. Additionally, 4C-seq from the viewpoint of *Rbm19*, the nearest gene 3' of *Tbx5*, suggests partially overlapping regulatory elements [47]. Together, this suggests that most of the cis-regulatory information for *Tbx5* expression is located in the 375kbp region between the CTCF sites demarcating the boundaries of the *Tbx3*/*Tbx5* and the *Tbx5*/*Rbm19* loci (Figure I.1) [47].

In addition to transcriptional regulation, *Tbx5* has also been shown to be regulated through microRNA-dependent mechanisms [49]. In a screen of candidate human microRNAs, MiR-10a and MiR-10b were shown to bind to the 3'-UTR of *Tbx5* and inhibit its translation [49]. More recently, it has been shown that regulation through these two microRNAs may play a role in adult conduction defects and pathological remodeling in disease-state hearts [50].

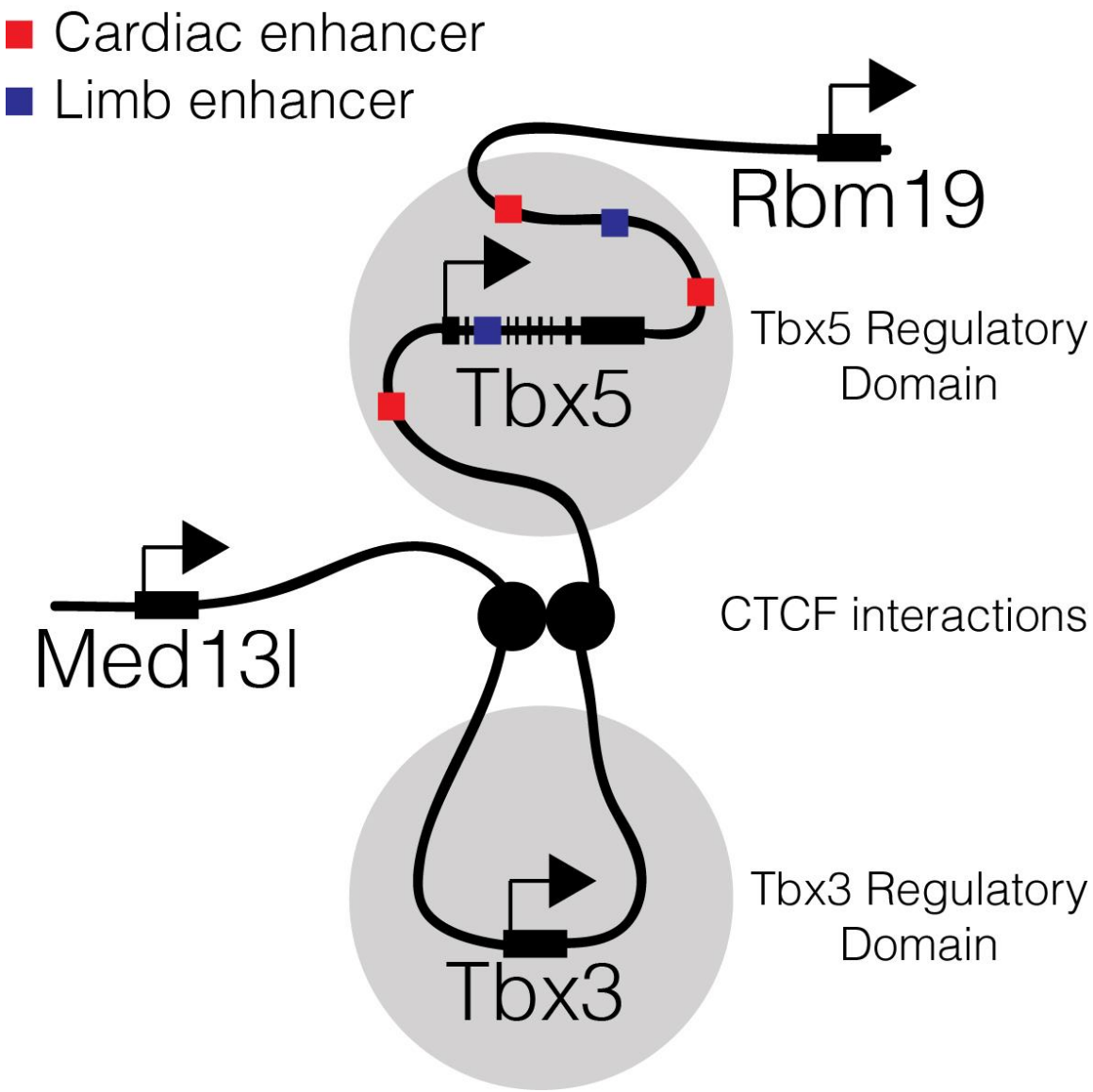


Figure I.1. Representation of *Tbx5* regulatory domain

Chromatin capture techniques suggest interactions between the promoter of *Tbx5* and *cis*-regulatory elements are limited to a 375kbp region demarcated by functional CTCF sites located between *Rbm19* and *Tbx5* and *Tbx5* and *Tbx3* [47].

TBX5 Haploinsufficiency: Holt-Oram Syndrome

Holt-Oram syndrome is an autosomal dominant disorder caused primarily by dominant mutations in *TBX5*. Holt-Oram syndrome is a clinical diagnosis that includes completely penetrant, variably expressed upper-limb malformations including pre-axial radial ray anomalies and congenital heart defects, typically septal and/or conduction defects [51]. Although sometimes difficult to detect, upper-limb malformations are fully penetrant, while structural cardiac defects occur in 76% of patients with Holt-Oram syndrome [51-54]. Holt-Oram syndrome affects 1 in 100,000-135,000 live births in European populations [55, 56], although defects can occur in any population [57-60]. Holt-Oram syndrome exhibits classic Mendelian inheritance for a dominant trait [1, 51, 52, 61], and the risk of non-affected parents with an affected proband giving rise to a second child with a de novo pathogenic mutation is the same as the average population [62]. Manifestations of upper-limb defects can include single or combinatorial abnormalities of the hand and digits, bones of the lower arm, humerus, or shoulder girdle [52, 54]. Defects of the hand and digits must include defects of the thumb for the Holt Oram diagnosis, while defects of the lower arm are associated primarily with the radius [51, 52, 54, 61]. Structural abnormalities of the heart can include secundum-type atrial septal defects, primum-type atrial septal defects, and/or ventricular septal defects [51, 52, 54]. Conduction system defects manifest as long PR interval, atrioventricular block, bundle branch block, bradycardia, sick sinus syndrome, and atrial fibrillation [51, 52, 54], and these conduction defects can occur in the absence of overt structural defects [52, 54]. Holt-Oram syndrome is not associated with defects of the lower limb, postaxial upper limb, gastrointestinal system, genitourinary, or nervous system, which if present suggest an alternative diagnosis [51, 52, 54, 61, 63].

Current evidence supports a model in which Holt-Oram syndrome is caused by *TBX5* haploinsufficiency [1, 2] with genetic abnormalities associated with *TBX5* coding or splice regulatory sequences underlying approximately 70% of patients meeting strict clinical diagnoses [61, 63]. In the latest collection of the Human Gene Mutation Database, there have been 103 reported mutations in coding, splicing, or regulatory sequences of *TBX5*, which result in Holt-Oram syndrome or other cardiac defects [64], recently reviewed in Yamak et al. 2015. Additionally, there have been 44 pathologic point mutations reported in the coding region of *TBX5* [64]. Similar to the 8 reported gross deletions [64], some of these are nonsense mutations resulting in highly truncated proteins that are thought to act as null alleles [1, 61, 65, 66]. Missense mutations have been reported throughout much of the T-box domain, typically resulting in transcriptional decrements [53, 61, 67, 68]. There are also several reported mutations that result in mis-splicing or alternative splicing [53, 61, 69-72]. Interestingly, duplications of *TBX5* are pathogenic, resulting in atypical Holt-Oram syndrome [57, 73]. Additionally, a patient with a homozygous, single-base pair mutation within a cis-regulatory element controlling *TBX5* displayed decreased *TBX5* expression and non-syndromic congenital heart disease, raising the possibility that the etiology of some of the remaining 30% of Holt-Oram patients may result from cis-regulatory element mutations [44].

While all cases of Holt-Oram syndrome result in both cardiac and forelimb defects, several cases have been reported in which defects in either the heart [1, 2, 74, 75] or the limb [2, 74, 75] appear more severe in one tissue than the other, suggesting a potential tissue-specific role for distinct domains of the protein [76]. While the underlying mechanisms of these biased defects have not been identified, some proposed models include differences in binding partners and binding motif recognition [26, 53, 76-78].

Animal Model of Holt-Oram Syndrome

Investigations into the role of *TBX5* in cardiac development have been undertaken in most major animal model systems: mouse, chick, frog, and zebrafish. Each system provides a unique set of tools for investigating the role of *TBX5* in the developmental etiology of Holt-Oram syndrome. The most well characterized model of Holt-Oram syndrome is the mouse heterozygous for a *Tbx5* knockout allele. The *Tbx5*^{tm1Jse} mouse allele contains loxP sites surrounding exon 3, which encodes a portion of the T-box DNA-binding domain, and upon Cre-mediated recombination, will generate truncated *Tbx5* transcripts [20]. Germline deletion of exon 3 generates the *Tbx5*$^{tm1.1Jse}$ mouse [20]. Heterozygous *Tbx5*$^{tm1.1Jse}$ mice exhibit the characteristic haploinsufficient phenotype of Holt-Oram syndrome, including anterior defects of the forelimb, septal defects of the heart, and defects of cardiac conduction [20, 42]. The mouse animal model has provided a robust system in which to study Holt-Oram syndrome *in vivo* and will continue to provide a platform by which to study the disease and the role of Tbx5 in cardiac and limb development.

TBX5 in Cardiac Morphologic Development

The morphologic cardiac defects associated with Holt-Oram syndrome are most commonly malformations of the septa dividing the left and right sides of the heart [51, 52, 54, 61]. From our current understanding, the ontogeny of the septa dividing the ventricular and atrial chambers is quite different [79], and yet defects in both arise from haploinsufficiency of *Tbx5* [1, 20, 35, 80-83].

Ventricular Septum

The morphology of the ventricular septum depends heavily on the localization of *Tbx5* expression during development. Both the left and the right ventricles contribute equally towards the formation of the interventricular septum, suggesting that a balance of left and right contributions may underlie development of the septum [84]. During development of the ventricular chambers, *Tbx5* is unilaterally expressed on the left side, including the left side of the ventricular septum [20, 35, 81]. Overexpression of *Tbx5* bilaterally results in malformation of the ventricular chambers and absence of the ventricular septum [80, 81, 85]. However, it remains unclear whether *Tbx5*-dependent transcriptional regulation alone controls ventricular septum formation [80, 81, 84]. Two additional T-box family genes, *Tbx18* and *Tbx20*, are unilaterally expressed in the left and right ventricles, respectively [81, 84]. The boundary between *Tbx5*-positive, *Tbx20*-negative and *Tbx5*-negative, *Tbx20*-positive myocardium appears to demarcate the location of ventricular septation and shifts in the expression levels can result in ventricular septum abnormalities; however, the exact mechanism by which the interface between *Tbx5* and the other T-box family genes instructs formation of the ventricular septum remains unclear [80, 81].

5.2 Atrial Septum

Haploinsufficiency of *TBX5* in Holt-Oram patients results in atrial septal defects in approximately half of cases [35], and *Tbx5* haploinsufficient mice exhibit atrial septal defects approximately 40% of the time [20]. The atrial septum is derived from the second heart field contributing to the inflow tract of the heart [86-90]. *Sonic hedgehog* (*Shh*), secreted from the pulmonary endoderm, signals through GLI-dependent transcription factors and is essential for

atrial septation [82, 83, 86, 87]. *Gli1* genetically interacts with *Tbx5* in the second heart field to co-activate downstream targets, including *Osr1* and *Foxf1* [82, 83, 86, 87]. Furthermore, deletion of one or both copies of *Tbx5* from SHH-receiving cells results in primum-type atrial septal defects, which can be rescued through constitutive activation of hedgehog-signaling [83]. This supports a model in which *Tbx5* acts upstream of active Shh-signaling in the second heart field, and that both *Tbx5* and activating GLI factors co-regulate transcription at the top of a hierarchy of atrial septation genes [82, 83].

TBX5 in Cardiac Conduction System Development

The cardiac conduction system is a highly specialized network of cardiomyocytes within the heart that generate and transmit electrical impulses throughout the heart to coordinate contraction. A majority of Holt-Oram syndrome patients present with conduction-system abnormalities [51, 52, 54]. Evidence suggests that *Tbx5* plays three key roles in the cardiac conduction system: specification of the conduction system during development, regulation of the conduction system transcriptome, and maintenance of conduction system identity in the adult [42, 91, 92]. *Tbx5* and *Nkx2-5* genetically interact to specify the ventricular cardiac conduction system in a ventricular sub-domain with the highest expression of both factors [42, 93]. Furthermore, *Tbx5* and *Nkx2-5* are required to co-regulate the transcriptional repressor *Id2*, which is required for proper formation and function of the conduction system [91]. Although the *Tbx5*-dependent transcriptome in the conduction system has not been well-characterized to-date, *Tbx5* is required for the regulation of critical conduction system ion channels, *Gja5* and *Scn5a*, and removal of *Tbx5* from the adult ventricular conduction system results in loss of these ion channels and altered ventricular conduction system function [20, 42, 92, 94].

Homozygous *Tbx5* mutations reveal novel roles of TBX5

While much attention has been focused on understanding *Tbx5* haploinsufficiency as it relates to human disease, homozygous deletion of *Tbx5* in animal models reveals novel requirements for *Tbx5* not uncovered by haploinsufficiency.

Complete Loss of Mammalian Tbx5

As *Tbx5* haploinsufficiency is associated with the phenotype of Holt-Oram syndrome in mice and *Tbx5* null embryos die in utero around E10.5, most studies examining the role of *Tbx5* in cardiac morphogenesis and transcription have focused on the *Tbx5* null heterozygote. In contrast, the null state provides the opportunity to understand critical roles of *Tbx5* not observed in heterozygotes [20, 42, 82, 83, 95-98]. In the mouse, germline deletion of both copies of *Tbx5*, $Tbx5^{tm1.1Jse}/^{tm1.1Jse}$, results in embryonic lethality by E10.5 [20]. These animals exhibit a grossly abnormal, linear heart tube [20] and complete absence of the forelimb buds [20, 95]. The complete absence of the forelimb buds indicates a role for *Tbx5* in limb bud initiation, shown to be downstream of fibroblast growth factor (FGF)-signaling [95, 98, 99]. Similar to the germline homozygous null embryos, homozygous hypomorphs for *Tbx5* show embryonic lethality prior to E11.5, with hypoplastic left ventricles and sinoatrial structures [97]. Distinctly, unlike the *Tbx5* homozygous null phenotype, homozygous hypomorphs still undergo heart looping and rudimentary formation of the left and right atrial chambers [97]. These observations indicate important roles for *Tbx5* at sequential stages of cardiac development, although the distinctions between these roles have yet to be elucidated.

Zebrafish *tbx5a*/*tbx5b*

The *heartstrings* mutation is the first published mutation of *tbx5a* in zebrafish, and was found as part of a screen for recessive lethal mutations affecting cardiac function [100]. The *tbx5a*/*heartstrings* mutants or morpholino knockdown of *tbx5a* recapitulate some aspects of Holt-Oram syndrome including forelimb defects and conduction defects [100, 101]. The hearts of these animals appear to develop normally during early cardiac development, only later displaying defects starting with the failure of heart looping and subsequent deterioration of chamber myocardium and heart failure [100, 101].

More recently, a second copy of *tbx5* (*tbx5b*) was found in the genome of zebrafish [102]. While *tbx5a* is expressed in the eye, heart, and forelimb during early development, *tbx5b* is only expressed robustly in the eye and heart [102], suggesting possible redundant functions for the two *tbx5* genes during early heart and eye development., Using morpholinos, *tbx5a*, *tbx5b*, or *tbx5a*/*tbx5b* double knockdowns all result in the *heartstrings* phenotype, i.e. normal heart tube formation, bradycardia, and progressive deterioration/heart failure [100, 103]. Different downstream targets have been identified for *tbx5a* and *tbx5b* [103]. While *tbx5b* knockdown does not result in patterning defects of chamber formation or sinus venosus seen in either the *tbx5a* mutants or knockdown experiments, *tbx5b* knockdown results in abnormal expansion of two morphogenesis markers, *hand2* and *vcana*, similar to *tbx5a* [100, 103]. However, known direct targets of mammalian *TBX5* or zebrafish *tbx5a*, such as *bmp4*, *nppa*, *tbx2b*, and *hey2* [20, 28, 104-106], were not disrupted with *tbx5b* knockdown, and neither *tbx5a* overexpression nor *tbx5b* overexpression rescue the reciprocal knockdown, suggesting the role of *tbx5b* is different than that of *tbx5a* [103]. The apparent differences in *tbx5a* and *tbx5b* function suggest that

following gene duplication there may have been evolutionarily-beneficial subfunctionalization of the two copies.

The *TBX5* Gene Regulatory Network

As a T-box transcription factor, the primary role of TBX5 is thought to be the regulation of target gene transcription. Historically, TBX5 has thought to act a positive regulator of transcriptional activity; however, recent evidence suggests that TBX5 may have a role in both transcriptional activation and repression (Figure I.2). In this section, we will explore what is known about both roles.

Positive transcriptional activation by TBX5 and co-factors: Cardiomyocyte Specific Factors, Chromatin Modification, and Maturation

TBX5 has long been known to act as a positive regulator of transcription in heart development and cardiomyocyte maturation [20, 42, 108, 109]. Some of the earliest identified direct targets of TBX5 were *NPPA* (encoding ANF) and *GJA5* (encoding cx40), both of which are expressed in differentiating cardiomyocytes during development and are markers of cardiac chamber differentiation [20, 108, 110, 111]. *Gja5* and *Nppa* are both highly sensitive markers of TBX5 activity with a nearly complete loss of *Gja5* in *Tbx5*$^{tm1.1Jse/+}$ embryos and a graded response of *Nppa* across a *Tbx5* allelic series [20, 97].

The first identified interaction partner of TBX5 was the tinman transcription factor NKX2-5 [108]. Identified by a classic yeast two-hybrid screen, TBX5 and NKX2-5 interact through the highly conserved C terminus of NKX2-5, relying on four key amino acids in an α-helix, P139, D140, R150, and Q151 [20, 96, 108]. The interaction between TBX5 and NKX2-5

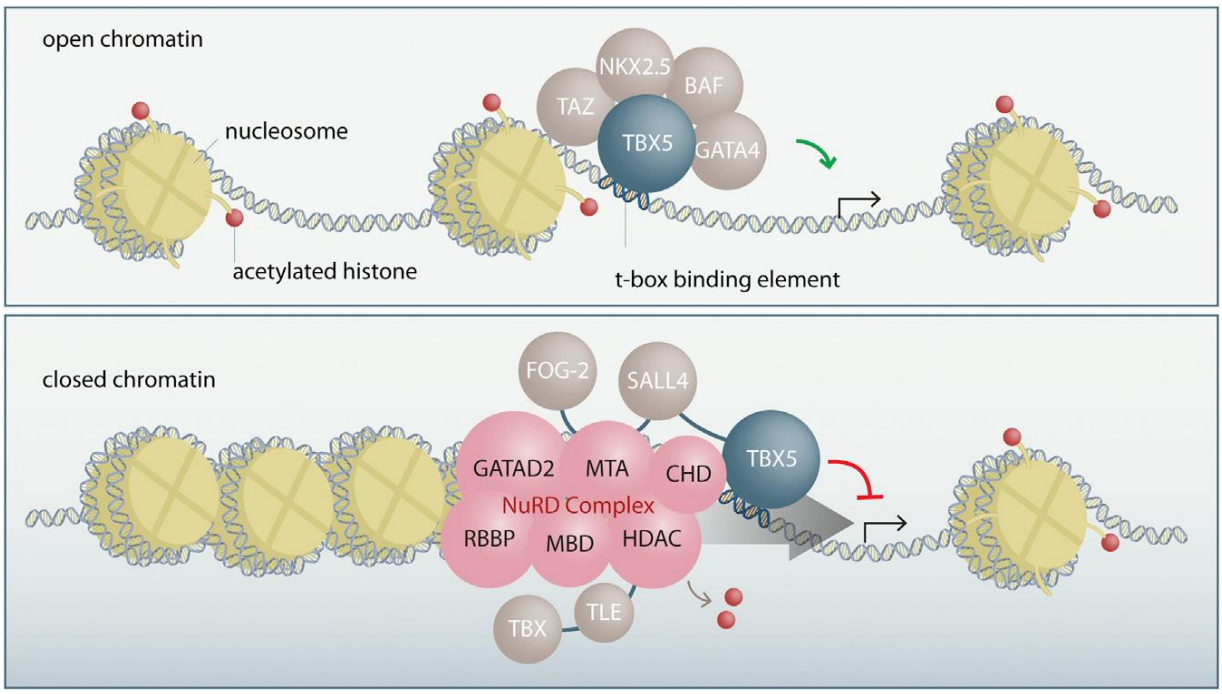


Figure I.2. Transcriptional regulation by TBX5

TBX5, through its interactions with other cardiac transcription factors, such as GATA4 and NKX2-5, and the BAF chromatin remodeling complex drive active transcription of target cardiac genes in regions of open chromatin (top panel). TBX5, through its interactions with the NuRD complex and other transcriptional repressors, such as SALL4, remodel chromatin to a closed state, which represses gene expression of non-cardiac genes (bottom panel). Reprinted from Developmental Cell, Vol 36(3), Boogerd & Evans, TBX5 and NuRD Divide the Heart, 242-244, 2016 with permission from Elsevier.

allows the proteins to synergistically activate targets such as *Nppa* through tandem transcription factor binding motifs [20, 108]. This interaction of TBX5 and NKX2-5 at tandem binding motifs induces bending in the DNA for transcriptional activity, a molecular mechanism by which synergistic activities of TBX5 and NKX2-5 interactions are determined at specific cis-regulatory elements [96]. Furthermore, the interaction between TBX5 and NKX2-5 is required to maintain fidelity in transcription factor binding throughout the genome, and the absence of either factor allows for inappropriate localization and activation of non-cardiac genes by the other [96]. To this point, several sites within TBX5 are important for the synergistic activation of *Nppa*, which can be abrogated by HOS mutation in the α-helix mentioned above as well as G80R and R237W [67, 78, 108].

Besides NKX2-5, TBX5 also shows direct interaction with other major cardiac transcription factors, including GATA4 [78, 112], GATA6 [112], TBX20 [113], MEF2C [114], and Myocardin [115]. The interaction between TBX5 and GATA4 was diminished by *GATA4* mutations causing non-syndromic congenital heart defects [78], as well as by *TBX5* Holt-Oram mutations causing heart defects but not those causing only limb defects [67, 78, 96]. Both GATA4-TBX5 and MEF2C-TBX5 interactions are required for synergistic activation of α-cardiac myosin heavy chain encoded by *MYH6* [112, 114], though GATA6-TBX5 protein interactions are not, suggesting that TBX5 interaction partners generate tissue and context specific gene expression [112].

TBX5 as a transcription factor appears to act as part of a multi-factor transcriptional complex for the activation and maintenance of cardiac lineage genes; however, the ability of transcription factors to regulate gene targets requires the ability of the factors to bind open chromatin. Addition of *Gata4*, *Mef2c*, and *Tbx5* to fibroblasts is sufficient to drive

reprogramming towards a cardiomyocyte fate [116, 117]. This ability to reprogram cells suggests that this core set of transcription factors may drive chromatin accessibility. In support of this supposition, it has been shown that TBX5 interacts with Baf60c and Brg1, encoded by *Smarcd3* and *Smarca4*, respectively, members of the SWI/SNF family of proteins involved in chromatin remodeling to drive mesodermal cells to cardiomyocyte fate *in vitro* and *in vivo* [118-120]. Furthermore, in *Tbx5* haploinsufficient mice, there is a loss of chromatin remodeling complexes at the promoters of *Tbx5*-dependent cardiac genes [118]. For another T-box family member, T-bet, encoded by *Tbx21*, it has been previously shown that T-bet, the Brg1-chromatin remodeling complex, and H3K27 demethylases physically interact [121], suggesting potential shared mechanisms with *Tbx5* in chromatin transitions seen in cardiac development [122].

TBX5-Mediated Repression: Inhibition of Non-cardiomyocyte Fate through Chromatin Remodeling

In addition to its positive role driving cardiac gene regulatory networks, recent genomic studies indicate that TBX5 acts as a direct transcriptional repressor during cardiac development where it is required for inhibition of inappropriate gene expression [14, 123]. While transcriptional repression by T-box factors in cardiac development and function has been well documented in the cases of TBX2 [18, 124] and TBX3 [19, 125-127], and it has been shown that TBX20 has roles in both transcriptional repression and activation [128-130], only recently has a repressive role for TBX5 been elucidated. Waldron et al. (2016) demonstrated that *Tbx5* inhibits non-cardiac gene regulatory programs, including neuronal networks during early cardiac development. Through biochemical and genetic interaction studies, it was shown that TBX5 protein interacts with the nucleosome remodeling and deacetylase (NuRD) complex during

embryonic development [14]. Similar to TBX20, the TBX5-NuRD interaction complex acts as an inhibitory mechanism by which TBX5 is able to repress non-cardiogenic gene expression in the heart [14, 130]. TBX5 physically interacts with the NuRD complex through an evolutionarily conserved α-helix domain located from amino acids 255-264 and disruption of this domain can result in Holt-Oram syndrome (e.g. S261C) [14, 74]. Using cardiomyocytes derived from murine embryonic stem cell differentiation, Luna-Zurita et al. (2016) demonstrated that TBX5 imparts specificity in cardiac transcription factor complexes by preventing off-target binding of other cardiac transcription factors. These findings suggest that the dual roles of T-box factors (i.e. TBX5 and TBX20 repression of non-cardiomyocyte fate) may be a more common theme in cardiac development than previously thought.

Direct targets of TBX5 regulation

As a member of the T-box family of transcription factors, TBX5 regulates transcription through direct interaction with DNA. In recent years, multiple groups have turned to chromatin immunoprecipitation with sequencing (ChIP-sequencing) to identify direct targets of TBX5 binding and regulation [13, 96]. The first ChIP-seq dataset was generated in the HL-1 atrial cardiomyocyte cell line using an overexpression construct of biotinylated TBX5, which identified over 56k binding sites within the genome [13]. The second set of data was generated using ChIP-exo technology in the context of mouse ES cell differentiation in both cardiac progenitors and cardiomyocytes, resulting in approximately 5k and 9k bindings sites, respectively [96]. While these datasets share many of the same locations, each also identifies many unique sites, suggesting that binding site information will need to be generated in each *Tbx5* expression context in order to understand the direct *Tbx5* transcriptome. For example,

whereas approximately 60% and 40% sites identified in the cardiac progenitor and cardiomyocyte mouse ES cell datasets are shared, only 4% of sites identified in the HL-1 dataset are shared with the mouse ES cell datasets. It is currently unclear to what degree the HL-1 dataset overestimates and the mouse embryonic stem cell datasets underestimate the total number of relevant binding sites, or whether both datasets overestimate functional binding events, HL-1 to a greater extent. Inclusion of additional markers, such as open chromatin, histone marks, and known TBX5 binding partners may allow the broad utilization of current datasets for identification of truly functional TBX5 binding sites.

To date, functionally confirmed direct targets of TBX5 are almost exclusively in genes implicated in cardiac proliferation, maturation, and function, including *Nppa*, *Gja5*, and *Scn5a*, [20, 42, 83, 91, 92, 97, 104, 108, 109, 131]. Interestingly, the direct targets mediating the morphogenesis requirement for TBX5 remain unknown. While some candidate target genes may act prior to morphological changes [82, 83], no direct mediators of morphology have been uncovered. TBX5 may therefore indirectly regulate these processes. With advances in genome-wide technology, understanding the basis by which TBX5 regulates morphological change will be key to understanding the role of TBX5 in cardiac development.

Concluding Remarks

The requirement of *TBX5* for normal human cardiac structure was identified over 20 years ago (Basson et al., 1997 and, Li et al, 1997); however, surprisingly, the mechanistic role of *TBX5* in cardiac development remains unclear. Little is known about the essential downstream targets of TBX5-mediated transcription in the context of cardiac development. Similarly, the complex temporal and spatial gene expression of *TBX5* has been mapped throughout

development and into adult life; however, the *cis*-regulatory architecture governing this expression is just beginning to be described. From a biochemical perspective, significant strides have been made in recent years to understand how TBX5 activates gene expression. However, the mechanisms by which TBX5 and its co-factors are targeted to specific loci, the temporal recruitment of TBX5 and its co-factors, the interplay between TBX5 and its co-transcriptional partners, and the mechanisms distinguishing active and repressive TBX5 activity are just recently coming into focus, and provide opportunities for exciting mechanistic studies. These areas of investigation will contribute to a broader understanding of the mechanisms underlying the requirement for *TBX5* in cardiac morphogenesis and more generally the transcriptional control of metazoan development.

References

1. Basson, C.T., et al., Mutations in human TBX5 [corrected] cause limb and cardiac malformation in Holt-Oram syndrome. Nature Genetics, 1997. 15(1): p. 30-5.

2. Li, Q.Y., et al., Holt-Oram syndrome is caused by mutations in TBX5, a member of the Brachyury (T) gene family. Nature Genetics, 1997. 15(1): p. 21-9.

3. Bollag, R.J., et al., An ancient family of embryonically expressed mouse genes sharing a conserved protein motif with the T locus. Nature Genetics, 1994. 7(3): p. 383-9.

4. Herrmann, B.G., et al., Cloning of the T gene required in mesoderm formation in the mouse. Nature, 1990. 343(6259): p. 617-22.

5. Pflugfelder, G.O., H. Roth, and B. Poeck, A homology domain shared between Drosophila optomotor-blind and mouse Brachyury is involved in DNA binding. Biochemical and Biophysical Research Communications, 1992. 186(2): p. 918-25.

6. Agulnik, S.I., R.J. Bollag, and L.M. Silver, Conservation of the T-box gene family from Mus musculus to Caenorhabditis elegans. Genomics, 1995. 25(1): p. 214-9.

7. Agulnik, S.I., et al., Evolution of mouse T-box genes by tandem duplication and cluster dispersion. Genetics, 1996. 144(1): p. 249-54.

8. Papaioannou, V.E., The T-box gene family: emerging roles in development, stem cells and cancer. Development, 2014. 141(20): p. 3819-33.

9. Kispert, A., B. Koschorz, and B.G. Herrmann, The T protein encoded by Brachyury is a tissue-specific transcription factor. EMBO Journal, 1995. 14(19): p. 4763-72.

10. Conlon, F.L., et al., Determinants of T box protein specificity. Development, 2001. 128(19): p. 3749-58.

11. Wilson, V. and F.L. Conlon, The T-box family. Genome Biology, 2002. 3(6): p. REVIEWS3008.

12. Mathelier, A., et al., JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. Nucleic Acids Research, 2016. 44(D1): p. D110-5.

13. He, A., et al., Co-occupancy by multiple cardiac transcription factors identifies transcriptional enhancers active in heart. Proceedings of the National Academy of Sciences of the United States of America, 2011. 108(14): p. 5632-7.

14. Waldron, L., et al., The Cardiac TBX5 Interactome Reveals a Chromatin Remodeling Network Essential for Cardiac Septation. Developmental Cell, 2016. 36(3): p. 262-75.

15. Jolma, A., et al., DNA-binding specificities of human transcription factors. Cell, 2013. 152(1-2): p. 327-39.

16. Muller, C.W. and B.G. Herrmann, Crystallographic structure of the T domain-DNA complex of the Brachyury transcription factor. Nature, 1997. 389(6653): p. 884-8.

17. Ghosh, T.K., et al., Characterization of the TBX5 binding site and analysis of mutations that cause Holt-Oram syndrome. Human Molecular Genetics, 2001. 10(18): p. 1983-94.

18. Carreira, S., et al., Brachyury-related transcription factor Tbx2 and repression of the melanocyte-specific TRP-1 promoter. Molecular Cell Biology, 1998. 18(9): p. 5099-108.

19. He, M., et al., Transcription repression by Xenopus ET and its human ortholog TBX3, a gene involved in ulnar-mammary syndrome. Proceedings of the National Academy of Sciences of the United States of America, 1999. 96(18): p. 10212-7.

20. Bruneau, B.G., et al., A murine model of Holt-Oram syndrome defines roles of the T-box transcription factor Tbx5 in cardiogenesis and disease. Cell, 2001. 106(6): p. 709-21.

21. Papaioannou, V.E. and L.M. Silver, The T-box gene family. Bioessays, 1998. 20(1): p. 9-19.

22. Ruvinsky, I., L.M. Silver, and J.J. Gibson-Brown, Phylogenetic analysis of T-Box genes demonstrates the importance of amphioxus for understanding evolution of the vertebrate genome. Genetics, 2000. 156(3): p. 1249-57.

23. Ruvinsky, I. and L.M. Silver, Newly identified paralogous groups on mouse chromosomes 5 and 11 reveal the age of a T-box cluster duplication. Genomics, 1997. 40(2): p. 262-6.

24. Collavoli, A., et al., TBX5 nuclear localization is mediated by dual cooperative intramolecular signals. Journal of Molecular and Cellular Cardiology, 2003. 35(10): p. 1191-5.

25. Zaragoza, M.V., et al., Identification of the TBX5 transactivating domain and the nuclear localization signal. Gene, 2004. 330: p. 9-18.

26. Camarata, T., et al., LMP4 regulates Tbx5 protein subcellular localization and activity. Journal of Cell Biologists, 2006. 174(3): p. 339-48.

27. Kulisz, A. and H.G. Simon, An evolutionarily conserved nuclear export signal facilitates cytoplasmic localization of the Tbx5 transcription factor. Molecular Cell Biology, 2008. 28(5): p. 1553-64.

28. Camarata, T., et al., Pdlim7 (LMP4) regulation of Tbx5 specifies zebrafish heart atrio-ventricular boundary and valve formation. Developmental Biology, 2010. 337(2): p. 233-45.

29. Stirnimann, C.U., et al., Structural basis of TBX5-DNA recognition: the T-box domain in its DNA-bound and -unbound form. Journal of Molecular Biology, 2010. 400(1): p. 71-81.

30. Georges, R., et al., Distinct expression and function of alternatively spliced Tbx5 isoforms in cell growth and differentiation. Molecular Cell Biology, 2008. 28(12): p. 4052-67.

31. Yamak, A., et al., Novel exons in the tbx5 gene locus generate protein isoforms with distinct expression domains and function. Journal of Biological Chemistry, 2015. 290(11): p. 6844-56.

32. Chapman, D.L., et al., Expression of the T-box family genes, Tbx1-Tbx5, during early mouse development. Developmental Dynamics, 1996. 206(4): p. 379-90.

33. Gibson-Brown, J.J., et al., Evidence of a role for T-box genes in the evolution of limb morphogenesis and the specification of forelimb/hindlimb identity. Mechanisms of Development, 1996. 56(1-2): p. 93-101.

34. Gibson-Brown, J.J., et al., Expression of T-box genes Tbx2-Tbx5 during chick organogenesis. Mechanisms of Development, 1998. 74(1-2): p. 165-169.

35. Bruneau, B.G., et al., Chamber-specific cardiac expression of Tbx5 and heart defects in Holt-Oram syndrome. Developmental Biology, 1999. 211(1): p. 100-8.

36. Horb, M.E. and G.H. Thomsen, Tbx5 is essential for heart development. Development, 1999. 126(8): p. 1739-51.

37. Showell, C., et al., Developmental expression patterns of Tbx1, Tbx2, Tbx5, and Tbx20 in Xenopus tropicalis. Developmental Dynamics, 2006. 235(6): p. 1623-30.

38. Takabatake, Y., T. Takabatake, and K. Takeshima, Conserved and divergent expression of T-box genes Tbx2-Tbx5 in Xenopus. Mechanisms of Development, 2000. 91(1-2): p. 433-7.

39. Hatcher, C.J., et al., Identification and localization of TBX5 transcription factor during human cardiac morphogenesis. Developmental Dynamics, 2000. 219(1): p. 90-5.

40. GTEx Consortium, The Genotype-Tissue Expression (GTEx) project. Nature Genetics, 2013. 45(6): p. 580-5.

41. Mele, M., et al., Human genomics. The human transcriptome across tissues and individuals. Science, 2015. 348(6235): p. 660-5.

42. Moskowitz, I.P., et al., The T-Box transcription factor Tbx5 is required for the patterning and maturation of the murine cardiac conduction system. Development, 2004. 131(16): p. 4107-16.

43. Devine, W.P., et al., Early patterning and specification of cardiac progenitors in gastrulating mesoderm. Elife, 2014. 3.

44. Smemo, S., et al., Regulatory variation in a TBX5 enhancer leads to isolated congenital heart disease. Human Molecular Genetics, 2012. 21(14): p. 3255-63.

45. Minguillon, C., et al., Hox genes regulate the onset of Tbx5 expression in the forelimb. Development, 2012. 139(17): p. 3180-8.

46. Adachi, N., et al., Regulatory evolution of Tbx5 and the origin of paired appendages. Proceedings of the National Academy of Sciences of the United States of America, 2016.

47. van Weerd, J.H., et al., A large permissive regulatory domain exclusively controls Tbx3 expression in the cardiac conduction system. Circulation Research, 2014. 115(4): p. 432-41.

48. Jin, F., et al., A high-resolution map of the three-dimensional chromatin interactome in human cells. Nature, 2013. 503(7475): p. 290-4.

49. Wang, F., et al., miR-10a and miR-10b target the 3'-untranslated region of TBX5 to repress its expression. Pediatric Cardiology, 2014. 35(6): p. 1072-9.

50. Torrado, M., et al., A MicroRNA-Transcription Factor Blueprint for Early Atrial Arrhythmogenic Remodeling. BioMed Research International, 2015. 2015: p. 263151.

51. Holt, M. and S. Oram, Familial heart disease with skeletal malformations. British Heart Journal, 1960. 22: p. 236-42.

52. Basson, C.T., et al., The clinical and genetic spectrum of the Holt-Oram syndrome (heart-hand syndrome). New England Journal of Medicine, 1994. 330(13): p. 885-91.

53. Basson, C.T., et al., Different TBX5 interactions in heart and limb defined by Holt-Oram syndrome mutations. Proceedings of the National Academy of Sciences of the United States of America, 1999. 96(6): p. 2919-24.

54. Newbury-Ecob, R.A., et al., Holt-Oram syndrome: a clinical genetic study. Journal of Medical Genetics, 1996. 33(4): p. 300-7.

55. Elek, C., M. Vitez, and E. Czeizel, [Holt-Oram syndrome]. Orvosi Hetilap, 1991. 132(2): p. 73-4, 77-8.

56. Barisic, I., et al., Holt Oram syndrome: a registry-based study in Europe. Orphanet Journal of Rare Diseases, 2014. 9: p. 156.

57. Kimura, M., et al., Novel TBX5 duplication in a Japanese family with Holt-Oram syndrome. Pediatric Cardiology, 2015. 36(1): p. 244-7.

58. Najjar, H., et al., Variability of the Holt-Oram syndrome in Saudi individuals. American Journal of Medical Genetics, 1988. 29(4): p. 851-5.

59. Al-Qattan, M.M. and H. Abou Al-Shaar, A novel missense mutation in the TBX5 gene in a Saudi infant with Holt-Oram syndrome. Saudi Medical Journal, 2015. 36(8): p. 980-2.

60. Ekure, E.N., et al., Holt-Oram syndrome with hypoplastic left heart syndrome in an African child. Nigerian Postgraduate Medical Journal, 2004. 11(3): p. 190-2.

61. McDermott, D.A., et al., TBX5 genetic testing validates strict clinical criteria for Holt-Oram syndrome. Pediatric Research, 2005. 58(5): p. 981-6.

62. McDermott, D.A., J.C. Fong, and C.T. Basson, Holt-Oram Syndrome, in GeneReviews(R), R.A. Pagon, et al., Editors. 1993: Seattle (WA).

63. Debeer, P., et al., Novel TBX5 mutations in patients with Holt-Oram syndrome. Clinical Orthopaedics and Related Research, 2007. 462: p. 20-6.

64. Stenson, P.D., et al., The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. Human Genetics, 2014. 133(1): p. 1-9.

65. Fan, C., et al., Novel TBX5 mutations and molecular mechanism for Holt-Oram syndrome. Journal of Medical Genetics
2003. 40(3): p. e29.

66. Gruenauer-Kloevekorn, C. and U.G. Froster, Holt-Oram syndrome: a new mutation in the TBX5 gene in two unrelated families. Annales de Génétique, 2003. 46(1): p. 19-23.

67. Boogerd, C.J., et al., Functional analysis of novel TBX5 T-box mutations associated with Holt-Oram syndrome. Cardiovascular Research, 2010. 88(1): p. 130-9.

68. Postma, A.V., et al., A gain-of-function TBX5 mutation is associated with atypical Holt-Oram syndrome and paroxysmal atrial fibrillation. Circulation Research, 2008. 102(11): p. 1433-42.

69. Heinritz, W., et al., Identification of new mutations in the TBX5 gene in patients with Holt-Oram syndrome. Heart, 2005. 91(3): p. 383-4.

70. Borozdin, W., et al., Expanding the spectrum of TBX5 mutations in Holt-Oram syndrome: detection of two intragenic deletions by quantitative real time PCR, and report of eight novel point mutations. Human Mutation, 2006. 27(9): p. 975-6.

71. Vianna, C.B., et al., Holt-Oram syndrome: novel TBX5 mutation and associated anomalous right coronary artery. Cardiology in the Young, 2011. 21(3): p. 351-3.

72. Cross, S.J., et al., The mutation spectrum in Holt-Oram syndrome. Journal of Medical Genetics, 2000. 37(10): p. 785-7.

73. Patel, C., et al., TBX5 intragenic duplication: a family with an atypical Holt-Oram syndrome phenotype. European Journal of Human Genetics, 2012. 20(8): p. 863-9.

74. Brassington, A.M., et al., Expressivity of Holt-Oram syndrome is not predicted by TBX5 genotype. American Journal of Human Genetics, 2003. 73(1): p. 74-85.

75. Yang, J., et al., Three novel TBX5 mutations in Chinese patients with Holt-Oram syndrome. American Journal of Medical Genetics, 2000. 92(4): p. 237-40.

76. Isphording, D., et al., T-box genes and congenital heart/limb malformations. Clinical Genetics, 2004. 66(4): p. 253-64.

77. Krause, A., et al., Tbx5 and Tbx4 transcription factors interact with a new chicken PDZ-LIM protein in limb and heart development. Developmental Biology, 2004. 273(1): p. 106-20.

78. Garg, V., et al., GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. Nature, 2003. 424(6947): p. 443-7.

79. Anderson, R.H., et al., Development of the heart: (2) Septation of the atriums and ventricles. Heart, 2003. 89(8): p. 949-58.

80. Koshiba-Takeuchi, K., et al., Reptilian heart development and the molecular basis of cardiac chamber evolution. Nature, 2009. 461(7260): p. 95-8.

81. Takeuchi, J.K., et al., Tbx5 specifies the left/right ventricles and ventricular septum position during cardiogenesis. Development, 2003. 130(24): p. 5953-64.

82. Hoffmann, A.D., et al., Foxf genes integrate tbx5 and hedgehog pathways in the second heart field for cardiac septation. PLoS Genetics, 2014. 10(10): p. e1004604.

83. Xie, L., et al., Tbx5-hedgehog molecular networks are essential in the second heart field for atrial septation. Developmental Cell, 2012. 23(2): p. 280-91.

84. Franco, D., et al., Left and right ventricular contributions to the formation of the interventricular septum in the mouse heart. Developmental Biology, 2006. 294(2): p. 366-75.

85. Liberatore, C.M., R.D. Searcy-Schrick, and K.E. Yutzey, Ventricular expression of tbx5 inhibits normal heart chamber development. Developmental Biology, 2000. 223(1): p. 169-80.

86. Hoffmann, A.D., et al., sonic hedgehog is required in pulmonary endoderm for atrial septation. Development, 2009. 136(10): p. 1761-70.

87. Goddeeris, M.M., et al., Intracardiac septation requires hedgehog-dependent cellular contributions from outside the heart. Development, 2008. 135(10): p. 1887-95.

88. Mommersteeg, M.T., et al., Two distinct pools of mesenchyme contribute to the development of the atrial septum. Circulation Research, 2006. 99(4): p. 351-3.

89. Snarr, B.S., et al., A spatiotemporal evaluation of the contribution of the dorsal mesenchymal protrusion to cardiac development. Developmental Dynamics, 2007. 236(5): p. 1287-94.

90. Wessels, A., et al., Atrial development in the human heart: an immunohistochemical study with emphasis on the role of mesenchymal tissues. Anatomical Record, 2000. 259(3): p. 288-300.

91. Moskowitz, I.P., et al., A molecular pathway including Id2, Tbx5, and Nkx2-5 required for cardiac conduction system development. Cell, 2007. 129(7): p. 1365-76.

92. Arnolds, D.E., et al., TBX5 drives Scn5a expression to regulate cardiac conduction system function. Journal of Clinical Investigation, 2012. 122(7): p. 2509-18.

93. Thomas, P.S., et al., Elevated expression of Nkx-2.5 in developing myocardial conduction cells. Anatomical Record, 2001. 263(3): p. 307-13.

94. van den Boogaard, M., et al., A common genetic variant within SCN10A modulates cardiac SCN5A expression. Journal of Clinical Investigation, 2014. 124(4): p. 1844-52.

95. Agarwal, P., et al., Tbx5 is essential for forelimb bud initiation following patterning of the limb field in the mouse embryo. Development, 2003. 130(3): p. 623-33.

96. Luna-Zurita, L., et al., Complex Interdependence Regulates Heterotypic Transcription Factor Distribution and Coordinates Cardiogenesis. Cell, 2016. 164(5): p. 999-1014.

97. Mori, A.D., et al., Tbx5-dependent rheostatic control of cardiac gene expression and morphogenesis. Developmental Biology, 2006. 297(2): p. 566-86.

98. Rallis, C., et al., Tbx5 is required for forelimb bud formation and continued outgrowth. Development, 2003. 130(12): p. 2741-51.

99. Hasson, P., J. Del Buono, and M.P. Logan, Tbx5 is dispensable for forelimb outgrowth. Development, 2007. 134(1): p. 85-92.

100. Garrity, D.M., S. Childs, and M.C. Fishman, The heartstrings mutation in zebrafish causes heart/fin Tbx5 deficiency syndrome. Development, 2002. 129(19): p. 4635-45.

101. Ahn, D.G., et al., T-box gene tbx5 is essential for formation of the pectoral limb bud. Nature, 2002. 417(6890): p. 754-8.

102. Albalat, R., M. Baquero, and C. Minguillon, Identification and characterisation of the developmental expression pattern of tbx5b, a novel tbx5 gene in zebrafish. Gene Expression Patterns, 2010. 10(1): p. 24-30.

103. Parrie, L.E., et al., Zebrafish tbx5 paralogs demonstrate independent essential requirements in cardiac and pectoral fin development. Developmental Dynamics, 2013. 242(5): p. 485-502.

104. Puskaric, S., et al., Shox2 mediates Tbx5 activity by regulating Bmp4 in the pacemaker region of the developing heart. Human Molecular Genetics, 2010. 19(23): p. 4625-33.

105. Plageman, T.F., Jr. and K.E. Yutzey, Differential expression and function of Tbx5 and Tbx20 in cardiac development. Journal of Biological Chemistry, 2004. 279(18): p. 19026-34.

106. Chi, N.C., et al., Foxn4 directly regulates tbx2b expression and atrioventricular canal formation. Genes & Development, 2008. 22(6): p. 734-9.

107. Boogerd, C.J. and S.M. Evans, TBX5 and NuRD Divide the Heart. Developmental Cell, 2016. 36(3): p. 242-4.

108. Hiroi, Y., et al., Tbx5 associates with Nkx2-5 and synergistically promotes cardiomyocyte differentiation. Nature Genetics, 2001. 28(3): p. 276-80.

109. Goetz, S.C., D.D. Brown, and F.L. Conlon, TBX5 is required for embryonic cardiac cell cycle progression. Development, 2006. 133(13): p. 2575-84.

110. Delorme, B., et al., Expression pattern of connexin gene products at the early developmental stages of the mouse cardiovascular system. Circulation Research, 1997. 81(3): p. 423-37.

111. Christoffels, V.M., et al., Chamber formation and morphogenesis in the developing mammalian heart. Developmental Biology, 2000. 223(2): p. 266-78.

112. Maitra, M., et al., Interaction of Gata4 and Gata6 with Tbx5 is critical for normal cardiac development. Developmental Biology, 2009. 326(2): p. 368-77.

113. Brown, D.D., et al., Tbx5 and Tbx20 act synergistically to control vertebrate heart morphogenesis. Development, 2005. 132(3): p. 553-63.

114. Ghosh, T.K., et al., Physical interaction between TBX5 and MEF2C is required for early heart development. Molecular Cell Biology, 2009. 29(8): p. 2205-18.

115. Wang, C., et al., Synergistic activation of cardiac genes by myocardin and Tbx5. PLoS One, 2011. 6(8): p. e24242.

116. Ieda, M., et al., Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. Cell, 2010. 142(3): p. 375-86.

117. Qian, L., et al., *In vivo* reprogramming of murine cardiac fibroblasts into induced cardiomyocytes. Nature, 2012. 485(7400): p. 593-8.

118. Takeuchi, J.K., et al., Chromatin remodelling complex dosage modulates transcription factor function in heart development. Nature Communications, 2011. 2: p. 187.

119. Lickert, H., et al., Baf60c is essential for function of BAF chromatin remodelling complexes in heart development. Nature, 2004. 432(7013): p. 107-12.

120. Takeuchi, J.K. and B.G. Bruneau, Directed transdifferentiation of mouse mesoderm to heart tissue by defined factors. Nature, 2009. 459(7247): p. 708-11.

121. Miller, S.A., S.E. Mohn, and A.S. Weinmann, Jmjd3 and UTX play a demethylase-independent role in chromatin remodeling to regulate T-box family member-dependent gene expression. Molecular Cell, 2010. 40(4): p. 594-605.

122. Wamstad, J.A., et al., Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. Cell, 2012. 151(1): p. 206-20.

123. Lewandowski, S.L., et al., Histone deacetylase 3 modulates Tbx5 activity to regulate early cardiogenesis. Human Molecular Genetics, 2014. 23(14): p. 3801-9.

124. Christoffels, V.M., et al., T-box transcription factor Tbx2 represses differentiation and formation of the cardiac chambers. Developmental Dynamics, 2004. 229(4): p. 763-70.

125. Lingbeek, M.E., J.J. Jacobs, and M. van Lohuizen, The T-box repressors TBX2 and TBX3 specifically regulate the tumor suppressor gene p14ARF via a variant T-site in the initiator. Journal of Biological Chemistry, 2002. 277(29): p. 26120-7.

126. Hoogaars, W.M., et al., The transcriptional repressor Tbx3 delineates the developing central conduction system of the heart. Cardiovascular Research, 2004. 62(3): p. 489-99.

127. Hoogaars, W.M., et al., Tbx3 controls the sinoatrial node gene program and imposes pacemaker function on the atria. Genes & Development, 2007. 21(9): p. 1098-112.

128. Stennard, F.A., et al., Murine T-box transcription factor Tbx20 acts as a repressor during heart development, and is essential for adult heart integrity, function and adaptation. Development, 2005. 132(10): p. 2451-62.

129. Sakabe, N.J., et al., Dual transcriptional activator and repressor roles of TBX20 regulate adult cardiac structure and function. Human Molecular Genetics, 2012. 21(10): p. 2194-204.

130. Kaltenbrun, E., et al., A Gro/TLE-NuRD corepressor complex facilitates Tbx20-dependent transcriptional repression. Journal of Proteome Research, 2013. 12(12): p. 5395-409.

131. Hatcher, C.J., et al., TBX5 transcription factor regulates cell proliferation during cardiogenesis. Developmental Biology, 2001. 230(2): p. 177-88.

CHAPTER II: EVOLUTIONARILY CONSERVED *TBX5-WNT2/2B* PATHWAY ORCHESTRATES CARDIOPULMONARY DEVELOPMENT

Abstract

Codevelopment of the lungs and heart underlies key evolutionary innovations in the transition to terrestrial life. Cardiac specializations that support pulmonary circulation, including the atrial septum, are generated by second heart field (SHF) cardiopulmonary progenitors (CPPs). It has been presumed that transcription factors required in the SHF for cardiac septation, e.g., *Tbx5*, directly drive a cardiac morphogenesis gene-regulatory network. Here, we report instead that TBX5 directly drives Wnt ligands to initiate a bidirectional signaling loop between cardiopulmonary mesoderm and the foregut endoderm for endodermal pulmonary specification and, subsequently, atrial septation. We show that *Tbx5* is required for pulmonary specification in mice and amphibians but not for swim bladder development in zebrafish. TBX5 is non–cell-autonomously required for pulmonary endoderm specification by directly driving *Wnt2* and *Wnt2b* expression in cardiopulmonary mesoderm. TBX5 ChIP-sequencing identified cis-regulatory elements at *Wnt2* sufficient for endogenous *Wnt2* expression domains *in vivo* and required for *Wnt2* expression in precardiac mesoderm *in vitro*. *Tbx5* cooperated with Shh signaling to drive *Wnt2b* expression for lung morphogenesis. *Tbx5* haploinsufficiency in mice, a model of Holt–Oram syndrome, caused a quantitative decrement of mesodermal-to-endodermal Wnt signaling and subsequent endodermal-to-mesodermal Shh signaling required for cardiac morphogenesis. Thus, *Tbx5* initiates a mesoderm–endoderm–mesoderm signaling loop in lunged vertebrates that provides a molecular basis for the coevolution of pulmonary and cardiac structures required for terrestrial life.

Significance

In the 20 years since the discovery of the genetic link between the transcription factor *TBX5* and congenital heart defects, few direct targets of *TBX5* in cardiac morphogenesis have been identified. In this work, we demonstrate that TBX5 directly regulates canonical Wnt ligands required for initiation of lung development. Lung endoderm forms a Hedgehog signaling source required for morphogenesis of both the lungs and the cardiac inflow septum. Our work expands the role of TBX5 to include a non–cell-autonomous component for atrial septation. We find the mesoderm–endoderm–mesoderm signaling loop initiated by TBX5 is evolutionarily conserved from amphibians to mammals. This work suggests that the evolutionary origin of lungs may have involved the recruitment of cardiac TBX5.

Introduction

Utilization of atmospheric oxygen revolutionized the ability of vertebrates to adapt to terrestrial life [1]. At the center of this revolution are the lungs, a foregut-derived gas-exchange structure [1, 2]. The derived cardiovascular system, utilizing pulmonary oxygen, must manage blood from both the body and the lungs simultaneously [2]. While most lungfish, amphibians, and reptiles exhibit a three-chambered heart with an atrial septum separating pulmonary and systemic circulation entering the heart [3], the two-sided, four-chambered crocodilian, avian, and mammalian hearts have independently evolved to completely separate pulmonary from systemic circulation [4]. The proper development and placement of the cardiac septa are critical for the efficient handling of blood, and defects in these structures comprise common forms of human congenital heart disease.

Recent work has highlighted the common developmental origin of multiple mesodermal derivatives in both the heart and the lung [5, 6]. This lateral plate mesodermal population has been termed the "second heart field" (SHF) or "cardiopulmonary progenitors" (CPPs). This population originates dorsal to the cardiac inflow tract and ventral to the anterior foregut and generates multiple structures in the heart, e.g., the atrial septum, and in the lungs, e.g., smooth muscle and vascular endothelium [5, 6]. This essential CPP region is labeled by expression of the canonical Wnt signaling ligand *Wnt2*, the Hedgehog (Hh) signaling-responsive transcription factor *Gli1*, and the T-box family transcription factor *Tbx5* [5, 7–9].

Mutations in *TBX5* have been implicated as the primary genetic cause of Holt–Oram syndrome (HOS), a human syndrome that includes cardiac septal defects [10–14]. Previous work has demonstrated that *Tbx5* is required in the posterior SHF (pSHF) for atrial septation [7, 9, 15]. In addition, Sonic hedgehog (Shh) signaling has been implicated in cardiac septation [7–9, 16]. *Shh*, expressed in the pulmonary endoderm (PE), activates GLI-mediated transcription in the CPPs [7, 8]. *Shh* and *Tbx5* genetically interact for cardiac septation, and constitutive activation of Hh signaling in CPPs rescues atrial septal defects caused by reduced *Tbx5* dose [7, 9]. Furthermore, TBX5 and GLI transcription factors directly cooperate at enhancers for genes required for cardiac septation [7, 9]. This has generated a model in which TBX5 and GLI transcription factors directly activate gene expression in the CPPs of the pSHF for cardiac morphogenesis.

CPPs are an important source of signals that induce the pulmonary lineage in the ventral foregut endoderm and contribute directly to the atrial septum and cardiopulmonary vasculature [17–19]. An evolutionarily conserved paracrine signaling cascade involving retinoic acid, Hh, Wnt signaling, and bone morphogenic protein (BMP) regulates the induction of pulmonary

progenitors from amphibians to mammals [5, 17–21]. *Tbx5* has been implicated in lung morphogenesis, both alone and in combination with *Tbx4* [22]. Midgestation conditional deletion of *Tbx5* in mouse embryos caused deficiency in lung-branching morphogenesis, and combined deletion of *Tbx4* and *Tbx5* in allelic combinations caused reduced WNT2 and BMP4 signaling [22].

We report that *Tbx5* is required non–cell-autonomously for the initiation of PE and lung formation. We find that *Tbx5* is required for the initiation of lung development in both mammals and amphibians but not for the swim bladder (SB) in zebrafish. Furthermore, we show that TBX5 directly drives the lung-inducing ligands *Wnt2* and *Wnt2b* in pSHF CPPs. TBX5-driven mesoderm-to-endoderm canonical Wnt signaling is required for the subsequent endoderm-to-mesoderm Shh signaling required for atrial septation. *Tbx5* haploinsufficiency diminishes both mesodermal *Wnt2* and endodermal *Shh* expression, suggesting that atrial septal defects in HOS may be caused in part by diminished Shh signaling rather than solely by a deficiency of a *Tbx5*-driven cell-autonomous SHF gene regulatory network (GRN). *Tbx5* thereby initiates a mesoderm–endoderm–mesoderm signaling loop, providing a molecular basis for the coevolution of pulmonary and cardiac development.

Results

Tbx5-*Dependent Transcriptional Profiling of the CPPs*

To investigate the role of *Tbx5* in CPPs, we performed RNA sequencing (RNA-seq) on microdissected tissue containing the CPPs from $Tbx5^{+/+}$ and $Tbx5^{-/-}$ mouse embryos at E9.5 (Figure II.1 A and B) [7, 23, 24]. Compared with $Tbx5^{+/+}$ CPPs, 5,486 genes were significantly dysregulated in $Tbx5^{-/-}$ CPPs [false-discovery rate (FDR) <0.05]. We restricted our

consideration of genes to those with a magnitude fold change ≥1.5 (*SI Appendix*, Dataset S1). This group contained 1,480 down-regulated genes and 1,588 up-regulated genes in the absence of *Tbx5* (Figure II.1C). The most significantly downregulated genes in *Tbx5*$^{-/-}$ CPPs were transcription factors and signaling factors critical for early lung development. Notably, expression of *Nkx2-1* and the long noncoding RNA *E030019B13Rik* or NANCI, the first markers of PE specification, was extinguished [17, 25–28]. Members of the Wnt and Shh signaling pathways, both required for early lung specification and morphogenesis, were also severely down-regulated (Figure II.1C). In addition, we observed 14 other genes among the down-regulated list that have been reported in the literature to be critical for lung development [6, 29–31]. As an early role for *Tbx5* in lung development has been suggested [22, 32], we validated the significant down-regulation of 15 of 16 "lung development" genes by qRT-PCR in independent samples (Figure II.1D). Together, these data suggested that Tbx5 might occupy a critical position in the GRN for lung induction.

Tbx5-*Dependent Lung Development Is Evolutionarily Conserved*

Utilizing the *Tbx5*$^{-/-}$ mouse embryos, we examined the requirement of *Tbx5* for PE specification. The earliest sign of pulmonary induction, *Nkx2-1* expression, was absent from the foregut endoderm at E9.5 by *in situ* hybridization (ISH) (Figure II.2A). Based on sagittal and coronal sections at E10.5, *Tbx5*$^{-/-}$ mice failed to demonstrate the earliest physical manifestation of lung morphogenesis, the outpouching of the foregut endoderm or lung buds (Figure II.2B). 3D reconstructions highlighting the endoderm further demonstrated the absence of lung bud initiation from the foregut in *Tbx5*$^{-/-}$ embryos (Figure II.2B).

Because of the fundamental role for *Tbx5* in lung development in mice, we asked whether this role is conserved across lunged vertebrates. Previously, an evolutionary link across amniotes has been made between *Tbx5* expression pattern in the heart and ventricular septation for efficient handling of blood [33]. We first examined expression of *Tbx5* by ISH in representative species of amphibians (*Xenopus laevis*), lizards (*Anolis sagrei*), crocodilians (*Alligator mississippiensis*), and birds (*Gallus gallus*). The expression domains of *Tbx5* are conserved across each of these species with expression found in the mesodermal derivatives of the lungs in each (Figure II.2C).

We hypothesized that *Tbx5* may be an evolutionarily ancient driver of lung specification. We examined whether the role of *Tbx5* in lung specification was evolutionarily conserved in amphibians, one of the earliest branching lineages of extant tetrapods [32, 34], using the experimentally tractable, *Xenopus laevis* and *tropicalis*. *Xenopus* embryos expressed *tbx5* in the heart and in the *wnt2b*-expressing lateral plate mesoderm surrounding the *nkx2-1*– and *shh*-expressing PE (*SI Appendix*, Figure II.S1). We examined the requirement of *tbx5* for lung development in *Xenopus tropicalis* by utilizing CRISPR to induce targeted biallelic frameshift mutations in the fifth exon of *tbx5* (termed "*tbx5* FS"), causing predicted truncations of the Tbx5 polypeptide due to premature translation termination (Figure II.2D). We observed a loss of *nkx2-1* in the foregut endoderm of *tbx5* FS embryos compared with controls by ISH of embryos at stage 35, when the lung lineage is being induced (Figure II.2E). We observed an identical loss of *nkx2-1* in *X. laevis* embryos injected with *tbx5* morpholinos (MOs) (Figure II.2F). *tbx5* FS embryos appear phenotypically similar to previously described *tbx5*-MO knockdowns at stage 42 [35, 36], including gross edema and loss of blood in the embryonic heart (Figure II.2G). Furthermore, similar to the *Tbx5*$^{-/-}$ mouse [22], *tbx5* FS embryos lacked lung buds as determined

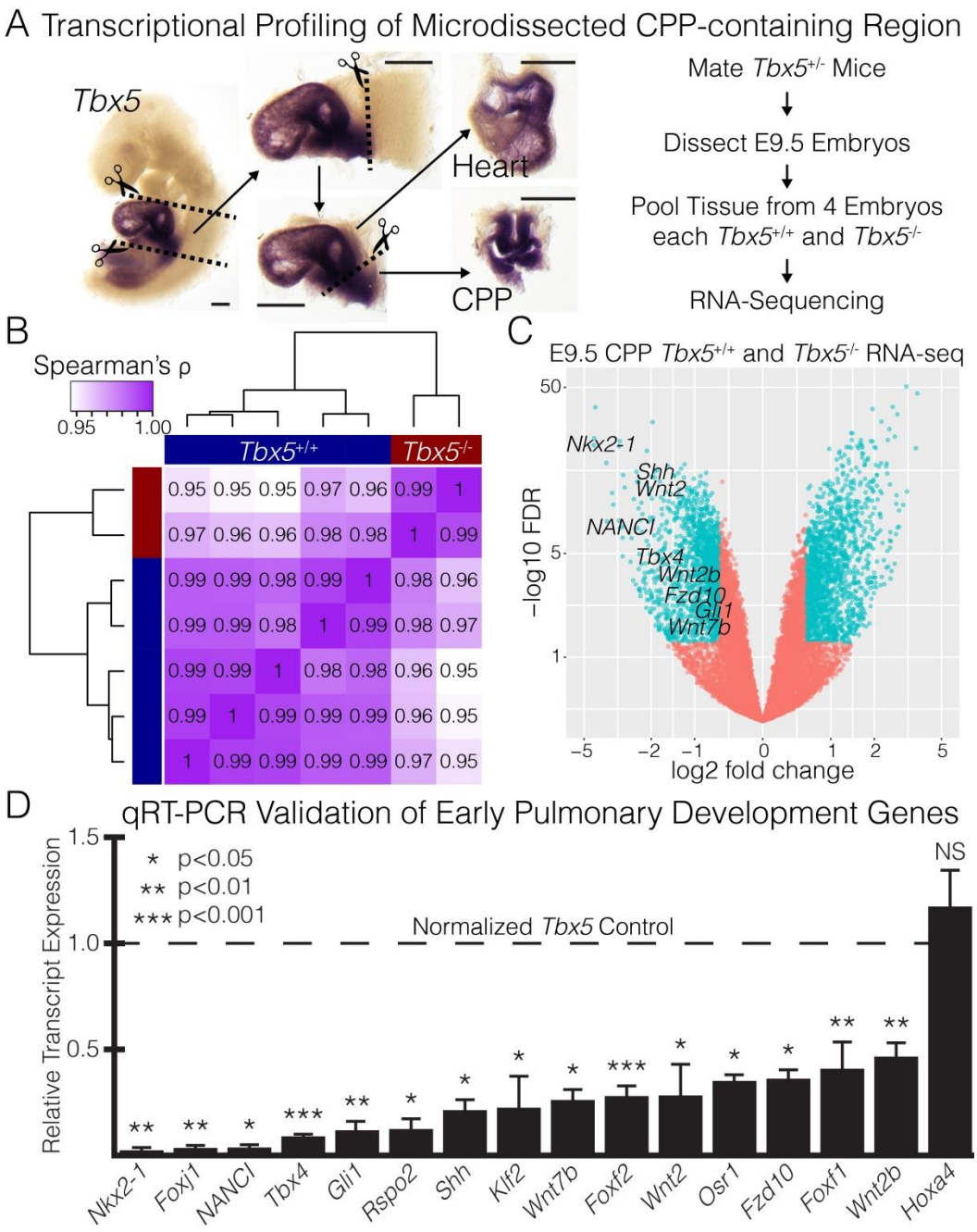


Figure II.1. Transcriptional profiling of microdissected CPPs identifies a critical role for *Tbx5* in pulmonary specification and lung development

(A, Left) Demonstration of microdissection methodology used for embryonic mouse experiments on an E9.5 embryo probed for *Tbx5* RNA by ISH. (Scale bars: 0.25 mm.) (Right) Transcriptional

(Figure II.1 continued) profiling strategy used to measure the *Tbx5*-dependent transcriptome in the CPP-containing tissue by RNA-seq. (B) Spearman's correlation of RNA-seq replicates. (C) Volcano plot of transcriptional profiling results with significantly dysregulated genes (teal) from the comparison of $Tbx5^{+/+}$ and $Tbx5^{-/-}$ CPPs. Early markers of the PE and canonical Wnt and Shh signaling are identified. (D) qRT-PCR validation of 16 early lung-development genes that were significantly dysregulated in the RNA-seq.

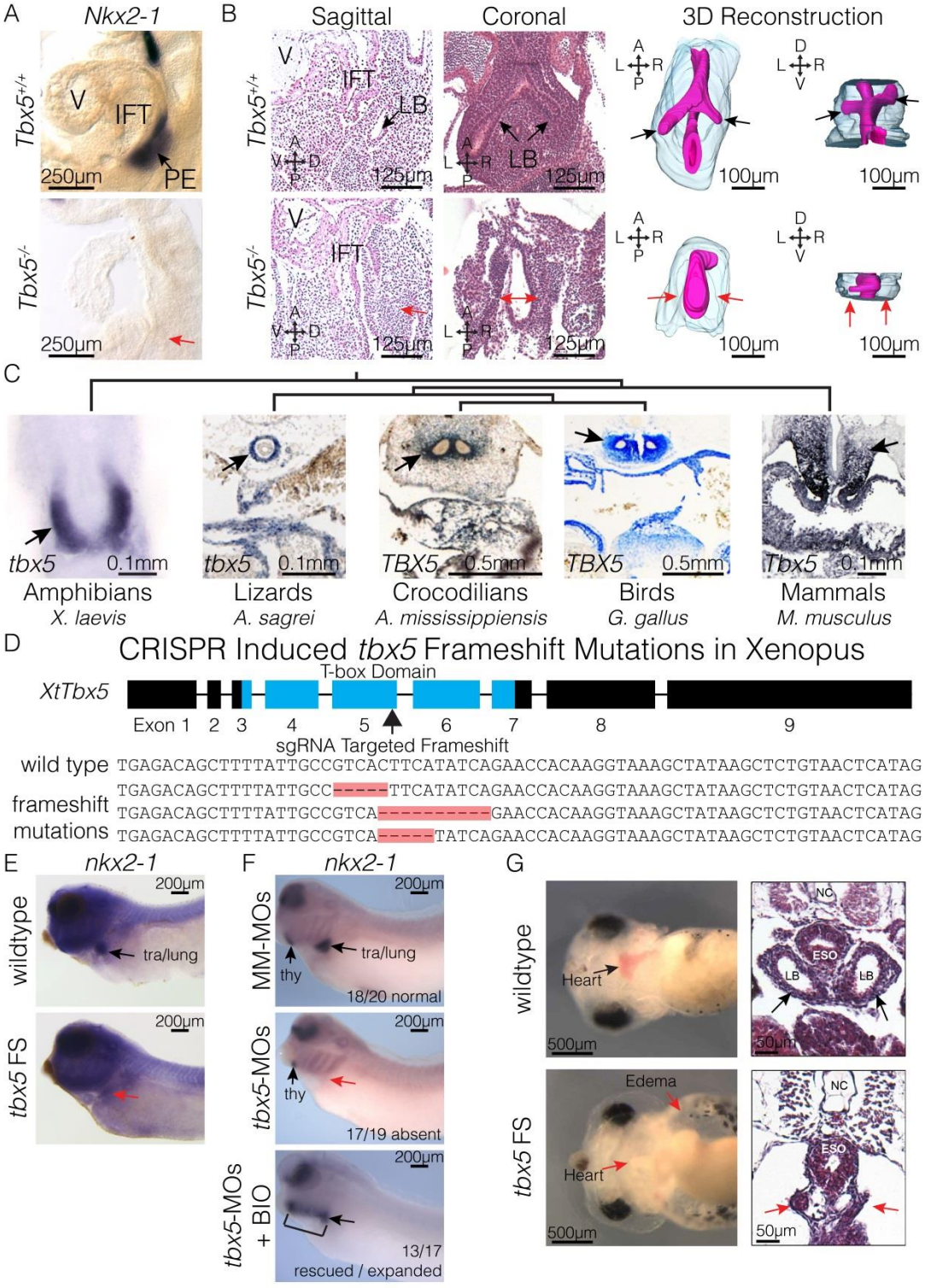


Figure II.2. *Tbx5* is required for lung development in mice and frogs

(Figure II.2 continued) (A) RNA ISH for *Nkx2-1* in E9.5 *Tbx5*$^{+/+}$ and *Tbx5*$^{-/-}$ embryos. The PE, inflow tract (IFT), and cardiac ventricle (V) are labeled. (B) Histology from both sagittal and coronal perspectives (Left) and 3D reconstruction (Right) of E10.5 lungs from *Tbx5*$^{+/+}$ and *Tbx5*$^{-/-}$ embryos. Black and red arrows point to the lung bud (LB) branches or lack of branches off the foregut. (C) ISH stains of *Tbx5* across vertebrate species. Arrows indicate expression in the mesodermal derivatives surrounding the PE. (D, Upper) Strategy for generating biallelic frameshift mutations using sgRNA targeted to the fifth exon of *X. tropicalis* to disrupt the T-box domain. (Lower) Examples of sequences recovered from tbx5 FS mutants. (E and F) RNA ISH of NF35 tadpoles for *nkx2-1* in wild type and tbx5 FS mutants (E) and tadpoles injected with mismatched morpholinos (MM-MOs), *tbx5*-MOs, or *tbx5*-MOs cotreated with BIO (F). (G) Live images (Left) and H&E-stained transverse sections (Right) of NF42 tadpoles, depicting anatomical defects induced by CRISPR-mediated mutation of *tbx5*.

by histologic section at stage 42 (Figure II.2 B and G). To further characterize lung development, expression of *sftpc*, encoding surfactant protein C and a marker of pulmonary epithelium, was examined [6, 37]. *Sftpc* expression was absent from the lung buds of *tbx5*-MO knockdowns at stage 42 but was rescuable by coinjection with a hormone-inducible version of Tbx5 (GR-Tbx5) (*SI Appendix*, Figure II.S2). Together, these data indicated that *Tbx5*-dependent PE specification is evolutionarily ancient and is conserved from amphibians to mammals.

The SB of ray-finned fish is an endoderm-derived out-pocket proposed to be a lung homolog and whose development shares many genetic pathways with lung development, including Wnt- and Hh-dependent signaling and transcriptomes [38–44]. The role of *Tbx5* in heart and limb development of ray-finned fish is conserved with tetrapods [45–47]. To address whether *Tbx5* is required for SB formation, zebrafish homozygous for the *tbx5a* mutant *heartstrings* (*hst*) allele [45] were examined at 96 hours postfertilization (hpf). Similar to clutchmate controls, homozygous *hst* mutants show SB formation (*SI Appendix*, Figure. II.S3). Zebrafish have two copies of the *Tbx5* gene, *tbx5a* and *tbx5b* [48]. To address potential redundancy, we utilized published MOs designed to target *tbx5a*, *tbx5b*, or *tbx5a* and *tbx5b* [47, 48]. In all conditions, early expression of *shha* in the SB bud at 72 hpf [39] and SB formation at 96 hpf was observed (*SI Appendix*, Figure II.S3). Together, our data suggest that, while *Tbx5* is required for the initiation and formation of the lungs, *tbx5a*/*b* is not required for the formation of the SB.

Non–Cell-Autonomous Requirement of Tbx5 *for PE Specification*

Previous work in mice has demonstrated that *Nkx2-1* expression in the foregut endoderm is regulated through canonical Wnt signaling, specifically *Wnt2* and *Wnt2b*. Furthermore,

temporal deletion of *Tbx5* from cultured embryos demonstrated decreased *Wnt2* and *Wnt2b* expression [22], making the Wnt ligands excellent candidates for direct TBX5 regulation. *Wnt2* and *Wnt2b* are coexpressed with *Tbx5* in the SHF mesoderm and are significantly down-regulated in $Tbx5^{-/-}$ CPPs (Figure II.1 C and D) [5, 17, 20]. We examined the epistatic requirement for *tbx5* and canonical Wnt signaling in *Xenopus* lung specification. We asked if lung specification in *tbx5*-MO embryos could be rescued by activating the canonical Wnt pathway, using treatment with a glycogen synthase kinase 3 (GSK-3) inhibitor, 6-bromoindirubin-30-oxime (BIO), which stabilizes β-catenin. We found that BIO treatment rescued and expanded *nkx2-1* expression in *tbx5*-MO embryos (Figure II.2E), suggesting that canonical Wnt signaling is downstream of *tbx5* in the lateral plate mesoderm of amphibians. Furthermore, we observed that expression of *Wnt2* and *Wnt2b* was extinguished in $Tbx5^{-/-}$ mouse embryos at E9.5 (Figure II.3A) and found by ISH that *wnt2b* was similarly lost from stage-35 *tbx5* FS *X. tropicalis* and from *tbx5*-MO–injected *X. laevis* (Figure II.3 B and C). These observations were consistent with a requirement for *Tbx5* upstream of canonical Wnt signaling for pulmonary specification.

Tbx5 *Is Required for Pulmonary* Shh *Signaling*

Hh signaling from the PE is required for both cardiac and lung morphogenesis [7, 9, 19, 49]. Specifically, *Shh* is expressed in the PE and is required for atrial septation and lung morphogenesis postinduction [8, 50, 51]. We predicted that the requirement of *Tbx5* for PE specification would also reflect a requirement for pulmonary Shh signaling. We observed by RNA-seq that *Shh* is dramatically downregulated in $Tbx5^{-/-}$ embryos, and we observed by ISH that *Shh* is specifically lost from the foregut/PE at E9.5 (Figures II.1A and II.3A). This suggests

that the epistatic relationship between *Tbx5* and Shh signaling [9] is an indirect feature of the requirement of *Tbx5* for pulmonary lung induction.

TBX5 haploinsufficiency in humans results in the congenital HOS, displaying radial forelimb and congenital heart defects, most commonly atrial septal defects [10–12]. *Shh* expression in the PE is required for morphologic development of the atrial septum [8, 16], and we have demonstrated a genetic interaction between *Tbx5* and *Shh* [7, 9]. We therefore hypothesized that a quantitative decrement in *Tbx5* would result in diminished Wnt and Shh signaling, contributing to the *Tbx5*-haploinsufficient phenotype. We examined the gene-expression level of the canonical Wnts, PE specification, and Hh signaling in mouse embryos with *Tbx5* haploinsufficiency (Figure II.3D). In the CPPs of $Tbx5^{+/-}$ embryos, we observed a significant down-regulation of *Wnt2* (0.67 ± 0.05 SEM, P = 0.0126) but not of *Wnt2b* or *Nkx2-1* (0.69 ± 0.15 SEM, P = 0.1707 and 0.80 ± 0.15 SEM, P = 0.5420, respectively). However, a significant down-regulation of *Shh* (0.46 ± 0.04 SEM, P = 3.693E-05) and the canonical Hh targets *Gli1* and *Hhip* (0.81 ± 0.04 SEM, P = 0.0347 and 0.63 ± 0.05 SEM, P = 0.0028, respectively) was observed in $Tbx5^{+/-}$ embryos compared with controls. Thus, *Tbx5* haploinsufficiency caused a decrement of both *Shh* expression in the PE and Shh signaling reception in CPPs.

A Mesoderm–Endoderm–Mesoderm Signaling Loop for Cardiopulmonary Development

Tbx5 and Shh signaling coordinately control gene expression in the CPPs for cardiac development [7, 9, 49]. We asked if *Tbx5* and Shh signaling interact to regulate lung development. Previously we showed that Shh reception in CPPs promotes *wnt2b* expression during lung induction in *Xenopus* [19]. We observed similar results in mice: *Wnt2b*, but not

Wnt2, in CPPs was *Shh* dependent at E9.5 (Figure II.3E). We further confirmed *Wnt2* expression in the embryonic mesoderm of the $Smo^{-/-}$ germline mutant, which ablates all Hh signaling independent of ligand (*SI Appendix*, Figure II.S4). These data suggest that *Wnt2* expression is upstream or independent of *Shh*, while *Wnt2b* is downstream of *Shh*.

Utilizing a suite of biochemical reagents and the *Xenopus* model (Figure II.4A), we investigated the interaction between Tbx5 and Shh signaling for *wnt2b* expression. RNA encoding a fusion protein between Tbx5 and the hormone-inducible region of the glucocorticoid receptor (GR-Tbx5), affording dexamethasone (DEX)-dependent regulation of nuclear import, was injected into the anterior mesendoderm (AME), which has active Hh signaling and gives rise to the foregut [52–55], or into the posterior mesendoderm (PME), which does not have active Hh signaling (Figure II.4 A and B). AME and PME tissue was explanted postgastrulation, DEX treated, and examined after 6 h (Figure II.4 C and D and *SI Appendix*, Figure II.S5). We found that GR-Tbx5 was sufficient to activate *wnt2b* in the Hh-positive AME but not in the Hh-negative PME (Figure II.4 C and D). We examined whether *wnt2b* activation by GR-Tbx5 was direct by coadministering DEX with cycloheximide (CHX) to block translation (Figure II.4 A–C). As with DEX administration alone, DEX/CHX coadministration induced wnt2b expression in the AME but not in the PME (Figure II.4B), suggesting that GR-Tbx5 directly activates *wnt2b* expression in Hh-positive tissue. To validate the requirement for Hh signaling, we coadministered DEX with the Hh antagonist cyclopamine (*SI Appendix*, Figure II.S5) [56]. Coadministration of DEX and cyclopamine significantly blunted the activation of *wnt2b* in AME tissue as compared with DEX alone (97.0-fold decrement, P = 5.48E-3) (Figure II.4D). Last, we examined whether Hh signaling in the AME was unique or whether treatment of PME tissue with the Hh agonist purmorphamine [57] was sufficient for the coinduction of *wnt2b* expression.

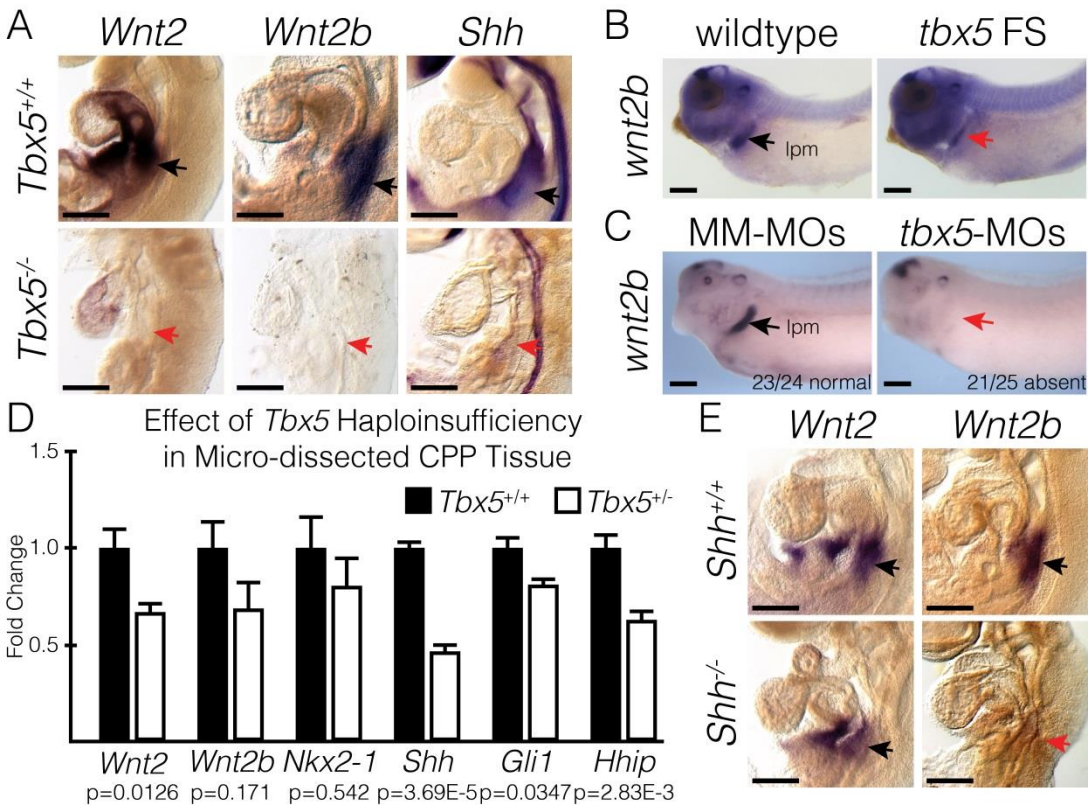


Figure II.3. *Tbx5* is required for *Wnt2*/*2b* and *Shh* expression

(A) RNA ISH for *Wnt2*, *Wnt2b*, and *Shh* in E9.5 $Tbx5^{+/+}$ and $Tbx5^{-/-}$ mouse embryos. Black and red arrows point to positive and negative staining, respectively, in the lung-forming region. (B and C) RNA ISH for *wnt2b* performed in wild-type or *tbx5* FS NF35 tadpoles (B) and in NF35 tadpoles injected with mismatched morpholinos (MM-MOs) or *tbx5*-MOs (C). The stained region corresponds with the lateral plate mesoderm (lpm). (D) qRT-PCR of microdissected CPP tissue from E9.5 *Tbx5*+/+ or *Tbx5*+/− mouse embryos. (E) RNA ISH for *Wnt2* and *Wnt2b* in E9.5 *Shh*+/+ and *Shh*−/− mouse embryos. (Scale bars: 250 μm.)

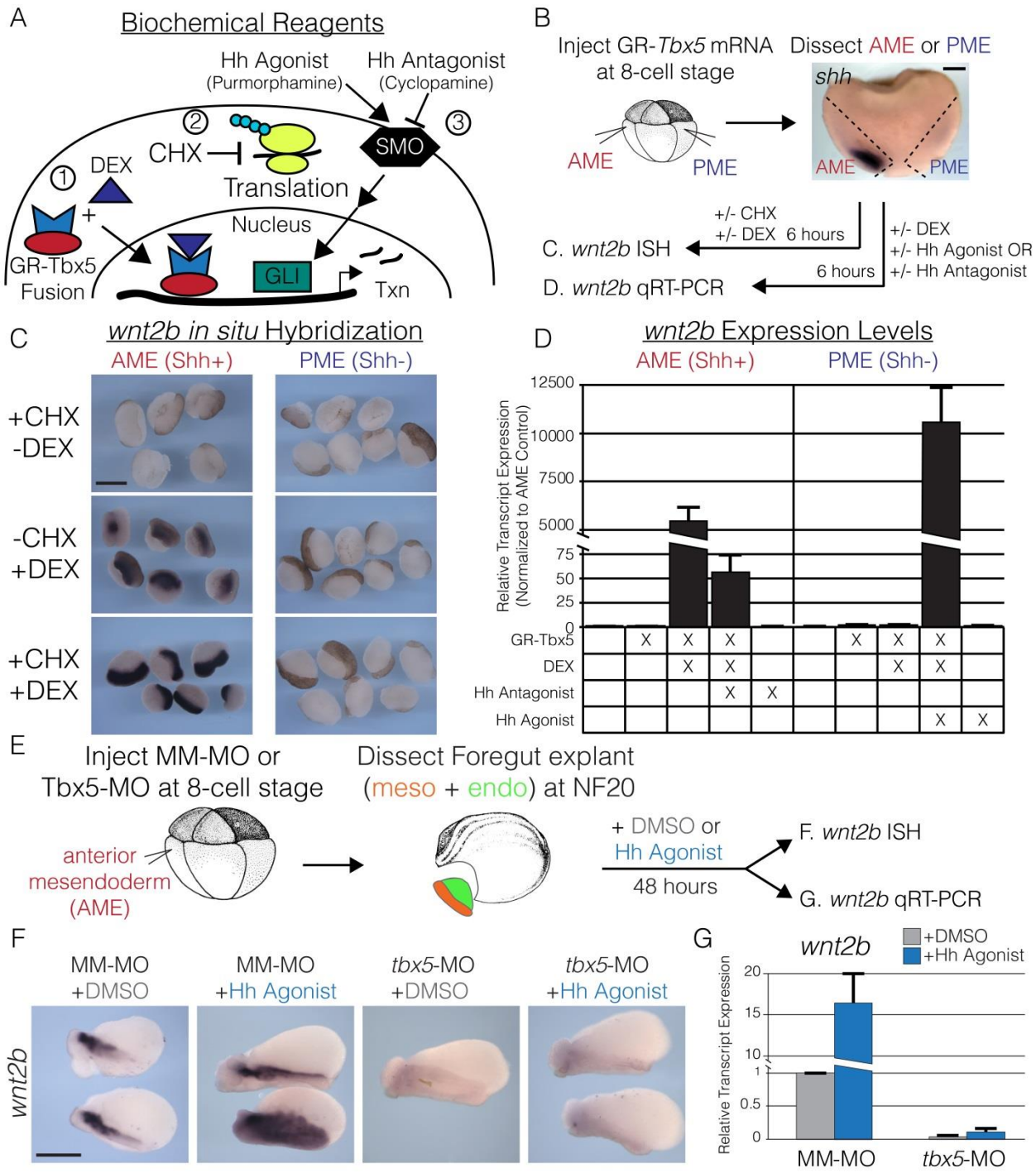


Figure II.4. *tbx5* directly regulates *wnt2b* expression for lung development in the presence of Hh signaling

(Figure II.4 continued) (A) To study the interaction of Tbx5 and Hh signaling, we utilized (1) a DEX-inducible GR-Tbx5 fusion protein; (2) CHX to inhibit protein synthesis; and (3) pharmacological agonists (purmorphamine) and antagonists (cyclopamine) of Smoothened (SMO) to activate or repress Hh signaling. (B) Strategy used in C and D for examining the regulation of *wnt2b* by Tbx5 in the presence or absence of Shh signaling in *Xenopus*. The AME (red) corresponds to *shh*-expressing tissue, and the PME (blue) corresponds to *shh*-negative tissue. (C) RNA ISH for *wnt2b* in AME or PME explants treated with CHX, DEX, or both. Note that the ISH signal is black; the brown color is pigment. (D) qRT-PCR of *wnt2b* in AME or PME explants, with or without GR-Tbx5, and treated with combinations of DEX, Hh agonist, and Hh antagonist. (E) Strategy used in F and G to examine the regulation of *wnt2b* by Hh signaling in the presence or absence of *tbx5*. AME explants were treated with DMSO (gray) or the Hh agonist purmorphamine (blue) for 48 h. (F) RNA ISH for wnt2b on DMSO- or Hh agonist-treated explants from embryos injected with mismatched morpholinos (MM-MO) or *tbx5*-MO. (G) qRT-PCR of *wnt2b* normalized to DMSO-treated MM-MO–injected explants. (Scale bars: B, 200 μm; C and E, 400 μm.)

PME tissue treated with purmorphamine alone significantly activated gli1 (16.7-fold activation, P = 0.015) but not *wnt2b* (Figure II.4D and *SI Appendix*, Figure II.S5). However, injection of GR-Tbx5 followed by coadministration of DEX and purmorphamine significantly activated *wnt2b* expression in the PME (P = 9.29E-3) (Figure II.4D). Together, these data suggested that Tbx5 directly activates *wnt2b* gene expression in the presence of active Hh signaling.

We next evaluated whether Tbx5 is required for the Hh pathway to activate expression of *wnt2b* and other markers of pulmonary development (Figure II.4 E–G and *SI Appendix*, Figure II.S5). We activated Hh signaling using the Hh agonist purmorphamine in control or *tbx5*-MO–injected AME foregut explants (Figure II.4E). Purmorphamine treatment of control embryos expanded the endogenous *wnt2b* expression domain, consistent with activation by Hh signaling (Figure II.4 F and G). However, purmorphamine was unable to promote *wnt2b* in explants from *tbx5*-MO embryos, which, similar to vehicle-treated *tbx5*-MO explants (Figure II.4 F and G), had little detectable *wnt2b*. Similar to $Tbx5^{-/-}$ mice (Figures II.1 and II.3), *tbx5*-MO explants displayed decreased expression of *nkx2-1* (P = 1.13E-9), *shh* (P = 6.81E-3), and *dhh* (P = 1.11E-5) in the foregut endoderm and decreased expression of *wnt2b* (P = 1.20E-3) and *gli1* (P = 1.09E-3) in the mesoderm (Figures II.4 F and G and *SI Appendix*, Figure II.S5). These explant data demonstrate that *tbx5* and Hh signaling are corequired for *wnt2b* expression and further demonstrate that *Tbx5* expression is required for *Shh* and *dhh* expression in the PE. Overall, these findings suggest a hierarchical series of signaling loops: TBX5 drives canonical Wnt mesoderm-to-endoderm signaling for pulmonary induction, and Shh signaling from the PE to mesoderm collaborates with mesodermal TBX5 to drive ongoing WNT2B mesoderm-to-endoderm signaling for pulmonary morphogenesis.

Identification of TBX5-Dependent Wnt2 *Enhancers*

To identify direct targets of TBX5 in coordinating cardiopulmonary specification, we performed TBX5 ChIP-seq on microdissected hearts including the *Wnt2*-expressing inflow tract from E9.5 mouse embryos (Figure II.1A). We identified 3,883 TBX5-bound regions at E9.5 (Figure II.5A and *SI Appendix*, Dataset S2). These locations segregated into 823 promoter-proximal (TBX5 summit ≤2 kbp from an annotated transcription start site) and 3,060 promoter-distal sites. To define TBX5 binding in a genomic context, we identified active promoters and cis-regulatory regions for H3K4me3 and H3K4me1 by ChIP-seq and in microdissected E9.5 heart and CPP tissue by assay for transposase-accessible chromatin sequencing (ATAC-seq) (Figures II.1A and II.5A) [58]. ATAC-seq, genome-wide and at TBX5-bound regions, showed similar signal (Pearson correlation coefficient = 0.96 and 0.93), H3K4me1 (Pearson correlation coefficient = 0.89 and 0.76), and H3K4me3 (Pearson correlation coefficient = 0.92 and 0.93) in both CPPs and the heart (*SI Appendix*, Figure II.S6). To identify direct TBX5 targets in the CPP tissue, we overlapped the 1,480 *Tbx5*-dependent genes (Figure II.1C) with the 3,880 ChIP sites (annotated to the nearest gene). This conservative approach identified 162 genes associated with 220 bound sites, including *Wnt2* and *Wnt2b* (Figure II.5 B and C and *SI Appendix*, Figure II.S7 and Dataset S3). This observation was consistent with the hypothesis that TBX5 directly regulates *Wnt2* and *Wnt2b* transcription [22, 32].

While both *Wnt2* and *Wnt2b* are *Tbx5* dependent and redundant for lung development [17, 22], only *Wnt2* is required for both lung and cardiac morphogenesis in mammals [18]. Therefore, we attempted to identify the TBX5-dependent cis-regulatory elements that control *Wnt2* expression. We compared our TBX5 ChIP-seq results with previously published TBX5 ChIP-seq from two *in vitro* systems [59, 60]. Using this approach, we identified a cluster of

TBX5-bound sites adjacent to the 3′ end of the neighboring gene, *St7*, which demonstrated the hallmarks of putative regulatory elements including chromatin accessibility and H3K4me1 signal (Figure II.5C). We cloned the regions corresponding to the ChIP-seq signal (mm10 chr6:17938154–17940081, chr6:17941997–17942724, and chr6:17952290–17953703) and named the putative regulatory elements "Wnt2 enhancer 1," "Wnt2 enhancer 2," and "Wnt2 enhancer 3" (W2E1–3), respectively (Figure II.5C). We performed ChIP-qPCR in IMR90 human lung fibroblast cells to validate TBX5 localization at these candidate enhancers (Figure II.5D). We observed a significant enrichment of TBX5 at W2E1 and W2E2 over IgG control (12.14 ± 3.81 SD, P = 0.03 and 10.95 ± 0.85 SD, P = 3.5E-5, respectively), while W2E3 did not show enrichment (0.58 ± 0.36 SD, P = 0.44).

We examined the enhancer activity and *Tbx5* dependence of W2E1–3 *in vitro* by luciferase reporter assay using HEK293T cells and exogenous expression of *Tbx5*, as previously described [7, 8, 61]. *Tbx5* expression activated W2E1 and W2E3 (3.44 ± 1.07 SEM, P = 0.0309 and 5.29 ± 0.93 SEM, P = 0.0099, respectively) but not W2E2 (1.27 ± 0.33 SEM, P = 0.4532) compared with a control vector (Figure II.5E). As W2E1 was both TBX5 bound and responsive to *Tbx5* expression, we examined the dependence of W2E1 activity on the presence of the canonical T-box motif AGGTG (*SI Appendix*, Figure II.S8) [59, 60, 62]. Mutation of the minimal canonical T-box motifs in W2E1 resulted in a 3.997-fold decrease compared with wild-type W2E1 (P = 0.0252), whereas mutation of T-box motifs within the control vector had no effect (P = 0.5237).

We examined the sufficiency of W2E1–3 for driving cardiac and SHF gene expression *in vivo*. Each enhancer was cloned upstream of the *Hsp68* minimal promoter driving *lacZ* expression and was utilized for the generation of transient transgenic mouse embryos, as

previously described [7, 63, 64]. W2E1 and W2E2 each drove robust activation of *lacZ* within the CPP and inflow tract domains of *Wnt2* expression with W2E1 driving robust activation of *lacZ* within many domains of *Wnt2* at E9.5 (Figure II.5F). W2E3, in contrast, activated *lacZ* in non–*Wnt2*-, non–*Tbx5*-expressing tissues. Taken together, our data suggested that W2E1 and W2E2 represent TBX5-responsive cis-regulatory elements for early cardiopulmonary *Wnt2* expression.

Requirement of TBX5-Dependent Regulation of Wnt2

To investigate the direct requirement of TBX5-driven enhancers for pulmonary mesoderm gene expression, we generated a mouse embryonic stem cell (mESC) line with doxycycline (DOX)-inducible *Tbx5* expression using the A2Lox.cre mESC line [65]. This line (*Tbx5*OE-mESC) was differentiated along a sequence of ES cells to mesoderm to lateral plate mesoderm to cardiac progenitor as previously described [66]. We observed a linear dose-response of *Tbx5* in cardiac progenitors (0 ng/μL to 500 ng/μL; 0.07x + 0.60, P = 3.29E-4) after 24 h (Figure II.6A). We observed a significant relationship between DOX dose and expression of *Wnt2* (P = 1.02E-2), *Wnt2b* (P = 1.55E-3), and *Tbx4* (P = 2.03E-2), another marker of pulmonary mesoderm. This observation suggested that pulmonary mesodermal markers are directly responsive to *Tbx5* expression levels in CPPs *in vitro*.

We examined the requirement of the TBX5-bound cis-regulatory elements W2E1 and W2E2 for *Tbx5*-dependent *Wnt2* expression. Specifically, we utilized CRISPR/cas9 to generate a 4.6-kbp deletion of W2E1 and W2E2 without disrupting the last exon of *St7* or its predicted splice acceptor from the mESC line overexpressing *Tbx5* (hereafter, the "*Tbx5*OE-mESC line")

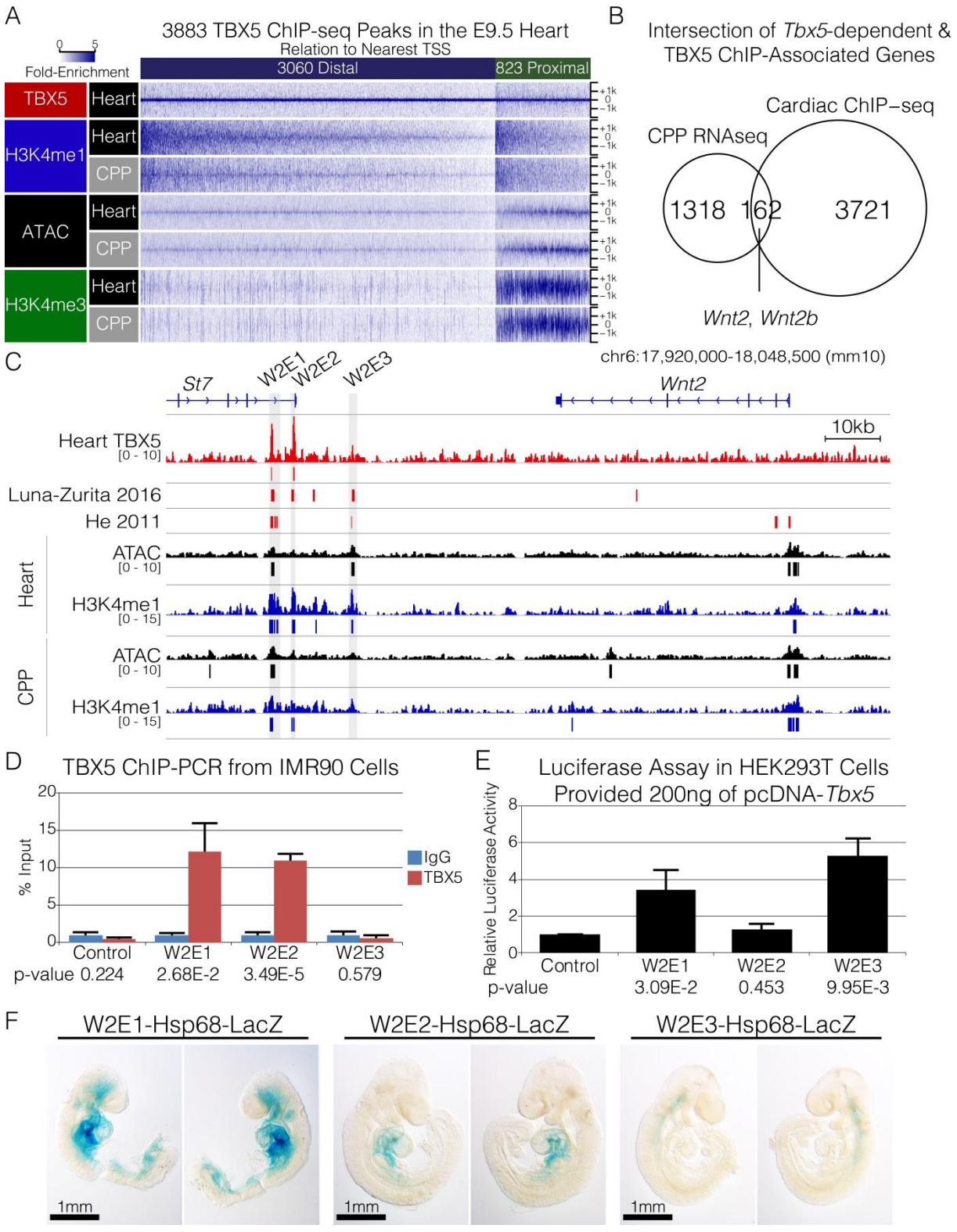


Figure II.5. Identification of TBX5-bound cis-regulatory elements for *Wnt2*

(Figure II.5 continued) (A, Upper) We identified 3,883 peaks by TBX5 ChIP-seq in the E9.5 heart that correspond to 3,060 distal and 823 proximal sites. (Lower) Heatmaps of fold enrichment plotted for TBX5, H3K4me1, and H3K4me3 by ChIP-seq and ATAC-seq from the heart and CPP microdissections at each of the 3,883 summits ± 2,000 bp. (B) Overlap of the 1,318 down-regulated genes in the CPP of *Tbx5*$^{-/-}$ embryos by RNA-seq and the 3,883 genes nearest to TBX5 ChIP-seq peaks. The 162 genes in the intersection include *Wnt2* and *Wnt2b*. (C) Genome browser view of *Wnt2* and *St7* (mm10 chr6:17,920,000–18,048,500) with TBX5 ChIP-seq (both from A and published in refs. 59 and 60), H3K4me1 ChIP-seq, and ATAC-seq in the heart and pSHF. Tracks depict fold-enriched signal, and bars below represent significant peak calls. Cloned enhancers W2E1, -2, and -3 are shaded in gray. (D) ChIP-PCR for TBX5 at W2E1, -2, and -3 in the human IMR90 lung fibroblast cell line. Significance is calculated relative to IgG control. (E) Luciferase assay examining activation of W2E1, -2, and -3 in HEK293T cells provided a vector containing *Tbx5* relative to a control vector. (F) Transgenic embryos were generated using an *Hsp68-LacZ* reporter construct upstream of W2E1, -2, or -3 and were stained at E9.5.

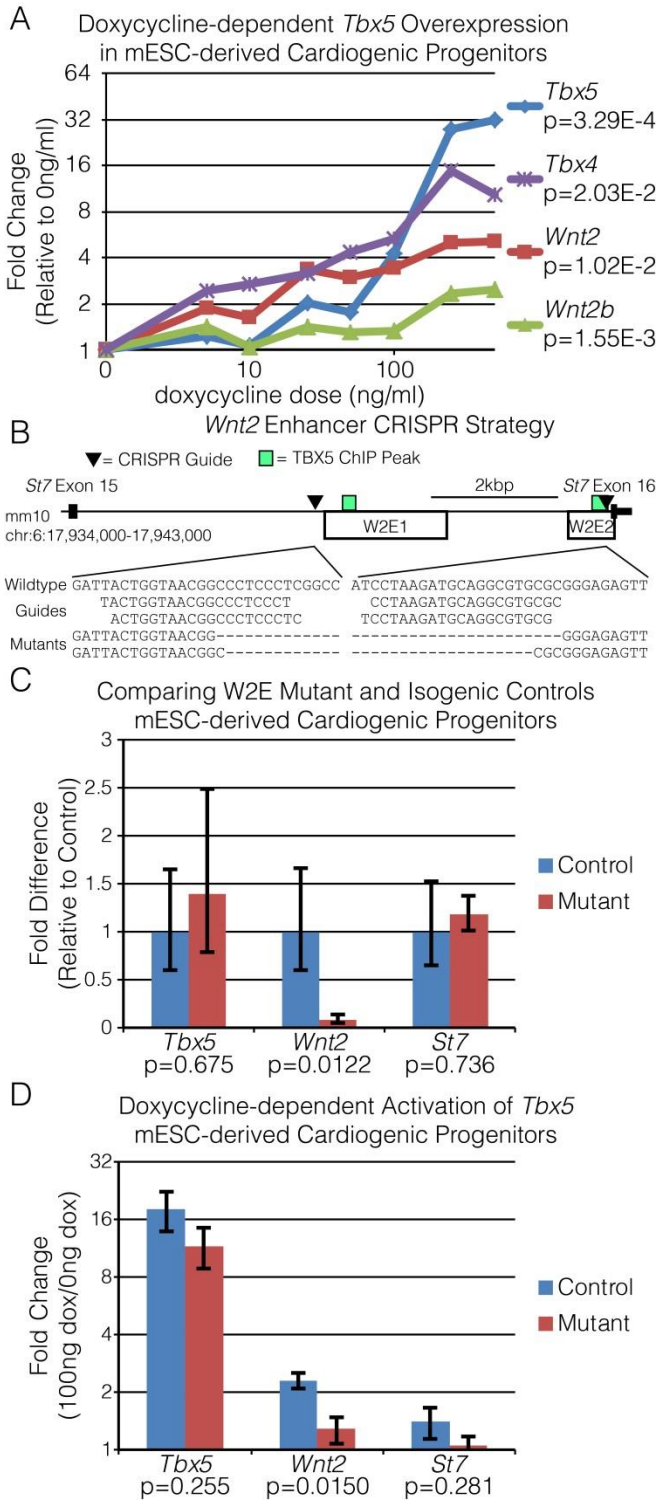


Figure II.6. *Cis*-regulatory elements are required for *Wnt2* expression

(Figure II.6 continued) (A) Dose-dependent gene-expression changes in mESC-derived cardiogenic progenitors harboring a DOX-dependent *Tbx5* construct (*Tbx5*OE-mESC) measured by qRTPCR. Cells were treated with DOX for 24 h before analysis. (B, Upper) sgRNAs were designed to induce a 4.5-kb deletion within the last intron of *St7*, removing a majority of W2E1 and W2E2 in the *Tbx5*OE-mESC via CRISPR-Cas9. This design maintains the last exon and the predicted splice branch for *St7* while removing the TBX5-bound sites in W2E1 and W2E2. (Lower) Amplification and sequencing across the target site demonstrate successful deletion. (C) Expression levels of *Tbx5*, *Wnt2*, and *St7* were compared in W2E mutants and isogenic controls following differentiation to cardiogenic progenitors by qRTPCR. (D) Changes in *Tbx5*, *Wnt2*, and *St7* expression following 24 h of 100 ng/mL DOX relative 0 ng/mL DOX in the W2E mutant and isogenic controls differentiated to cardiogenic precursors.

(Figure II.6B). Two homozygous deletion clones (W2E mutants) and two control clones (W2E controls) were generated and evaluated. Following differentiation of clones to cardiac progenitors, the W2E mutants demonstrated a significant reduction in *Wnt2* gene expression compared with W2E controls (91.8% reduction, P = 0.0122), while there was no significant difference in *Tbx5* or *St7* (Figure II.6C). To examine the requirement of W2E1/W2E2 for *Tbx5*-dependent activation of *Wnt2*, we induced *Tbx5* overexpression and evaluated the response of *Tbx5*, *Wnt2*, and *St7* expression. We observed that the W2E mutants had significantly reduced *Wnt2* expression in response to *Tbx5* overexpression compared with the W2E control lines (2.31-fold versus 1.28-fold activation, P = 0.0150); no significant differences between the mutant and control lines were observed for *Tbx5* or *St7* (Figure II.6D). Taken together, these results demonstrate that W2E1 and W2E2 are required for *Wnt2* expression and are necessary for *Tbx5*-responsive *Wnt2* expression in mESC-derived CPPs. These results demonstrate direct molecular control of *Wnt2* by TBX5 in an *in vitro* model of early cardiopulmonary development.

Discussion

TBX5 has been genetically implicated in human cardiac septal defects for over 20 y. Based on its strong expression in the heart, *TBX5* was assumed to directly drive a cardiac GRN for cardiac septation. Recently, work by the I.P.M. laboratory has determined that the role of *TBX5* resides in the SHF (Figure II.7). We assumed that probing pSHF CPPs for *Tbx5*-dependent target genes would uncover a direct cardiac progenitor GRN for cardiac morphogenesis. Instead, we observed a primary role for *Tbx5* in the initiation of lung development and, secondarily, the establishment of PE-to-mesoderm signaling for cardiac septation.

We report that *Tbx5* is required for the initiation of lung development through canonical Wnt signaling (Figure II.7). We observed that *Tbx5* directly regulates the transcription of both *Wnt2* and *Wnt2b*, wingless-family signaling molecules redundantly required for the earliest aspects of pulmonary development [17]. Specifically, we identified *cis*-regulatory elements for *Wnt2* that are required for TBX5-responsive transcription and that drive transcription in the CPPs and inflow tract of the heart. A role for canonical Wnt signaling in inflow tract development is conserved between *Drosophila* and mammals [67]. *Drosophila Wingless* (*wg*) is required for the formation of the *Drosophila* cardiac inflow tract [67], suggesting that the preexisting role of canonical Wnt signaling in inflow tract development may have been coopted for lung development and inflow septation later in vertebrates and early tetrapods. Although the early requirement of *Tbx5* for heart and limb development has been well documented across vertebrate species [13–15, 45–47, 68–70], the role of *Tbx5* in lung development has not been examined outside of mammals. Overall, our work suggests a fundamental role for *Tbx5* in tetrapod lung development and the possibility that the evolutionary origin of lungs may have involved the recruitment of TBX5 from an ancestral cardiac expression domain.

Although a recent model suggested that the lungs and SB are evolutionarily derived from a common structure, we find that *tbx5* is not required for the development of both [43, 44]. Previous work in zebrafish demonstrates that depletion of *wnt2* and *wnt2bb* causes SB agenesis, similar to their requirement in lung specification [17, 38]. However, we find that although *Tbx5* is required for lung formation, *tbx5a*/*b* is not required for SB formation in the zebrafish. These observations suggest that *tbx5a*/*b*-independent regulation of Wnt signaling is required for the initiation of SB development. One question worth future investigation is whether the Tbx5-positive lateral plate mesoderm gives rise to SB components or whether the dorsally derived SB

forms from a mesodermal contribution distinct from the ventrally derived lungs. Additionally, zebrafish are part of the derived teleost fish, and further characterization across ray-finned fish is required. We note that the Senegal bichir (*Polypterus senegalus*), a member of the early-diverging Actinopterygii, has a ventral-sided lung structure for air breathing and was reported to express both *tbx5* and *tbx4* in the early lung structure [43]. Additional work is required to resolve the evolutionary relationship between the SB and lungs.

Integrating our observations that *Tbx5* is required for pulmonary specification with previous work demonstrating a role for *Tbx5* in lung morphogenesis [22] suggests sequential roles for Wnt signaling during lung development: an early requirement for initiation and a later requirement for branching morphogenesis. In contrast to the complete loss of lung development in the *Tbx5* germline-null mouse, conditional removal of *Tbx5* at E8.5 caused malformation of lung bud branching and disruption of canonical Wnt signaling in explant cultures [22]. Consistent with a dual-role hypothesis, a partial decrement of canonical Wnt signaling allows lung initiation but causes defects in lung-branching morphogenesis [17, 18], while a complete failure of lung initiation has been observed only by homozygous removal of both *Wnt2* and *Wnt2b* [17]. We conclude that *Tbx5* initiates a multistep mesoderm–endoderm–mesoderm signaling loop (Figure II.7). TBX5 directly drives WNT2 and WNT2B mesoderm-to-endoderm signaling for pulmonary induction. Secondarily, PE-to-mesoderm Shh signaling collaborates with mesodermal TBX5 for ongoing WNT2B mesoderm-to-endoderm signaling and later pulmonary morphogenesis (Figure II.7).

This study and previously published work suggest that *TBX5* mutations may be associated with lung defects. Although rare, lung defects have been described in patients with HOS. Two cases of structural lung disease have been associated with “atypical” HOS: one case

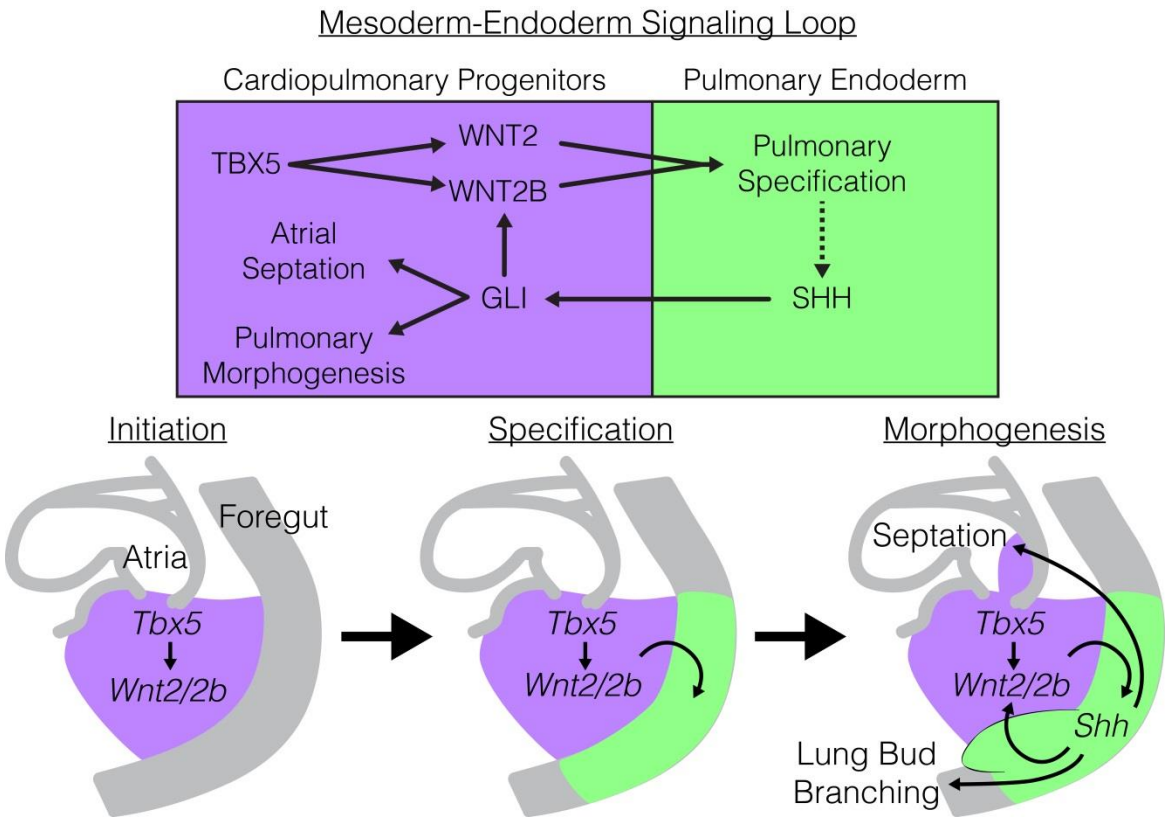


Figure II.7. *Tbx5* is required for a mesoderm–endoderm bidirectional signaling loop for cardiopulmonary development

Model of genetic interaction between *Tbx5*, canonical Wnt signaling, and Hh signaling for PE specification, pulmonary morphogenesis, and cardiac septation. TBX5, expressed in the CPPs (purple) initiates the bidirectional signaling loop through direct activation of *Wnt2* and *Wnt2b* expression. Canonical Wnt signaling drives pulmonary specification in the foregut endoderm and *Nkx2-1* expression. SHH, derived from the PE (green) signals back to the CPPs where it cooperatively activates *Wnt2b* but not *Wnt2*. Shh signaling drives both atrial septation and lung bud morphogenesis through previously described downstream targets.

of right lung agenesis [71] and one case of horseshoe lungs [72]. Additionally, a screen of patients with esophageal atresia and tracheoesophageal fistula identified a patient with HOS [73]. Further, rare genetic variants at the *TBX5*/*RBM19* locus have been associated with lung function in smokers by a genome-wide association study [74]. These studies suggest the intriguing possibility that *TBX5* may play a role in both lung development and adult lung function, similar to its requirement for both cardiac development and adult cardiac function [15, 61, 63, 75].

Cardiac septa are observed in all lunged vertebrates. We have previously demonstrated that GLI-dependent transcription downstream of PE Shh signaling and *Tbx5* cooperate in mesodermal SHF CPPs to drive atrial septation [7–9, 49]. Here, we demonstrate that *Tbx5* is required for the initiation of Shh signaling through the specification of PE. Previous work demonstrated that atrial septal defects caused by removal of *Tbx5* from the CPPs were rescued by concomitant activation of Hh signaling in those cells, providing epistatic evidence that *Tbx5* acts upstream of Hh signaling for atrial septation [9]. In this context, our current results suggest that Shh signaling from the PE to the CPPs is a direct requirement for atrial septation, while *Tbx5* may be dispensable following the initiation of lung development and subsequent Shh signaling. Our work demonstrates that *Tbx5* haploinsufficiency causes reduced *Wnt2* expression and subsequently reduced expression of *Shh* in the PE, resulting in reduced expression of quantitative markers of Shh reception in the cardiopulmonary mesoderm. Therefore, this quantitative decrement in CPP Hh signaling may contribute to the causation of cardiac septal defects in HOS patients.

The linked mesoderm–endoderm–mesoderm molecular pathways for lung development and cardiac inflow septation are conserved between amphibians and mammals. Remarkably, amphibians with evolutionary loss of lungs exhibit much reduced atrial septation, consistent with

necessary instructive cross-talk between these structures [3, 76]. We posit that *Tbx5*–*Wnt2*/*Wnt2b* signaling provides a molecular basis for the link between lung formation and the cardiac specializations required for pulmonary blood flow observed in lunged vertebrates.

Materials and Methods

Ethics Statement

All murine and zebrafish experiments were performed under University of Chicago Institutional Animal Care and Use Committee (IACUC) protocols no. 71737 and no. 71112. *X. laevis* and *X. tropicalis* adults were housed according to Cincinnati Children's Hospital Medical Center or University of North Carolina, Chapel Hill IACUC protocols. Handling of lizards (*Anolis sagrei*) and harvest of tissues complied with national and institutional guidelines and were approved by the IACUC of the University of Amsterdam (DAE101617).

Mouse Lines

Tbx5 germline mutant animals ($Tbx5^{+/-}$) were produced by crossing the $Tbx5^{tm1Jse}$ allele ($Tbx5^{flox}$) with a germline cre-recombinase and were out-crossed for multiple generations with CD-1 animals as previously described [15, 23]. Additionally, the Shh^{tm1Amc} [77] and Smo^{tm1Amc} [78] germline mutants have been previously described.

Xenopus *Experiments and CRISPR-Based Genome Editing*

Ovulation, *in vitro* fertilization, and dejellying of embryos were performed as described [79]. The pCS2+GR Tbx5 plasmid [70] was used to synthesize mRNA for injection using the Ambion mMessage mMachine SP6 RNA synthesis kit. GR-Tbx5 RNA (125 pg) was injected

into either the dorsal or ventral marginal zone (targeting the AME or PME, respectively) at the eight-cell stage. Validated Tbx5 translation-blocking MOs [35] were injected at the eight-cell stage (3.5 ng of each MO). See *SI Appendix* for full details. A small guide RNA (sgRNA) designed to target exon 5 of the *X. tropicalis* locus (GGGGTTCTGATATGAAGTGA) was coinjected at 200 pg with 2 ng Cas9 protein (PNA Bio) in 2-nL drops into one-cell-stage wild-type *X. tropicalis* embryos [80]. To screen rapidly for altered loci, an ∼500-bp genomic fragment asymmetrically flanking the sgRNA target sequence was amplified by PCR and subjected to digestion by T7 endonuclease (New England Biolabs).

Zebrafish Lines and Experiments

Zebrafish were maintained under standard laboratory conditions [81]. MO injections were performed as described [82]. *tbx5a* MO (3.7 ng) [47] and 5 ng of *tbx5b* translation-blocking MO [48] were injected into each embryo. Lines used were *AB and *heartstrings* (*hst*) mutants [45].

Transcriptional Profiling by RNA-Seq

RNA-seq was performed on microdissected CPP tissue at E9.5. Microdissected tissues from four embryos were pooled, and total RNA was extracted from five $Tbx5^{+/+}$ and two $Tbx5^{-/-}$ pools and was sequenced using the Illumina HiSeq 2500 platform by the Genomics Core Facility at the University of Chicago. Analysis was performed as previously described [23]. See *SI Appendix* for full details.

qRT-PCR

For mice, RNA was extracted from microdissected tissue as was done for RNA-seq. The reverse-transcription reaction was performed using SuperScript III First-Strand Synthesis SuperMix (Invitrogen). qRT-PCR was performed using Power SYBR Green PCR master mix (Applied Biosystems) and was run on an AB7500 machine (Applied Biosystems). Gene-expression level was normalized by *Gapdh*. For *Xenopus*, RNA was collected from three biological replicates containing four explants each. RNA was extracted using the Direct-zol RNA MiniPrep Plus kit (R2070; Zymo Research), and cDNA was generated using SuperScript VILO Master Mix (11755050; Thermo Fisher). Real-time PCR reactions were carried out using PowerUp SYBR Green Master Mix (A25742; Thermo Fisher) on ABI StepOnePlus qPCR machines (Applied Biosystems). *Ornithine decarboxylase* (*odc*) was used as a reference gene.

ISH

Mouse embryonic ISH was performed as previously described [7, 83, 84]. ISH of *Xenopus* embryos and explants was performed as described [79]. ISH was performed on stage-11 lizards as previously described [85]. The alligator and chicken sections come from a stained series used in previous publications [86, 87], but the sections shown have not been published before. See *SI Appendix* for full details.

Histology and 3D Reconstruction

All mouse and zebrafish embryonic histology was performed by the University of Chicago Human Tissue Resource Center. All tissues were fixed in formalin, embedded in paraffin wax, and sectioned to 10-μM thickness. Tissue was counterstained with H&E.

Reconstructions of embryonic lung histology were performed using AMIRA (5.3.2). See *SI Appendix* for full details.

ChIP and Analysis

Chromatin extract was prepared from microdissected tissue from E9.5 CD-1 mouse embryos (2× from50 embryo pools each) obtained from Charles River or from pelleted IMR90 cells (4× from 5 million cells each). For immunoprecipitation, the chromatin extract was incubated with anti-TBX5 antibody (sc-17866; lot no. G1516; Santa Cruz Biotechnology), anti-H3K4me3 (no. 305-34819; lot no. 14004; Wako Chemicals), or anti-H3K4me1 (ab8895; lot no. GR257926-1; Abcam). High-throughput sequencing libraries from ChIP and input DNA were prepared using the NEBNext Ultra DNA Library Prep Kit (E7370S; New England Biolabs) and were sequenced using Illumina HiSeq instruments by the Genomics Core Facility at the University of Chicago. ChIP-seq analysis was performed using a typical pipeline involving Bowtie2 [88] and MACS2 [89, 90]. See *SI Appendix* for full details.

ATAC-Seq and Analysis

ATAC-seq was performed as previously described [58] on an Illumina HiSeq system by the Genomics Core Facility at the University of Chicago. Analysis was performed in a similar manner to ChIP-seq. See *SI Appendix* for full details.

Luciferase Assays

pCDNA3.1 expression vector for *Tbx5* were previously described [63]. W2E1–3 were cloned into the pGL4.23 vector (Promega). Expression and reporter vectors were transfected into

HEK293T cells using FuGENE (Promega). Cells were cultured for 48 h after transfection and then were lysed and assayed using the Dual-Luciferase Reporter Assay System (Promega).

Transient Transgenics

Transient transgenic experiments were performed at E9.5 as previously described [7, 63, 64]. W2E1–3 were subcloned into the *Hsp68-LacZ* vector. The resulting construct was digested with NotI enzyme to remove the backbone, gel-purified, and injected into fertilized mouse eggs at the University of Chicago Transgenics Core Facility.

Tbx5OE-mESC Generation, CRISPR, and in Vitro *Differentiation*

The inducible *Tbx5*OE-mESC line was generated as previously published [65]. To generate the W2E mutants, we transfected the mESC with pSpCas9(BB)-2-Puro (PX459) plasmid vectors containing guides designed to generate an ~4.6- kbp deletion of W2E1 and W2E2. Following clone selection and expansion, two homozygous deletion clones (W2E mutants) and two wild-type clones (W2E controls) were evaluated. Cardiac stem cell differentiation was based on the original protocol from the laboratory of Keller and coworkers [66] with some modifications. For all mESC experiments utilizing overexpression, cells were treated with DOX (Sigma D9891) at the cardiac progenitor-like stage (day 6) and were harvested for RNA 24 h later. For CRISPR cell line evaluation, a dose of 100 ng/mL was used. See *SI Appendix* for full details.

Acknowledgements

We thank Lorenzo Pesce for use of the Beagle2 super computer partly supported by NIH Grant 1S10OD018495-01. This work was funded by NIH Grants R01 HL092153 and R01 HL124836 (to I.P.M.), R01 HD089275 (to F.L.C.), R01 HL126509 (to F.L.C. and I.P.M.), R01 DK070858, R01 HL114898, and P01 HD093363 (to A.M.Z.), R01 HD072598 (to R.K.H.), R01 HD084409 and P40 OD010997 (to M.E.H.), U01 HL100407 (to M.K.), R21 AG054770 (to K.I.), and R21 LM012619 (to X.H.Y.); support was also provided by NIH Grants T32 GM007183 (J.D.S.), T32 HL007381 (J.D.S., A.B.R., R.D.N.,M.R., and A.D.H.), T32 HD055164 (S.L.), and T32 GM007197 (A.D.H. and E.A.T.B.A.) and by Regenerative Medicine Minnesota Grant RMM 102516 001 (to S.S.-K.C.).

Contributions

The work presented in this chapter was highly collaborative in nature. Scott A. Rankin (Cincinnati Children's Hospital Medical Center, CCHMC), Aaron M. Zorn (CCHMC), Christopher E. Slagle (University of North Carolina, Chapel Hill, UNC-CH), Frank L. Conlon (UNC-CH), William Thomas (Marine Biological Laboratory, MBL), and Marko E. Horb (MBL) designed, perform, and analyze all experiments utilizing the *Xenopus* system. Erin A. T. Boyle Anderson (University of Chicago, UC) and Robert K. Ho (UC) helped design, perform, and analyze all experiments utilizing the zebrafish system. Ariel B. Rydeen (UC), Sunny Sun-Kin Chan (University of Minnesota, UM), Megan Rowton (UC), Sonja Lazarevic (UC), and Michael Kyba (UM) helped design, perform, and analyze all experiments utilizing the mESC system. Junghun Kweon (UC) and Xinan H. Holly (UC) helped design, perform, and analyze experiments related to RNA-seq. Rangarajan D. Nadadur (UC) helped perform ATAC-seq. Luis

Luna-Zurita (Gladstone Institute) provided ChIP-seq and Kohta Ikegami (UC) helped design, perform, and analyze ChIP-seq. Bjarke Jensen (University of Amsterdam) provided multi-species comparison. Jenna Bekeny (UC) helped design, perform, and analyze ChIP-PCR, qRT-PCR, and luciferase assays. Ivan P. Moskowitz (UC) helped design and analyze the work presented in this chapter.

References

1. Hsia CC, Schmitz A, Lambertz M, Perry SF, Maina JN (2013) Evolution of air breathing: Oxygen homeostasis and the transitions from water to land and sky. Compr Physiol 3:849–915.

2. Farmer CG (1999) Evolution of the vertebrate cardio-pulmonary system. Annu Rev Physiol 61:573–592.

3. Jensen B, Spicer DE, Sheppard MN, Anderson RH (2017) Development of the atrial septum in relation to postnatal anatomy and interatrial communications. Heart 103:456–462.

4. Jensen B, Wang T, Christoffels VM, Moorman AF (2013) Evolution and development of the building plan of the vertebrate heart. Biochim Biophys Acta 1833:783–794.

5. Peng T, et al. (2013) Coordination of heart and lung co-development by a multipotent cardiopulmonary progenitor. Nature 500:589–592.

6. Herriges M, Morrisey EE (2014) Lung development: Orchestrating the generation and regeneration of a complex organ. Development 141:502–513.

7. Hoffmann AD, et al. (2014) Foxf genes integrate tbx5 and hedgehog pathways in the second heart field for cardiac septation. PLoS Genet 10:e1004604.

8. Hoffmann AD, Peterson MA, Friedland-Little JM, Anderson SA, Moskowitz IP (2009) Sonic hedgehog is required in pulmonary endoderm for atrial septation. Development 136:1761–1770.

9. Xie L, et al. (2012) Tbx5-hedgehog molecular networks are essential in the second heart field for atrial septation. Dev Cell 23:280–291.

10. Basson CT, et al. (1994) The clinical and genetic spectrum of the Holt-Oram syndrome (heart-hand syndrome). N Engl J Med 330:885–891.

11. Newbury-Ecob RA, Leanage R, Raeburn JA, Young ID (1996) Holt-Oram syndrome: A clinical genetic study. J Med Genet 33:300–307.

12. Holt M, Oram S (1960) Familial heart disease with skeletal malformations. Br Heart J 22:236–242.

13. Li QY, et al. (1997) Holt-Oram syndrome is caused by mutations in TBX5, a member of the Brachyury (T) gene family. Nat Genet 15:21–29.

14. Basson CT, et al. (1997) Mutations in human TBX5 [corrected] cause limb and cardiac malformation in Holt-Oram syndrome. Nat Genet 15:30–35.

15. Bruneau BG, et al. (2001) A murine model of Holt-Oram syndrome defines roles of the T-box transcription factor Tbx5 in cardiogenesis and disease. Cell 106:709–721.

16. Goddeeris MM, Schwartz R, Klingensmith J, Meyers EN (2007) Independent requirements for hedgehog signaling by both the anterior heart field and neural crest cells for outflow tract development. Development 134:1593–1604.

17. Goss AM, et al. (2009) Wnt2/2b and beta-catenin signaling are necessary and sufficient to specify lung progenitors in the foregut. Dev Cell 17:290–298.

18. Zhang Y, et al. (2008) A Gata6-Wnt pathway required for epithelial stem cell development and airway regeneration. Nat Genet 40:862–870.

19. Rankin SA, et al. (2016) A retinoic acid-hedgehog cascade coordinates mesoderm-inducing signals and endoderm competence during lung specification. Cell Rep 16:66–78.

20. Harris-Johnson KS, Domyan ET, Vezina CM, Sun X (2009) Beta-catenin promotes respiratory progenitor identity in mouse foregut. Proc Natl Acad Sci USA 106:16287–16292.

21. Domyan ET, et al. (2011) Signaling through BMP receptors promotes respiratory identity in the foregut via repression of Sox2. Development 138:971–981.

22. Arora R, Metzger RJ, Papaioannou VE (2012) Multiple roles and interactions of Tbx4 and Tbx5 in development of the respiratory system. PLoS Genet 8:e1002866.

23. Waldron L, et al. (2016) The cardiac TBX5 interactome reveals a chromatin remodeling network essential for cardiac septation. Dev Cell 36:262–275.

24. Burnicka-Turek O, et al. (2016) Cilia gene mutations cause atrioventricular septal defects by multiple mechanisms. Hum Mol Genet 25:3011–3028.

25. Minoo P, et al. (1995) TTF-1 regulates lung epithelial morphogenesis. Dev Biol 172:694–698.

26. Minoo P, Li C, Liu HB, Hamdan H, deLemos R (1997) TTF-1 is an epithelial morphoregulatory transcriptional factor. Chest 111(Suppl):135S–137S.

27. Herriges MJ, et al. (2014) Long noncoding RNAs are spatially correlated with transcription factors and regulate lung development. Genes Dev 28:1363–1379.

28. Lazzaro D, Price M, de Felice M, Di Lauro R (1991) The transcription factor TTF-1 is expressed at the onset of thyroid and lung morphogenesis and in restricted regions of the foetal brain. Development 113:1093–1104.

29. Cardoso WV, Lü J (2006) Regulation of early lung morphogenesis: Questions, facts and controversies. Development 133:1611–1624.

30. Warburton D, et al. (2010) Lung organogenesis. Curr Top Dev Biol 90:73–158.

31. McCulley D, Wienhold M, Sun X (2015) The pulmonary mesenchyme directs lung development. Curr Opin Genet Dev 32:98–105.

32. Rankin SA, Zorn AM (2014) Gene regulatory networks governing lung specification. J Cell Biochem 115:1343–1350.

33. Koshiba-Takeuchi K, et al. (2009) Reptilian heart development and the molecular basis of cardiac chamber evolution. Nature 461:95–98, and erratum (2009) 461:550.

34. Rankin SA, et al. (2015) A molecular atlas of Xenopus respiratory system development. Dev Dyn 244:69–85.

35. Brown DD, et al. (2005) Tbx5 and Tbx20 act synergistically to control vertebrate heart morphogenesis. Development 132:553–563.

36. Goetz SC, Brown DD, Conlon FL (2006) TBX5 is required for embryonic cardiac cell cycle progression. Development 133:2575–2584.

37. Kimura S, et al. (1996) The T/ebp null mouse: Thyroid-specific enhancer-binding protein is essential for the organogenesis of the thyroid, lung, ventral forebrain, and pituitary. Genes Dev 10:60–69.

38. Poulain M, Ober EA (2011) Interplay between Wnt2 and Wnt2bb controls multiple steps of early foregut-derived organ development. Development 138:3557–3568.

39. Winata CL, et al. (2009) Development of zebrafish swimbladder: The requirement of hedgehog signaling in specification and organization of the three tissue layers. Dev Biol 331:222–236.

40. Yin A, Korzh S, Winata CL, Korzh V, Gong Z (2011) Wnt signaling is required for early development of zebrafish swimbladder. PLoS One 6:e18431.

41. Cass AN, Servetnick MD, McCune AR (2013) Expression of a lung developmental cassette in the adult and developing zebrafish swimbladder. Evol Dev 15:119–132.

42. Zheng W, et al. (2011) Comparative transcriptome analyses indicate molecular homology of zebrafish swimbladder and mammalian lung. PLoS One 6:e24019.

43. Tatsumi N, et al. (2016) Molecular developmental mechanism in polypterid fish provides insight into the origin of vertebrate lungs. Sci Rep 6:30580.

44. Longo S, Riccio M, McCune AR (2013) Homology of lungs and gas bladders: Insights from arterial vasculature. J Morphol 274:687–703.

45. Garrity DM, Childs S, Fishman MC (2002) The heartstrings mutation in zebrafish causes heart/fin Tbx5 deficiency syndrome. Development 129:4635–4645.

46. Ng JK, et al. (2002) The limb identity gene Tbx5 promotes limb initiation by interacting with Wnt2b and Fgf10. Development 129:5161–5170.

47. Ahn DG, Kourakis MJ, Rohde LA, Silver LM, Ho RK (2002) T-box gene tbx5 is essential for formation of the pectoral limb bud. Nature 417:754–758.

48. Parrie LE, Renfrew EM, Wal AV, Mueller RL, Garrity DM (2013) Zebrafish tbx5 paralogs demonstrate independent essential requirements in cardiac and pectoral fin development. Dev Dyn 242:485–502.

49. Zhou L, et al. (2015) Tbx5 and Osr1 interact to regulate posterior second heart field cell cycle progression for cardiac septation. J Mol Cell Cardiol 85:1–12.

50. Bellusci S, et al. (1997) Involvement of sonic hedgehog (Shh) in mouse embryonic lung growth and morphogenesis. Development 124:53–63.

51. Pepicelli CV, Lewis PM, McMahon AP (1998) Sonic hedgehog regulates branching morphogenesis in the mammalian lung. Curr Biol 8:1083–1086.

52. Peyrot SM, Wallingford JB, Harland RM (2011) A revised model of Xenopus dorsal midline development: Differential and separable requirements for notch and Shh signaling. Dev Biol 352:254–266.

53. Bowes JB, et al. (2010) Xenbase: Gene expression and improved integration. Nucleic Acids Res 38:D607–D612.

54. James-Zorn C, et al. (2018) Navigating Xenbase: An integrated Xenopus genomics and gene expression database. Methods Mol Biol 1757:251–305.

55. Karpinka JB, et al. (2015) Xenbase, the Xenopus model organism database; new virtualized system, data types and genomes. Nucleic Acids Res 43:D756–D763.

56. Chen JK, Taipale J, Cooper MK, Beachy PA (2002) Inhibition of hedgehog signaling by direct binding of cyclopamine to smoothened. Genes Dev 16:2743–2748.

57. Wu X, Ding S, Ding Q, Gray NS, Schultz PG (2002) A small molecule with osteogenesis inducing activity in multipotent mesenchymal progenitor cells. J Am Chem Soc 124: 14520–14521.

58. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ (2015) ATAC-seq: A method for assaying chromatin accessibility genome-wide. Curr Protoc Mol Biol 109:21.29.1–21.29.9.

59. He A, Kong SW, Ma Q, Pu WT (2011) Co-occupancy by multiple cardiac transcription factors identifies transcriptional enhancers active in heart. Proc Natl Acad Sci USA 108:5632–5637.

60. Luna-Zurita L, et al. (2016) Complex interdependence regulates heterotypic transcription factor distribution and coordinates cardiogenesis. Cell 164:999–1014.

61. Nadadur RD, et al. (2016) Pitx2 modulates a Tbx5-dependent gene regulatory network to maintain atrial rhythm. Sci Transl Med 8:354ra115.

62. Jolma A, et al. (2013) DNA-binding specificities of human transcription factors. Cell 152:327–339.

63. Arnolds DE, et al. (2012) TBX5 drives Scn5a expression to regulate cardiac conduction system function. J Clin Invest 122:2509–2518.

64. Smemo S, et al. (2012) Regulatory variation in a TBX5 enhancer leads to isolated congenital heart disease. Hum Mol Genet 21:3255–3263.

65. Iacovino M, et al. (2011) Inducible cassette exchange: A rapid and efficient system enabling conditional gene expression in embryonic stem and primary cells. Stem Cells 29:1580–1588.

66. Kattman SJ, et al. (2011) Stage-specific optimization of activin/nodal and BMP signaling promotes cardiac differentiation of mouse and human pluripotent stem cell lines. Cell Stem Cell 8:228–240.

67. Trujillo GV, et al. (2016) The canonical Wingless signaling pathway is required but not sufficient for inflow tract formation in the Drosophila melanogaster heart. Dev Biol 413:16–25.

68. Agarwal P, et al. (2003) Tbx5 is essential for forelimb bud initiation following patterning of the limb field in the mouse embryo. Development 130:623–633.

69. Rallis C, et al. (2003) Tbx5 is required for forelimb bud formation and continued outgrowth. Development 130:2741–2751.

70. Horb ME, Thomsen GH (1999) Tbx5 is essential for heart development. Development 126:1739–1751.

71. Tseng YR, et al. (2007) Holt-Oram syndrome with right lung agenesis caused by a de novo mutation in the TBX5 gene. Am J Med Genet A 143A:1012–1014.

72. Qin X, Wei W, Fangqi G (2015) Horseshoe lung associated with Holt-Oram syndrome. Iran J Pediatr 25:e251.

73. de Jong EM, Felix JF, de Klein A, Tibboel D (2010) Etiology of esophageal atresia and tracheoesophageal fistula: "Mind the gap". Curr Gastroenterol Rep 12:215–222.

74. Wain LV, et al.; UK Brain Expression Consortium (UKBEC); OxGSK Consortium (2015) Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): A genetic association study in UK Biobank. Lancet Respir Med 3:769–781.

75. Zhu Y, et al. (2008) Tbx5-dependent pathway regulating diastolic function in congenital heart disease. Proc Natl Acad Sci USA 105:5519–5524.

76. Lewis ZR, Hanken J (2017) Convergent evolutionary reduction of atrial septation in lungless salamanders. J Anat 230:16–29.

77. St-Jacques B, et al. (1998) Sonic hedgehog signaling is essential for hair development. Curr Biol 8:1058–1068.

78. Zhang XM, Ramalho-Santos M, McMahon AP (2001) Smoothened mutants reveal redundant roles for Shh and Ihh signaling including regulation of L/R symmetry by the mouse node. Cell 106:781–792.

79. Sive HL, Grainger RM, Harland RM (2000) Early Development of Xenopus laevis: A Laboratory Manual (Cold Spring Harbor Lab Press, Cold Spring Harbor, NY), p ix,
338 p.

80. Tandon P, Showell C, Christine K, Conlon FL (2012) Morpholino injection in Xenopus. Methods Mol Biol 843:29–46.

81. Westerfield M (2007) The Zebrafish Book (Univ Oregon Press, Eugene, OR), 5th Ed.

82. Nasevicius A, Ekker SC (2000) Effective targeted gene 'knockdown' in zebrafish. Nat Genet 26:216–220.

83. Wilkinson DG, Nieto MA (1993) Detection of messenger RNA by *in situ* hybridization to tissue sections and whole mounts. Methods Enzymol 225:361–373.

84. Takada S, et al. (1994) Wnt-3a regulates somite and tailbud formation in the mouse embryo. Genes Dev 8:174–189.

85. Jensen B, et al. (2012) Identifying the evolutionary building blocks of the cardiac conduction system. PLoS One 7:e44231.

86. Jensen B, et al. (2018) Specialized impulse conduction pathway in the alligator heart. eLife 7:e32120.

87. Poelmann RE, et al. (2014) Evolution and development of ventricular septation in the amniote heart. PLoS One 9:e106569.

88. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359.

89. Zhang Y, et al. (2008) Model-based analysis of ChIP-seq (MACS). Genome Biol 9:R137.

90. Feng J, Liu T, Qin B, Zhang Y, Liu XS (2012) Identifying ChIP-seq enrichment using MACS. Nat Protoc 7:1728–1740.

SI Appendix

Extended Methods

Xenopus *Experiments*

Xenopus explants were microdissected in 1× Modified Barth's saline (MBS) + 50 μg/mL gentamycin sulfate (1676045; MP Biochemicals) and then were cultured in 0.5× MBS + 0.1% fatty acid-free BSA (BP9704; Fisher) + 50 μg/mL gentamycin sulfate. Concentrations of small molecules or reagents were as follows: 20 μM purmorphamine (catalog no. 540220; EMD Millipore/Calbiochem); 1 μM DEX (D4902; Sigma); and 5 μg/mL CHX (C1988; Sigma). For DEX+CHX cotreatment, explants were incubated in CHX alone for 45 min before culture in DEX+CHX for 6 h.

Transcriptional profiling by RNA-seq

Embryos for RNA-seq were generated by crossing *Tbx5*$^{+/-}$ animals together in timed matings and embryos harvested and micro-dissected as shown in Figure II.1 A at E9.5. After genotyping, micro-dissected tissues from four embryos were pooled. Total RNA was extracted from five *Tbx5*$^{+/+}$ and two *Tbx5*$^{-/-}$ pools. 51-bp single-ended sequencing libraries (TruSeq RNA Sample prep kit v2; Part no. RS-122- 2001) were prepared and sequenced using the Illumina HiSeq2500 platform by the Genomics Core Facility at the University of Chicago (Invitrogen, 2013). We focused on the 38-bp reads on the right-side with a general declining pattern of quality scores as expected and ensured a larger than 30 quality score per base. Around 17-26 million RNA-seq reads were generated for each replicate and aligned to the GRCm38/mm10 build of the *Mus musculus* genome using TopHat v2.0.6 [1, 2]. Gene-level expression was quantified as read counts per exon using featureCounts in the Bioconductor package RSubRead

[3]. Reads overlapping exons in annotation build 38.1 of NCBI RefSeq database were included. Counts were converted to log2 counts, fit to a generalized linear model, and normalized to a gene- and sample specific normalization factor generated by the model with the DESeq2 package [4]. Data was plotted using the ggplot2 v2.2.1 [5] and scales v0.4.1 [6] packages in R [7]. The raw data was deposited in GEO with an accession number GSE75077.

In situ *Hybridization*

Mouse embryonic *in situ* hybridization was performed mainly as previously described [8-10]. RNA probes were generated using a digoxigenin RNA labeling kit (Roche), and hybridized overnight at 70°C onto bleached (6% Hydrogen Peroxide for 1 hour), proteinase-digested (10μg/ml proteinase K at 37°C for 8-10 minutes), and fixed (0.2% glutaraldehyde/4% PFA, 20 minutes) E9.5 embryos. Following hybridization, embryos were washed 2x30 minutes at 70°C in Solution I (50% formamide, 4X SSC, 1% SDS), 1x10 minutes at 70°C in a 1:1 solution of Solution I and Solution II (500mM NaCl, 10mM Tris 7.5, 0.1% Tween-20), 3x5 minutes Solution II at room temperature, RNase A treated (Sigma) in Solution II (37°C for 1 hour), washed in Solution II 1x5 minutes (room temperature), washed 1x5 minutes in Solution III (50% formamide, 2X SSC, 0.1% Tween-20 at room temperature), washed 2x30 minutes in Solution III at 65°C, and then washed 3x5 minutes in MABTL (100mM Maleic acid, 150mM NaCl, 0.1% Tween-20, and 2mM levamisole, pH 7.5, room temperature). Embryos were blocked for 3 hours in 10% sheep serum, 2% Boehringer Mannheim Blocking Reagent (BMBR) in MABTL. DIG-labeled probes were detected by anti-digoxin-AP Fab fragments (1:4000 dilution, Roche) in MABTL+2% BMBR (4°C overnight), washed 3x5 minutes in MABTL, washed 5-8x60 minutes MABTL, and washed overnight in MABTL (room temperature). Embryos were washed 3x10

minutes in NTMTL (100mM NaCl, 100mM Tris, pH9.5, 50mM MgCl2, 0.1% Tween-20, and 2mM levamisole) and precipitated by BM purple AP substrate (covered, room temperature, Roche). *Wnt2b* and *Nkx2-1* probes were provided by Ed Morrisey (Perelman School of Medicine, University of Pennsylvania). *Shh* probes were provided by Elizabeth Grove (University of Chicago). The stained section of *Tbx5* come from a series used in a previous publication [11], but the shown sections have not been previously published.

In situ hybridization of *Xenopus* embryos and explants was performed mostly as described [12]. Embryos and explants were fixed overnight (12-16 hours) at 4°C in MEMFA (0.1 M MOPS, 2 mM EGTA, 1 mM MgSO4, 3.7% formaldehyde), washed in MEMFA buffer without the formaldehyde 3x5 minutes, dehydrated directly into 100% ethanol, washed 4-6 times in 100% ethanol, and stored at -20°C. The following minor modifications to the *in situ* protocol were used: proteinase K (ThermoFisher AM2548) on day 1 was used at 2ug/mL for 10 minutes on explants, and at 3.5ug/mL for 10 minutes on whole embryos; the RNAse A step was consisted of 0.5ug/mL for 15 minutes on day 2; and finally the anti-DIG-alkaline phosphatase antibody (Sigma 11093274910) was used at a 1:5000 dilution in MAB buffer (maleic acid buffer, 100 mM Maleic acid, 150 mM NaCl, pH7.5) + 10% heat-inactivated lamb serum (Gibco 16070096) + 2% blocking reagent (Sigma 11096176001) on day2/3. Anti-sense DIG labeled *in situ* probes were generated using linearized plasmid full-length cDNA templates with the 10X DIG RNA labeling mix (Sigma 11277073910) according to manufacturer's instructions.

We used a stage 11 lizard, staged according to Sanger et al. [13]. *In situ* hybridization for *Tbx5* was performed as previously described [14]. The sections of the American alligator and chicken come from stained section series used in previous publications [15, 16], but the shown sections have not been published before.

3D Reconstruction

Reconstructions of embryonic lung histology were performed using AMIRA (version 5.3.2) on a Linux platform. Manual review of each image for a stack was performed and corrections were made when necessary. LabelFields for mesenchymal and endodermal tissue were generated from the same tissue using separate CastField and LabelVoxel modules. SurfaceGen modules were used to generate surfaces from these LabelFields. Tissue models for the two different tissues were initially aligned using the Landmark (2sets) module. A minimum of three landmarks were used to align the separate models. Final alignments were fine-tuned manually using the Transform editor.

Chromatin Immunoprecipitation and Analysis

To prepare chromatin extract, micro-dissected tissue (Figure II.1 A) from E9.5 CD-1 mouse embryos (2x from 50 embryo pools each) obtained from Charles River or pelleted IMR90 cells (4x from 5 million cells each) were cross-linked in PBS containing 1% formaldehyde at 25°C for 10 minutes. The reaction was quenched by 125 mM glycine. The cross-linked tissues were incubated in LB1 (50 mM HEPES-KOH, pH 7.5; 140 mM NaCl; 1 mM EDTA; 10% Glycerol; 0.5% NP-40; 0.25% Triton X-100) with protease and phosphatase inhibitors on ice for 10 minutes. The tissues were then sonicated in LB3 (10 mM Tris-HCl, pH 8.0; 100 mM NaCl; 1 mM EDTA; 0.5 mM EGTA; 0.1% sodium deoxycholate; 0.5% N-lauroyl sarcosine; 1% Triton X-100) with protease and phosphatase inhibitors. Chromatin extract was then cleared by centrifugation at 14,000g, 4°C for 10 minutes.

For immunoprecipitation, the chromatin extract was incubated with 5ug of the anti-TBX5 antibody (Santa Cruz Biotechnology sc-17866; Lot #G1516), 1µg of anti-H3K4me3 (Wako

Chemicals #305-34819; Lot #14004), or 1μg of anti-H3K4me1 (Abcam ab8895; Lot #GR257926-1) at 4°C for >12 hours in a total volume of 200 μL. The immune-complexes were captured by Protein G-conjugated magnetic beads (Life Technologies, 1003D), washed in sequence by LB3, LB3 with 1 M NaCl, LB3 with 0.5 M NaCl, LB3, and TE (10 mM Tris-HCl, pH 8.0; 1 mM EDTA). The captured chromatin was eluted in ChIP Elution Buffer (10 mM Tris-HCl, pH 8.0; 1 mM EDTA; 1% SDS; 250 mM NaCl) at 65°C. After RNase and proteinase K treatment and reverse cross-linking, DNA were purified. High-throughput sequencing libraries from ChIP and input DNA were prepared using NEBNext Ultra DNA Library Prep Kit (New England Biolabs, E7370S). During library preparation, adaptor-ligated DNA fragments of 200-650 bp in size were selected before PCR amplification using Sera-Mag magnetic beads (GE, 6515-2105-050-250). DNA libraries were sequenced using Illumina Hi-seq instruments (single-end 50 base) by the Genomics Core Facility at the University of Chicago.

Raw sequencing reads were aligned to the mm10 genome using Bowtie2 [17] and SAMtools [18, 19] requiring a minimum mapping quality of 10 (-q 10). Pooled peak calling was performed using default settings of MACS2 callpeak [20, 21] with a q-value set to 0.05 and tag size set to 6 (-q 0.05 -s 6). A fold-enrichment track was generated using MACS2 using the bdgcmp function (-m FE) for visualization on the genome browser. Peaks overlapping with ENCODE blacklist sites [22] were removed. TBX5 peaks were annotated in the genome browser using the R [7] package ChIPseeker [23] and TxDb.Mmusculus.UCSC.mm10.knownGene [24] with the promoter defined as within 2kb of the TSS. Density plots, heatmaps, and peak overlaps were generated using the R packages ChIPpeakAnno [25, 26], GenomicRanges [27], trackViewer [28], rtracklayer [29], RColorBrewer [30], and org.Mm.eg.db [31]. Correlation analysis between tissues was performed using multiBigwigSummary and plotCorrelation

functions of deepTools [32]. The raw data was deposited in GEO with an accession number GSE119885.

ATAC-seq and Analysis

ATAC-seq (Assay for Transposase-Accessible Chromatin using sequencing) was performed as previously described [33]. Nuclei were isolated from micro-dissected tissue (Figure II.1 A) from E9.5 CD-1 mouse embryos (2x from 5 embryo pools each) from Charles River according to published protocols and transposition reaction performed by published methods [33]. Libraries were amplified and normalized with the Illumina Nextera DNA Library prep kit (FC-121–1031) according to the manufacturer's protocols. Libraries were quantitated using the Agilent Bio-analyzer, pooled in equimolar amounts, and sequenced with 50-bp single-end reads on the Illumina HiSeq following the manufacturer's protocols through the Genomics Core Facility at the University of Chicago.

Sequencing reads were aligned to the mm10 genome using Bowtie2 [17] and SAMtools [18, 19] in a similar manner to the ChIP-seq. Peak calling was performed using MACS2 callpeak [20, 21] using the settings --nomodel --shift -100 --extsize 200 -q 0.05. A fold-enrichment track was generated using MACS2 using the bdgcmp function (-m FE) for visualization on the genome browser. The raw data was deposited in GEO with an accession number GSE119885.

Tbx5*OE-mESC Generation, CRISPR, and* In Vitro *Differentiation*

The inducible *Tbx5*OE-mESC line was generated as previously published [34]. *Tbx5* cDNA, without the stop codon, was Gateway cloned into the p2Lox plasmid, which was transfected into the A2Lox.cre mESC, which contains a doxycycline-inducible loxP-flanked cre-

recombinase upstream of the *Hprt* locus. Using cre-mediated cassette exchange, the cre locus was replaced with the *Tbx5* cDNA, generating the doxycycline-inducible *Tbx5*OE-mESC. Multiple clones were generated and characterized for growth and differentiation. The best clone was expanded and frozen down for subsequent experimentation.

In order to generate the W2E mutants, we transfected the mESC with pSpCas9(BB)-2-Puro (PX459) plasmid vectors containing one of four guides (TACTGGTAACGGCCCTCCCT, ACTGGTAACGGCCCTCCCTC, CCTAAGATGCAGGCGTGCGC, and TCCTAAGATGCAGGCGTGCG) designed to generate an approximately 4.6kbp deletion of W2E1 and W2E2 without disrupting the last exon of *St7* or the predicted splice acceptor. Following clone selection and expansion, two homozygous deletion clones (W2E mutants) and two wildtype clones (W2E controls) were successfully generated and evaluated. Stem cell differentiation to the cardiac lineage was based on the original protocol out of the lab of Gordon Keller [35] with some modifications. In brief, cells were treated with 1.5ng/ml BMP4 (R&D), 8ng/ml Activin A (R&D), 6ng/ml bFGF (R&D), and 200ug/ml Transferin (Sigma) on day 2 of the differentiation, and treated with 20ng/ml PDGFAA (eBioscience) on day 3. Live cells, positive for PDGFRα (Invitrogen) and Flk-1 (BioLegend), were sorted at day 3.75. Following cell sorting on day 4, cells were cultured in StemPro (LifeTechnologies) with 2mMol Glutamax, 1mM ascorbic acid, 1ug/ml SB431542 and 20mg/ml Normocine (Invitrogen).

Tbx5 overexpression dosage was determined through a doxycycline dose responsive curve using the following concentrations: 0, 5, 10, 25, 50, 100, 250, 500ng/mL. For all mESC experiments utilizing overexpression, *Tbx5*OE-mESC cells were treated with doxycycline (Sigma D9891) at the cardiac progenitor-like stage (day 6) then harvested for RNA 24 hours later. For CRISPR cell line evaluation, a dose of 100ng/mL was used. RNA isolation was

performed with the NucleoSpin RNA kit (Macherey-Nagel) and cDNA was generated using qScript cDNA Supermix (Quantabio).

SI Appendix Datasets

The following three datasets can be found at the PNAS website under the original publication, Steimle et al. 2018, "Evolutionarily conserved *Tbx5–Wnt2/2b* pathway orchestrates cardiopulmonary development" or at the url, https://www.pnas.org/content/115/45/E10615.

Dataset S1 – List of all significantly dysregulated genes (FDR < 0.05, FC ≥ 1.5 or FC ≤ 1/1.5) identified by RNA-seq in the micro-dissected CPPs from *Tbx5*+/+ and *Tbx5*-/- embryos at E9.5.

Dataset S2 – List of all 3883 peaks called by MACS2 for the TBX5 ChIP-seq, including positional information (mm10), fold-change over input, significance, and fold enrichment.

Dataset S3 – List of all 162 genes that were both significantly down and the closest gene to a TBX5 ChIP-seq peak. Some of these dysregulated genes were associated with more than one TBX5 ChIP-seq peak.

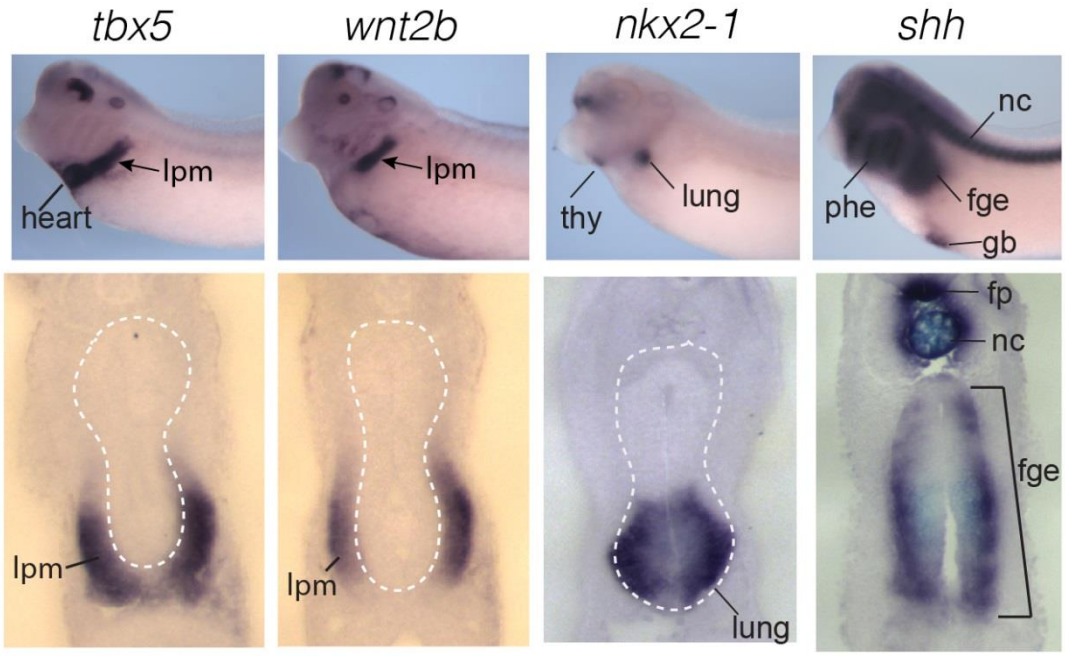


Figure II.S1. *Xenopus* whole mount *in situ* hybridization

RNA *in situ* hybridization was performed in NF35 *Xenopus* embryos for *tbx5*, *wnt2b*, *nkx2-1*, and *shh* (whole mount side-view on top, transverse cross-section on bottom). lpm = lateral plate mesoderm, thy = thyroid, nc = notochord, phe = pharyngeal endoderm, fge = foregut endoderm, gb = gall bladder, fp = floor plate.

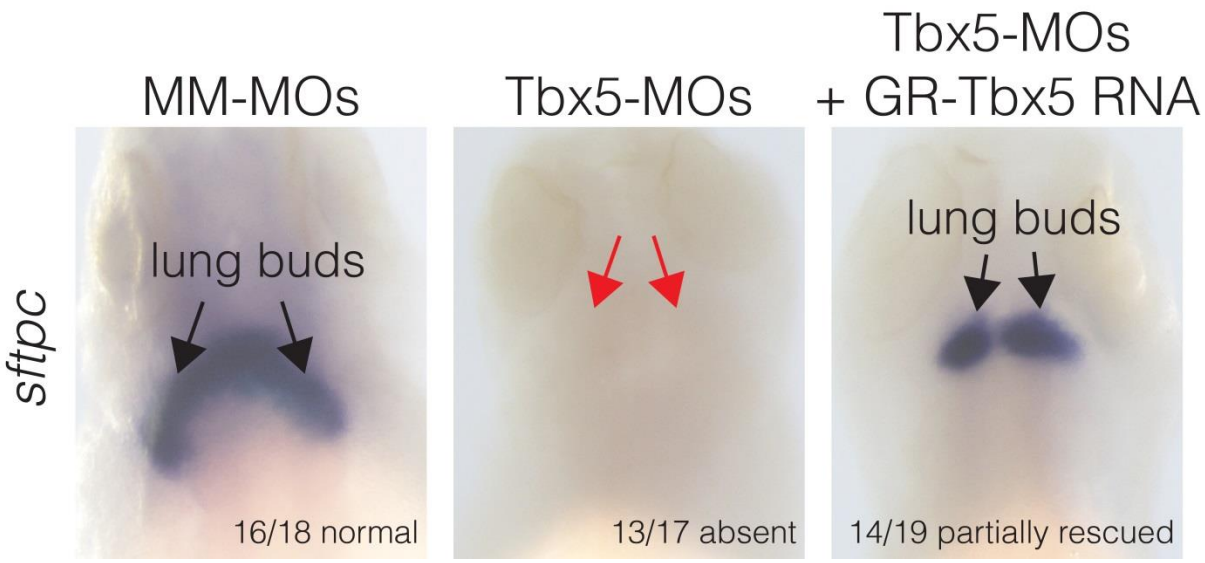


Figure II.S2. Expression of *sftpc* is *tbx5*-dependent in *Xenopus*

RNA *in situ* hybridization was performed for the pulmonary epithelium marker *sftpc* in NF42 *Xenopus* embryos injected with mismatched morpholinos (MM-MOs), Tbx5-MOs, or Tbx5-MOs and GR-Tbx5 RNA. Black arrows indicate the branching lung structures as indicated by *sftpc* expression, while the red indicates absence.

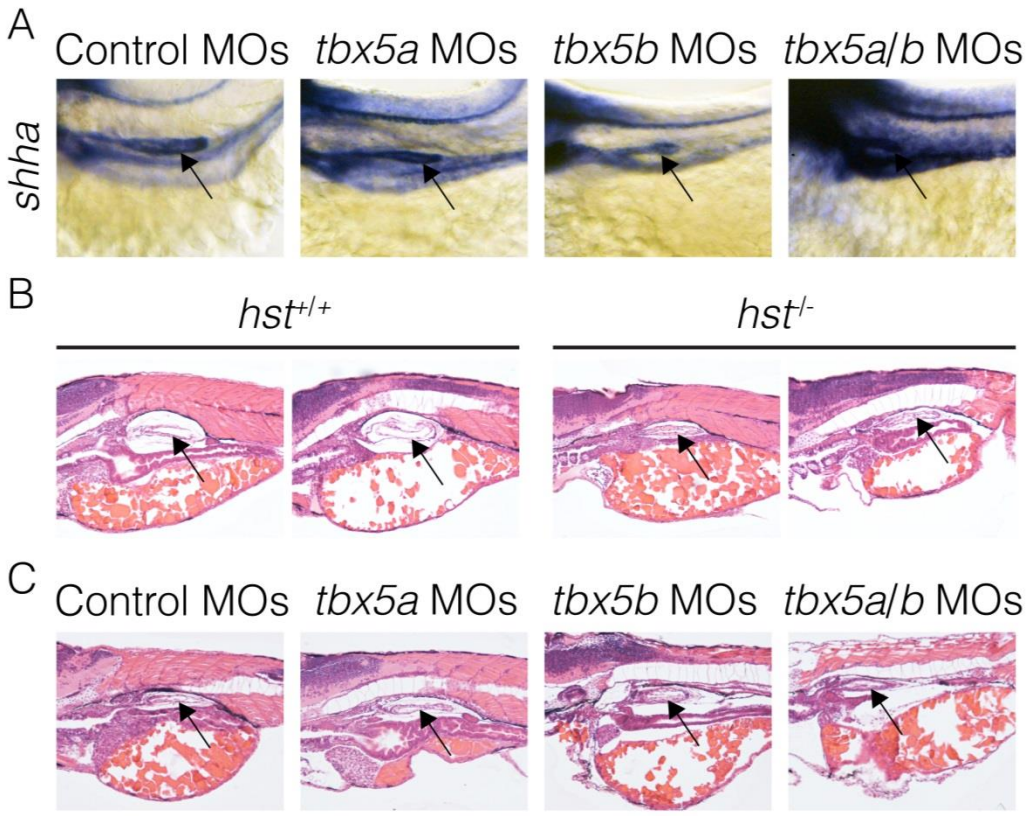


Figure II.S3. Initiation of the swim bladder occurs independent of *tbx5a/b* in zebrafish

A. RNA *in situ* hybridization was performed for *shha* in 72hpf zebrafish embryos injected with controls, *tbx5a*, *tbx5b*, or *tbx5a* and *tbx5b* morpholinos (MOs). B. Zebrafish embryos were collected at 96hpf and histologically sectioned and stained with H&E comparing *heartstrings* (*hst*) controls and homozygous mutants. C. Similarly, zebrafish embryos were collected at 96hpf and histologically sectioned and stained with H&E comparing control, *tbx5a*, *tbx5b*, and *tbx5a* and *tbx5b* MOs. In all images, the black arrows indicate the forming swim bladder.

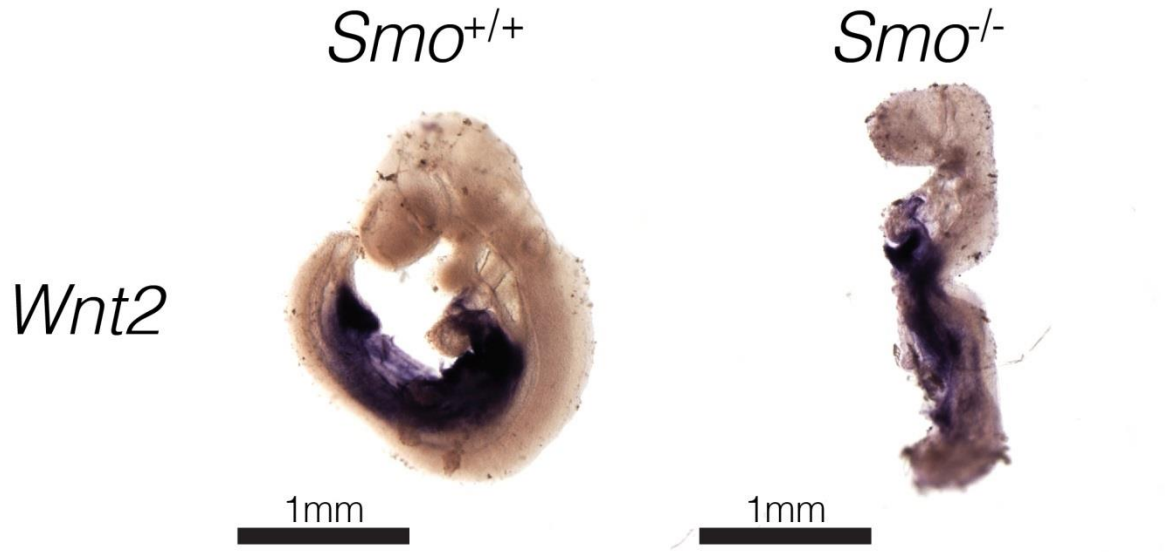


Figure II.S4. *Wnt2* expression is independent of Hedgehog signaling

Whole mount *in situ* hybridization for *Wnt2* was performed on E9.5 *Smo*+/+ and *Smo*-/- embryos.

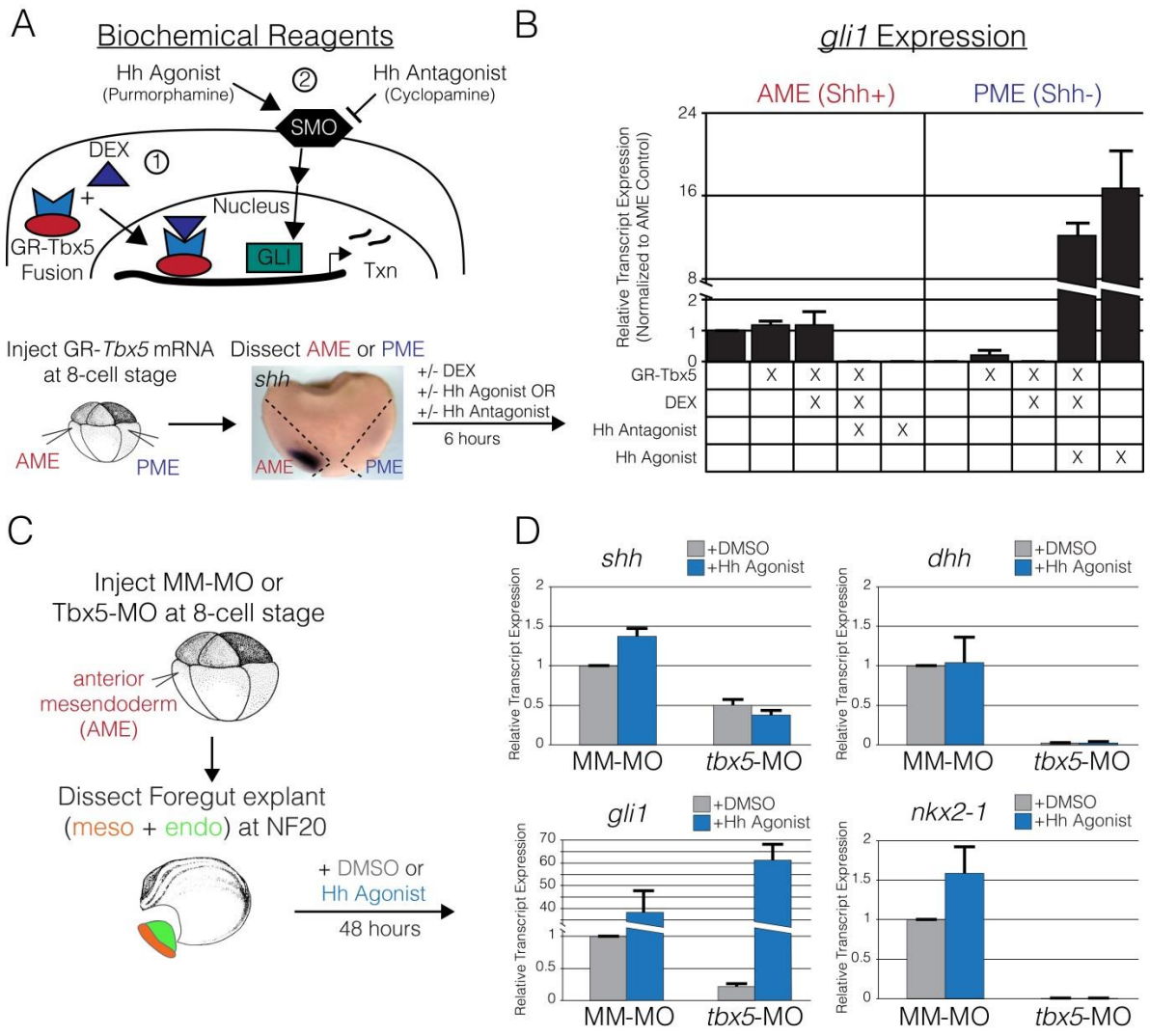


Figure II.S5. Utilizing *Xenopus* system to understand the genetic interaction between *Tbx5* and Shh signaling

A. To study the interaction of Tbx5 and Hedgehog signaling, we utilized: (1) a dexamethasone (DEX) inducible glucocorticoid receptor-Tbx5 (GR-Tbx5) fusion protein, and (2) pharmacological agonists (purmorphamine) and antagonists (cyclopamine) of SMO to activate or repress Hh signaling similar to Figure II.4 (top). Experimental details for testing Hedgehog response in anterior mesendoderm (AME, red, Shh+) and posterior mesendoderm (PME, blue, Shh-) tissues measured by *gli1* expression (bottom). Dissected AME or PME tissue was cultured

(Figure II.S5 continued) for 6 hours in combinations of dexamethasone (DEX, translocates GR-Tbx5 fusion protein to the nucleus), purmorphamine (Hh agonist), and cyclopamine (Hh antagonist). B. Expression of *gli1* examined by qRT-PCR in culture conditions as indicated by the checked boxes. All conditions were normalized to the AME controls (far left). C. Experimental plan for investigating markers of lung development in embryos injected with mistmatched morpholinos (MM-MOs) or Tbx5-MOs. Embryos were cultured in DMSO (grey) or Hh agonist (purmorphamine, blue) for 48 hours. D. qRT-PCR of *shh*, *dhh*, *gli1*, and *nkx2-1* from Tbx5-morphant or control embryonic explants treated with DMSO or Hh agonist.

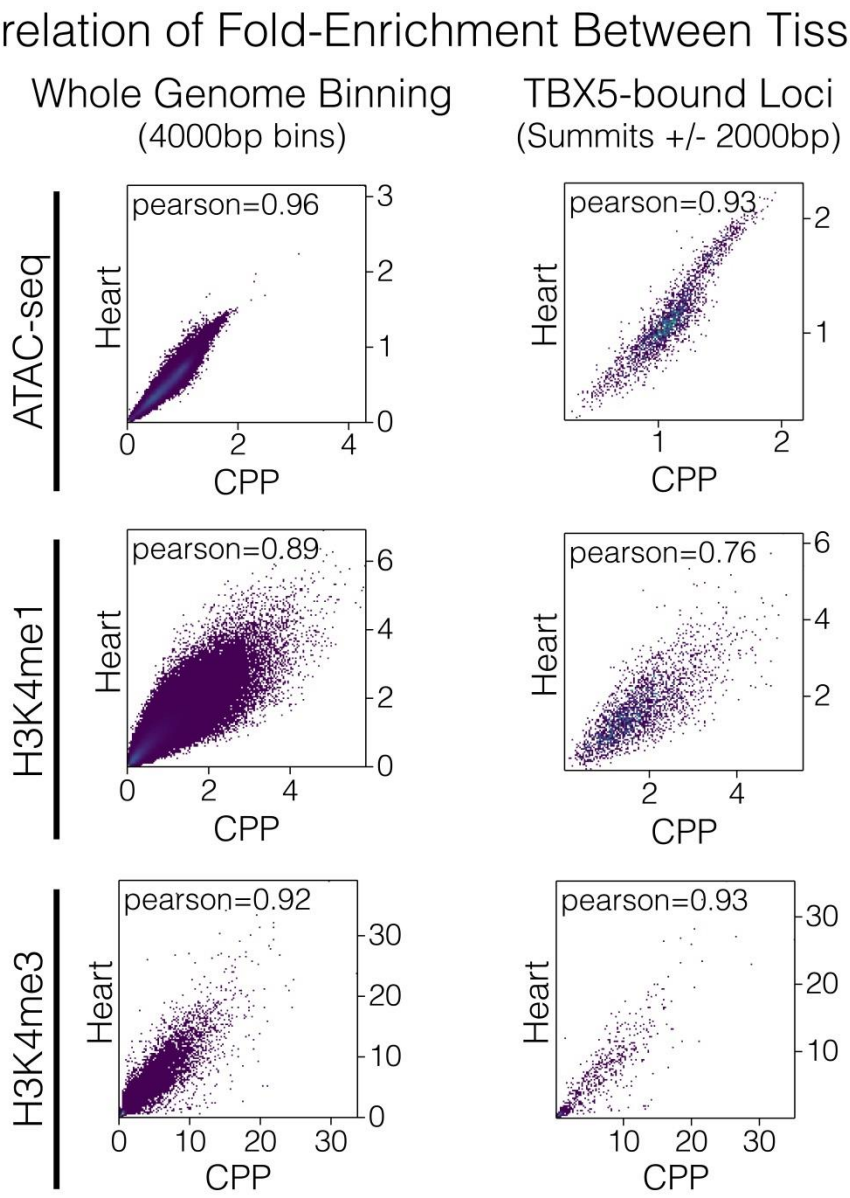


Figure II.S6. Genomic correlation between tissues

Fold-enrichment signals from the CPP micro-dissected tissue were plotted against fold-enrichment signals from the heart for each ATAC-seq, H3K4me1 ChIP-seq, and H3K4me3 ChIP-seq. On the left, the signal was calculated from 4000bp bins of the whole genome, while on the right, signal was determined from the 2000bp window to either side of a TBX5 ChIP-seq summit.

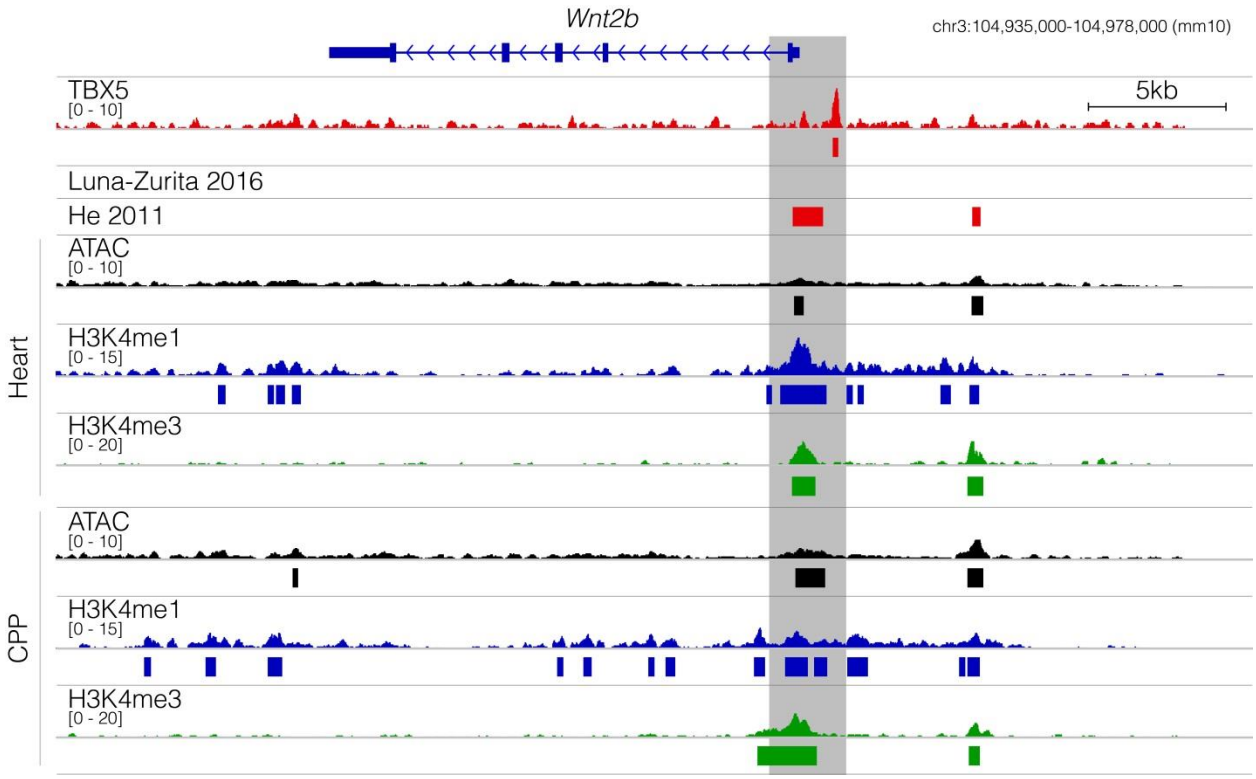
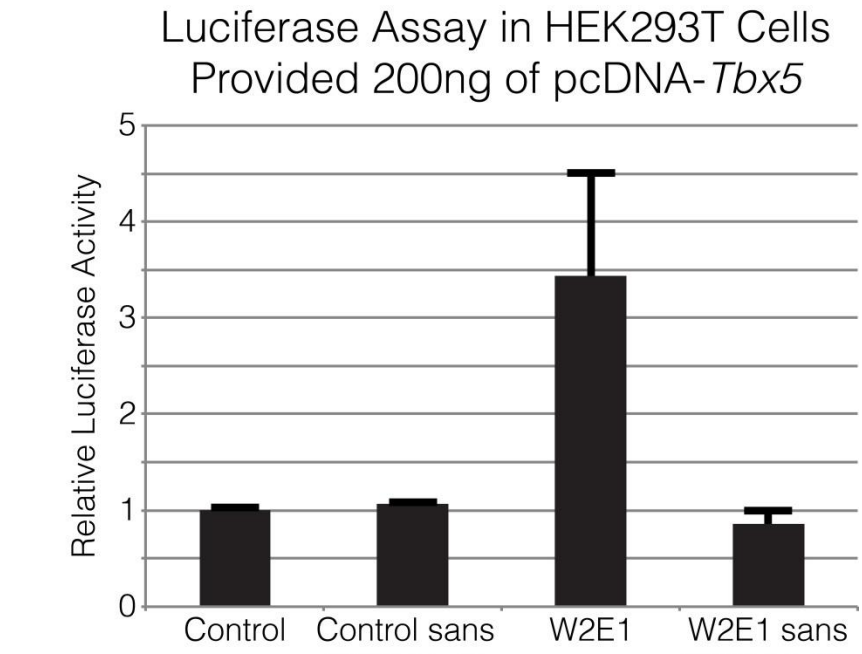


Figure II.S7. *Wnt2b* genomic locus

Genome browser view of *Wnt2b* and surrounding locus (chr3:104,935,000-104,978,000) with TBX5 ChIP-seq signals from the E9.5 heart and literature [42, 43], H3K4me1 and H3K4me3 ChIP-seq signals from the heart and CPP micro-dissected tissues, and ATAC-seq from the heart and CPP micro-dissected tissues. Significant peak calls below fold-enrichment signal tracks.



p-value:	Control sans TBX5 Motifs	W2E1	W2E1 sans TBX5 Motifs
Relative to Control	0.524	0.0309	0.308
Relative to W2E1			0.0252

Figure II.S8. Putative T-box motifs of W2E1 are required for TBX5-dependent activation in HEK293T cells

Luciferase assay examining TBX5 activation of the random sequence control, W2E1, and mutant variations of each lacking the canonical T-box motif. Data was normalized to the control vector.

SI Appendix References

1. Kim D, *et al.* (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14(4):R36.

2. Trapnell C, Pachter L, & Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25(9):1105-1111.

3. Liao Y, Smyth GK, & Shi W (2013) The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res* 41(10):e108.

4. Love MI, Huber W, & Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15(12):550.

5. Wickham H (2009) *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag New York).

6. Wickham H (2017) scales: Scale Functions for Visualization), R package version 0.5.0.

7. R Core Team (2016) R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).

8. Hoffmann AD, *et al.* (2014) Foxf genes integrate tbx5 and hedgehog pathways in the second heart field for cardiac septation. *PLoS Genet* 10(10):e1004604.

9. Wilkinson DG & Nieto MA (1993) Detection of messenger RNA by *in situ* hybridization to tissue sections and whole mounts. *Methods Enzymol* 225:361-373.

10. Takada S, *et al.* (1994) Wnt-3a regulates somite and tailbud formation in the mouse embryo. *Genes Dev* 8(2):174-189.

11. Xie L, *et al.* (2012) Tbx5-hedgehog molecular networks are essential in the second heart field for atrial septation. *Dev Cell* 23(2):280-291.

12. Sive HL, Grainger RM, & Harland RM (2000) *Early development of Xenopus laevis : a laboratory manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.) pp ix, 338 p.

13. Sanger TJ, Losos JB, & Gibson-Brown JJ (2008) A developmental staging series for the lizard genus Anolis: a new system for the integration of evolution, development, and ecology. *J Morphol* 269(2):129-137.

14. Jensen B, *et al.* (2012) Identifying the evolutionary building blocks of the cardiac conduction system. *PLoS One* 7(9):e44231.

15. Jensen B, *et al.* (2018) Specialized impulse conduction pathway in the alligator heart. *Elife* 7.

16. Poelmann RE, *et al.* (2014) Evolution and development of ventricular septation in the amniote heart. *PLoS One* 9(9):e106569.

17. Langmead B & Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4):357-359.

18. Li H, *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078-2079.

19. Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27(21):2987-2993.

20. Zhang Y, *et al.* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9(9):R137.

21. Feng J, Liu T, Qin B, Zhang Y, & Liu XS (2012) Identifying ChIP-seq enrichment using MACS. *Nat Protoc* 7(9):1728-1740.

22. Encode Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489(7414):57-74.

23. Yu G, Wang LG, & He QY (2015) ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* 31(14):2382-2383. 16

24. Bioconductor Core Team and Bioconductor Package Maintainer (2016) TxDb.Mmusculus.UCSC.mm10.knownGene: Annotation package for TxDb object(s)), R package version 3.4.0.

25. Zhu LJ, *et al.* (2010) ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics* 11:237.

26. Zhu LJ (2013) Integrative analysis of ChIP-chip and ChIP-seq dataset. *Methods Mol Biol* 1067:105-124.

27. Lawrence M, *et al.* (2013) Software for computing and annotating genomic ranges. *PLoS Comput Biol* 9(8):e1003118.

28. Ou J, Wang Y-X, & Zhu LJ (2016) trackViewer: A bioconductor package with minimalist design for drawing elegant tracks or lollipop plot), R package version 1.10.2.

29. Lawrence M, Gentleman R, & Carey V (2009) rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics* 25(14):1841-1842.

30. Neuwirth E (2014) RColorBrewer: ColorBrewer Palettes), 1.1-2.

31. Carlson M (2016) org.Mm.eg.db: Genome wide annotation for Mouse), 3.4.0.

32. Ramirez F, *et al.* (2016) deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* 44(W1):W160-165.

33. Buenrostro JD, Wu B, Chang HY, & Greenleaf WJ (2015) ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol* 109:21 29 21-29.

34. Iacovino M, *et al.* (2011) Inducible cassette exchange: a rapid and efficient system enabling conditional gene expression in embryonic stem and primary cells. *Stem Cells* 29(10):1580-1588.

35. Kattman SJ, *et al.* (2011) Stage-specific optimization of activin/nodal and BMP signaling promotes cardiac differentiation of mouse and human pluripotent stem cell lines. *Cell Stem Cell* 8(2):228-240.

CHAPTER III: REGIONAL AND TBX5-DEPENDENT GENE EXPRESSION IN THE ATRIA: IMPLICATIONS FOR PULMONARY VEIN DEVELOPMENT AND ATRIAL FIBRULLATION

Abstract

Atrial fibrillation, characterized by irregular atrial depolarization leading to an uncoordinated contraction of the atrial muscle, is the most common cardiac arrhythmia. The T-box transcription factor *TBX5* regulates atrial rhythm and has been genetically associated with human atrial fibrillation by both familial and genome-wide association studies. Over the last 30 years, the pulmonary veins and peri-pulmonary region of the left atrium have been implicated as a physical locus that frequently initiates atrial fibrillation. Regional gene expression differences within the atria have been reported that may reflect distinct ontogeny of these regions and impact arrhythmia mechanisms. In this study, we investigated transcriptional differences between regions of the adult left atrium in the mouse. We performed and compared transcriptional profiling of the left atrial appendage and peri-pulmonary vein atrium of the adult mouse. We identified novel markers and transcriptional differences between the two regions, including of multiple known AF genes. Furthermore, we compared *Tbx5*-dependent genes between each region and identified *Hcn4* as a novel dysregulated gene in the peri-pulmonary vein atrium. Regional and *Tbx5*-dependent gene expression differences between the left atrial appendage and peri-pulmonary vein atrium are considered in the context of the genetic basis of atrial fibrillation.

Introduction

Atrial fibrillation and the pulmonary vein

Atrial fibrillation (AF) is the most common sustained cardiac arrhythmia, affecting more than 7 million Americans and 33 million people worldwide [1]. AF is characterized by an irregular pattern of atrial depolarization, resulting in rapid and disorganized atrial conduction and lack of effective atrial chamber contraction. The rhythm abnormality in AF manifests with circulatory deficits and systemic thromboembolism that greatly increase morbidity and mortality. Patients with AF have an increased risk of developing major complications including heart failure and stroke. In addition, the prevalence of AF is expected to rise significantly as the population ages. AF has become a major clinical and economic burden, owing to the limitations and side effects associated with current AF therapies. Although AF most often manifests in the context of pre-existing cardiac pathologies, idiopathic or lone AF forms indicated a heritable component [2]. Over the last decade, genome-wide association studies (GWAS) have identified over one hundred AF-associated loci [3, 4]. These advances in understanding the genetic basis of AF portend an understanding of the complex molecular pathways provoking AF.

The current paradigm of AF requires two arrhythmogenic components: 1) ectopic (triggered) atrial myocardial activity such as early or delayed afterdepolarizations and 2) a fibrillogenic substrate that propagates these triggers throughout the atrial chambers [5]. The most common source of triggered activity is the pulmonary veins, which connect the lungs to the left atria [6, 7]. It has been shown that the pulmonary vein (PV) and peripulmonary vein (PPV) myocardium can demonstrate automaticity which can trigger atrial depolarization [8]. A pacemaker current has been observed in the PV-left atrial junction in atrial fibrillation, and gene expression contributing to myocardial automaticity, such as the HCN4 channel protein, is

observed in the pulmonary veins [9]. A common treatment for AF is to isolate the PV myocytes from the atrial myocardium by catheter ablation, severing the electrical connectivity between these two populations and preventing inappropriate PV depolarizations from initiating global atrial depolarization and atrial arrhythmias [6, 7].

An important advance in the understanding of the pathophysiology of AF has been the demonstration that the pulmonary vein plays an important role in the initiation of AF [6, 10]. Other initiating foci include the superior vena cava and the left atrium; however, the pulmonary vein remains the most common source of focal activity. This observation has stimulated consideration of the distinctions between the myocardium of the pulmonary/peri-pulmonary vein region and other regions of the atrium. However, the gene expression profile promoting focal activity around the PVs in humans has been difficult to elucidate, owing to the limitations of tissue samples from patients. The molecular identity of the PV myocardium, including specific ion channels and transcription factors expressed in this tissue, has only recently been described. Analysis of a whole-lung transcriptome data set uncovered a set of 24 transcripts, including sarcomeric structural proteins, genes regulating sarcomere assembly, ion transport proteins and hormone signaling in the pulmonary cardiomyocyte gene network [11]. Some of these genes have been linked to AF in humans, suggesting that perturbation of this transcriptional network might lead to altered calcium handling, altered myocardium contractile force and electrophysiological properties which may contribute to the initiation of atrial arrhythmias.

TBX5 *and pulmonary vein development*

Holt-Oram syndrome is an autosomal dominant disorder caused by mutations in the T-box family transcription factor *TBX5* [12-14], occurring in approximately 1 in 100,000 live births

[15, 16]. The clinical diagnosis of Holt-Oram syndrome includes completely penetrant upper-limb malformations and variable congenital heart defects, most commonly septal and cardiac conduction defects [17]. Specifically, the conduction defects can manifest as long PR interval, AV and bundle branch block, bradycardia, sick sinus syndrome, and atrial fibrillation with or without overt structural defects [18, 19]. The observation that conduction abnormalities can occur in the absence of structural heart defects was an early indication that TBX5 may play a direct role in controlling gene expression essential for normal cardiac conduction. Recently, GWAS have implicated *TBX5* in altered cardiac conduction speed (PR and QRS intervals) and AF in structurally normal hearts. Although strict clinical diagnosis of Holt-Oram Syndrome does not involve defects of the pulmonary veins [14, 17], pulmonary vein abnormalities are often associated with Holt-Oram Syndrome and *TBX5* mutations [20-22]. Furthermore, homozygous mutation of *Tbx5* in the mouse germline [23] results in complete morphological failure of the cardiopulmonary progenitors, which would normally generate portions of the atria, the atrial septum, pulmonary vein myocardium and smooth muscle cells, and airways smooth muscle cells [24, 25]. These observations suggested that *TBX5* may drive gene expression important for the development of the pulmonary vasculature in the embryo and for the suppression of atrial arrhythmias in the pulmonary veins or peripulmonary vein atrial myocardium in the adult.

Here, we describe regional differences between the transcriptome of the left atrial appendage and the peri-pulmonary vein of the left atrium from adult mice. RNA-sequencing revealed previously known and novel regional differences, including several genes implicated in atrial fibrillation. We further compared *Tbx5*-depedent transcripts between the left atrial appendage and the peri-pulmonary vein of the left atrium from adult mice. Important distinctions were observed between the two regions including *Hcn4*, which demonstrated differential *Tbx5*-

dependence. Our data suggests that underlying regional differences in the transcriptome of the atrial myocardium may enlighten susceptibility to atrial fibrillation.

Results

Comparison of peri-pulmonary vein and atrial appendage transcriptomes

We investigated transcriptional differences between regions of the left atrium in 7-week-old mice. The left atrium was microdissected into two parts, the left atrial appendage (LAA) and the peri-pulmonary vein (PPV) (Figure III.1A). Transcriptional profiling was performed by RNA-sequencing (RNA-seq; n = 4 in each group) as previously described [26]. Principal component analysis (PCA) demonstrated that atrial region was responsible for the first principal component, accounting for 32% of the variation between samples (Figure III.1B). Differential expression testing between the PPV and LAA samples identified 747 significantly different genes ($|\log2FC| > 1$; FDR < 0.01) with 582 expressed more strongly in the PPV samples and 165 expressed more strongly in the LAA samples (Figure III.1C, full table available at GEO accession GSE128870). Gene ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were examined using metascape (http://metascape.org) for both the LAA and PPV genes [27]. GO terms associated with the PPV indicated a preponderance of genes associated with immune cells. We observed GO terms that were unambiguously related to the immune system and only containing genes associated with the PPV, including phagocytosis (GO:0006909), inflammatory response (GO:0006954), humoral immune response (GO:0006959), lymphocyte differentiation (GO:0030098), and cytokine-cytokine receptor interaction (mmu04060). Although these 141 genes are of potential interest, they were separated from the remaining 441 PPV genes to focus

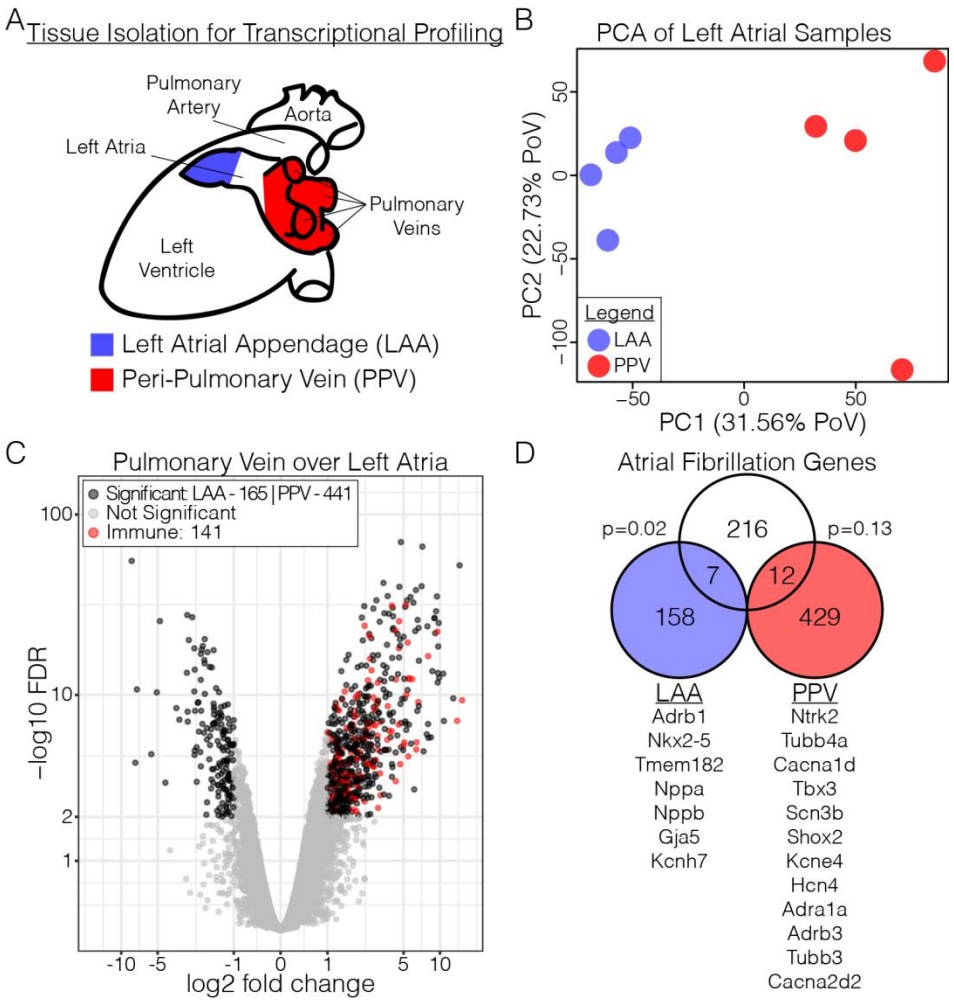


Figure III.1. Transcriptional comparison of the left atrial appendage and peripulmonary vein atrium

A. The left atrial appendage (LAA, blue) and the peri-pulmonary vein region of the left atrium (PPV, red) was microdissected from 7-week-old adult mice. B. Principal component analysis comparing the LAA (blue) and PPV (red) transcriptomes. C. Volcano plot showing log2(fold-change PPV/LAA). 165 were associated with LAA and 582 were associated with the PPV (|log2FC|>1 and FDR< 0.01). Of the PPV genes, 141 genes were associated with GO terms related to the immune system (red) and removed from downstream analysis. D. Overlap of genes

(Figure III.1 continued) associated with atrial fibrillation and either the LAA (7 genes) or PPV (12 genes). P-values derived from Fisher's exact test.

on those genes previously verified as myocardially expressed or those with unknown roles or expression domains (Figure III.1C).

Because of the relationship between the left atria, pulmonary vein, and AF, we examined region-specific gene expression against a list of 235 AF associated genes [3, 28-30]. Of the 606 genes that demonstrated regionalized expression in our RNA-seq, 19 have been associated with AF (Figure III.1D). Specifically, *Adrb1*, *Nkx2-5*, *Tmem182*, *Nppa*, *Nppb*, *Gja5*, and *Kcnh7* are more strongly expressed in the LAA, while *Ntrk2*, *Tubb4a*, *Cacna1d*, *Tbx3*, *Scn3b*, *Shox2*, *Kcne4*, *Hcn4*, *Adra1a*, *Adrb3*, *Tubb3*, and *Cacna2d2* have higher expression in the PPV.

Tbx5-*dependent transcriptome of peri-pulmonary vein atrium identifies genes involved with cardiac conduction*

We hypothesized that *Tbx5*-dependent gene expression may vary between different regions in the left atria. Previous work has demonstrated *Tbx5* regulates a gene regulatory network for atrial identity and its removal from the adult mouse results in atrial fibrillation [26]. *Tbx5* is robustly expressed in both the LAA and PPV tissues with an average TPM of 79.7 ± 9.9SEM and 97.1 ± 18.4SEM, respectively putting *Tbx5* within the top 8% of all genes expressed in either area of the left atrium (TPM > 1). To investigate differential *Tbx5*-dependent regional gene expression in the left atrium of the adult mouse, we performed *Tbx5*-dependent transcriptional profiling of the PPV tissue and compared it to previously described *Tbx5*-dependent transcriptional profiling of the left atrial appendage [26]. We performed RNA-seq on the PPV of an adult-specific conditional knockout of *Tbx5* [26]. Control (*R26creERT2*/*creERT2*; *Tbx5*+/+) and conditional mutant (*R26creERT2*/*creERT2*; *Tbx5flox*/*flox*) animals were treated with tamoxifen at 6-weeks-old and tissue was harvested one week later, prior to the onset or

atrial arrhythmias (Figure III.1A) [26]. PCA analysis demonstrated clustering along PC1 and PC2 of samples (cumulative variation > 50%). PC1 demonstrated clustering associated with genotype (control versus conditional *Tbx5* mutant) and PC2 demonstrated clustering associated with LA region (LAA versus PPV), indicating that genotype differences drove the majority of variance followed by regional atrial differences (Figure III.2A).

We identified 338 significantly dysregulated genes (|log2FC| > 1; FDR < 0.01) comparing the PPV between *Tbx5* conditional mutant and controls (Figure III.2B, GEO accession GSE128870). Of these, 149 were up-regulated and 189 were down-regulated in the *Tbx5* mutants (Figure III.2B). Utilizing metascape [27], we identified 7 representative GO terms associated with the up-regulated genes and 2 representative GO terms associated with the down-regulated genes (Figure III.2C). The down-regulated GO terms are reminiscent of the LAA previously reported [26]. In particular, examining the full set of GO terms under "multicellular organismal signaling," we identified multiple GO terms associated with the regulation of cardiac conduction, including ion channels, action potential, and transmembrane signaling (Figure III.2D).

We identified 1492 significantly dysregulated genes in the LAA conditional mutant and control, with 770 up-regulated and 722 down-regulated, similar to our prior analysis of this dataset [26]. Although the numbers of significantly dysregulated genes were different between the PPV and LAA, the dysregulation between the LAA and PPV experiments were significantly correlated (p-value < 2.2E-16, Figure III.2E), suggesting that the removal of *Tbx5* had a broadly similar effect in the LAA and PPV regions. In order to identify genes demonstrating distinct *Tbx5*-dependent effects between the atrial subregions, we correlated the *Tbx5*-dependent changes between the LAA and PPV samples. We identified the 28 genes with residuals in the outermost

one-thousandth quantiles (Table III.1). Amongst those genes, we identified 11 genes that were significantly different in both experiments with 10/11 demonstrating changes in the same direction, 4 genes that were significantly increased in only the LAA samples (*Cemip*, *Col9a2*, *Itgb7*, *Lars2*), 7 genes that were significantly different in only the PPV samples (Increased: *Gm42417*, *Lrtm1*; Decreased: *Alox8*, *Cntn2*, *Hcn4*, *Scarna10*, *Vsnl1*), and 6 genes that were not significantly different in either sample. Interestingly, among the genes that were in the outermost one-thousandth quantiles, we identified *Hcn4*, which encodes the potassium channel essential for the depolarizing "funny current" of the sinoatrial node. *Hcn4* was significantly downregulated only in the PPV and was also significantly more expressed in the PPV samples than the LAA (Figure III.1D).

Discussion

The genetics of atrial fibrillation implicates developmental transcription factors

Most genetic loci implicated in AF encode ion channels that affect arrhythmia trigger or substrate. Initial transcriptional studies of AF have focused on a restricted set of candidates, associating remodeling with changes in ion channels and components of cellular signaling cascades [31, 32]. Early transcriptomic or proteomic investigation of AF has been limited to animal models and/or chronic AF [33-36]. For example, animal models based on disruption of specific ion channels do not exhibit spontaneous AF in the absence of concomitant cardiac pathophysiology [37]. However, a recent transcriptomic study using the left atrial appendage of patients with AF revealed that AF susceptibility was associated with decreased expression of transcription factors involved in inflammation, oxidation, and the cellular stress response [38]. This study further suggested that remodeling of ion channel expression occurs later in the onset

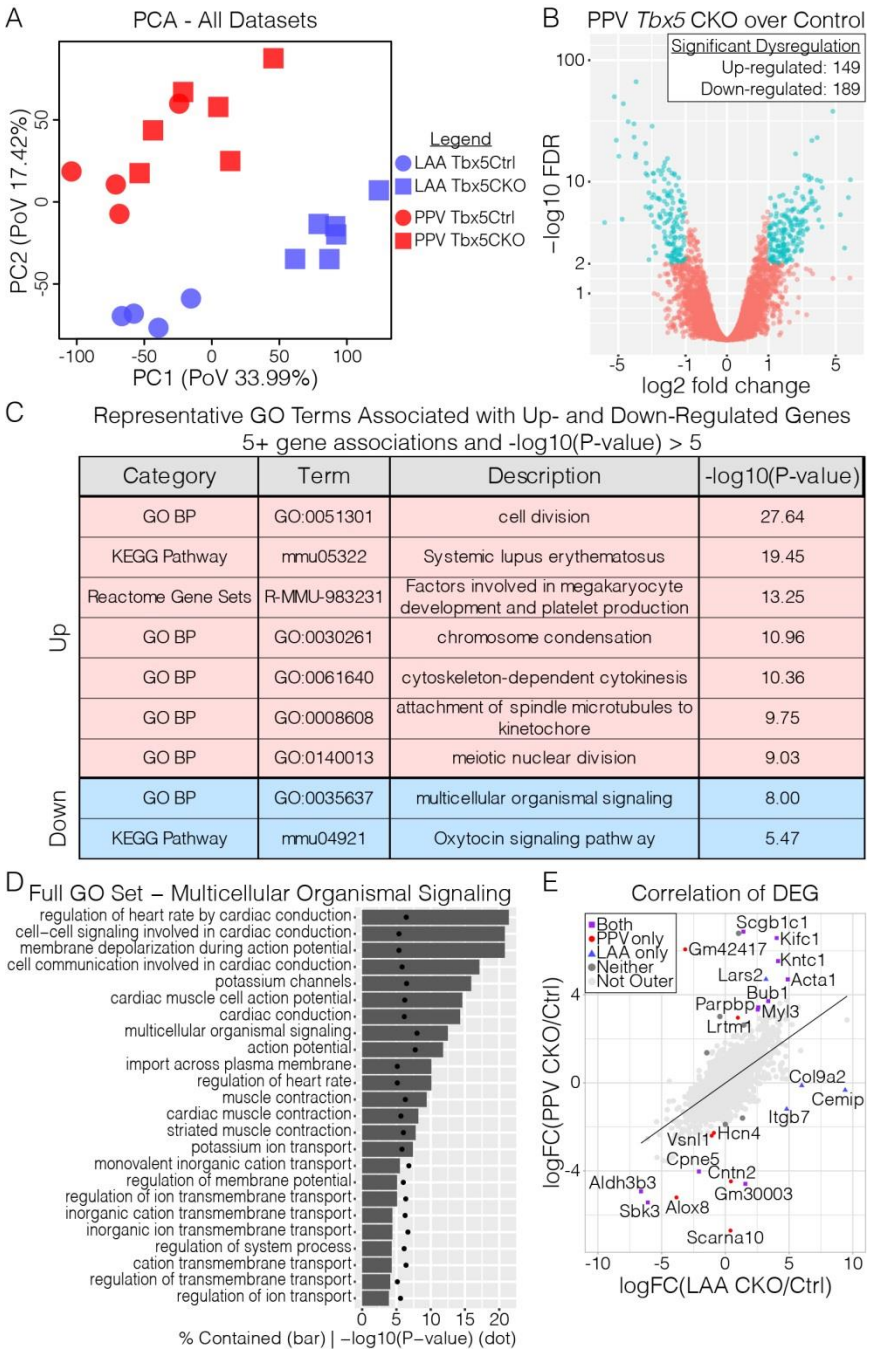


Figure III.2. *Tbx5* conditional knockout from the peri-pulmonary vein reveals *Tbx5*-dependence of multiple unique features

A. Principal component analysis comparing the left atrial appendage (LAA, blue) and peri-pulmonary vein (PPV, red) transcriptomes from control (*R26creERT2*/*creERT2*; *Tbx5*+/+,

(**Figure III.2. continued)** "Tbx5Ctrl", circles) and *Tbx5* conditional knockout (*R26creERT2*/*creERT2*; *Tbx5flox*/*flox*, "Tbx5CKO", squares) adult mice. B. Volcano plot comparing the PPV *Tbx5* CKO and control samples. 149 genes were significantly upregulated and 189 were significantly down-regulated (|log2FC|>1 and FDR< 0.01). C. Summary GO terms representing all GO terms associated with up- and down-regulated genes (red and blue, respectively) with greater than 5 associated genes and -log10(Pvalue) greater than 5. D. Full GO term list represented by "multicellular organismal signaling" with percent of genes identified (bar) and corresponding -log10(P-value) (dot). E. Correlation of fold changes between the LAA and PPV samples. Colors and shapes represent whether the genes residuals were in the outermost one-thousandth quantiles (purple square, red circle, blue triangle, dark grey circle) or not (light grey circle). Furthermore, colors and shapes represent whether those outliers were significantly different in both datasets (purple square), the PPV only (red circle), the LAA only (blue triangle), or neither (dark grey circle).

of the disease. Accumulating evidence suggest that transcription factors are important contributors to the pathogenesis of AF [39].

GWAS of AF have identified numerous transcription factor loci that play fundamental roles during cardiac development, including *TBX5*, *NKX2-5*, *GATA4*, and *PITX2*. The implication of transcription factor genes essential for cardiac development in in AF raises the fundamental question of whether AF is a function or a result of a developmental defect, for example of the pulmonary veins, or a later manifestation of altered transcriptional control of cardiac rhythm genes in the adult. The first GWAS of AF identified a susceptibility locus on chromosome 4q25 adjacent to *PITX2* [40]. There is evidence that *Pitx2c* affects adult gene expression in the atrium. *PITX2* is expressed both in the left atrium and the pulmonary myocardial sleeve and that its levels are decreased in patients with AF [41, 42]. Atrial specific deletion of *Pitx2c* leads to hallmarks of AF, which is also observed in *Pitx2c* heterozygous mice or adult-specific deletion of this gene [41, 43, 44]. Furthermore, microarray analysis of *Pitx2c* heterozygous mice revealed impaired gap and tight junction gene expression, consistent with the hypothesis that structural genes are remodeled later in the onset of AF [44].

We previously demonstrated that adult-specific deletion of *Tbx5* causes spontaneous and sustained AF with atrial gene regulatory network dysfunction [26]. TBX5 drives the atrial expression of *Pitx2*, and TBX5 and PITX2 co-modulate the expression of cardiac rhythm effector genes, including *Ryr2* and *Atp2a2*. These findings indicated that interactions between TBX5 and PITX2 provide tight control of an atrial rhythm gene regulatory network and that perturbation of this network triggered AF susceptibility. This example suggested that cardiac TFs implicated in AF by genetic association may coregulate a gene regulatory network for atrial rhythm homeostasis in the adult atrium.

Gene Symbol	LAA - Tbx5CKO/Ctrl	PPV - Tbx5CKO/Ctrl	Short Description
Acta1	Increased	Increased	skeletal muscle alpha-actin associated with myopathies (65)
Bub1	Increased	Increased	spindle assembly checkpoint and centromeric cohesion (66)
Kifc1	Increased	Increased	mitotic kinesin protein involved in centromsome duplication (67)
Kntc1	Increased	Increased	kinetochore-binding mitotic checkpoint gene (68)
Myl3	Increased	Increased	ventricular and slow skeletal muscle myosin light chain (69)
Parpbp	Increased	Increased	inhibits inappropraiate homologous recombination in mitosis (70)
Scgb1c1	Increased	Increased	Secretoglobin protein associated with anti-inflammatory function (71)
Aldh3b3	Decreased	Decreased	removes lipid-derived aldehydes generated by oxidative stress (72)
Cpne5	Decreased	Decreased	calcium-dependent, phospholipid-binding proteins (73)
Sbk3	Decreased	Decreased	uncharacterized cardiac expressed, SH3 domain binding kinase (74)
Gm30003	Increased	Decreased	lung-expressed predicted gene (75)
Cemip	Increased	NS	depolymerizes hyaluronic acid (76)
Col9a2	Increased	NS	Type IX collagen found with type II fibrillar collagen (77)
Itgb7	Increased	NS	leukocyte associated integrin beta subunit (78)
Lars2	Increased	NS	mitochondrial leucyl-tRNA synthetase (79)
Gm42417	NS	Increased	predicted gene (80)
Lrtm1	NS	Increased	transmembrane protein with homology to *Slit3* (81)
Alox8	NS	Decreased	peroxidizes polyunsaturated fatty acids (82)
Cntn2	NS	Decreased	membrane junction protein associated with axons (83)
Hcn4	NS	Decreased	hyperpolarization-activated current / "funny" current (84)
Scarna10	NS	Decreased	small Cajal body-specific RNA 10 (85)
Vsnl1	NS	Decreased	neuronal calcium sensor (86)
Gm13938	NS	NS	antisense lncRNA (80)
Gm33051	NS	NS	kidney-expressed antisense lncRNA (75)
Igf2	NS	NS	fetal growth hormone signaling and tissue differentiation (87)
Rab3c	NS	NS	calcium triggered vesicle transport (88)
Slc27a2	NS	NS	fatty acid transport (89)
Snhg3	NS	NS	competing endogenous RNA to miR-182-5p (90)

Table III.1. Genes displaying distinct *Tbx5*-dependent expression patterns

The 28 genes identified as outliers in the outermost one-thousandth quartiles when comparing the *Tbx5*-dependent gene changes in the LAA and the PPV. The genes are listed with whether they were significantly increased (red), decreased (blue), or not significant (NS) in each of the two comparisons. Furthermore, a short description of each gene is provided.

Transcription factors and AF – A pulmonary vein developmental relationship?

The requirement of developmental TFs in the adult for normal cardiac rhythm control does not rule out the possibility that developmental defects caused by their deficiency in the embryo also contributes to AF susceptibility. During embryonic development, the cardiopulmonary progenitors give rise to parts of the atria, pulmonary vein, and lungs [25]. Distinct functional characteristics between the PPV and LAA regions may reflect distinct developmental ontogeny. The myocardium of the pulmonary veins has a distinct developmental origin from that of the atrial free walls [45]. Subsequently, the myocardium surrounding the pulmonary veins is distinct from that of other regions in the atrium [46]. These developmental distinctions may set the stage for regional gene expression differences within the mature atrium of the adult that normally maintains atrial rhythm but which make the PPV region susceptible to trigger formation when faced with environmental or genetic insult.

Mutations in *NKX2-5, PITX2C, and TBX5* increase susceptibility to AF, and all of these transcription factors (TF) play critical roles during cardiovascular development [26, 47-50]. Each of these TFs are expressed in the pulmonary veins, suggesting that perturbations of the transcription program in the PV myocardium may underlie the pathogenesis of AF. The caudal myocardium, including the developing pulmonary vein myocardium, expresses *Tbx5*, *Pitx2c*, *Nkx2-5*, and *Gja5* (Cx40) at high levels [51, 52] consistent with a fast-conducting atrial myocardial phenotype. In a *Nkx2-5* haploinsufficient mouse model, *Gja5* levels are decreased whereas *Hcn4* levels are increased in the pulmonary vein myocardium. *Nkx2-5* and *Pitx2c* are both required for the development and maintain identity of the pulmonary myocardium [45, 53]. The changes in gene expression observed in *Nkx2-5* haploinsufficient mice could be reflective of a loss of fast conduction and acquisition of pacemaking conduction phenotype. This type of

transition provided a conceivable model for how genetic changes could result in automaticity in the PPV region that resulted in arrhythmia initiation.

Absence of the lungs and pulmonary vasculature in *Tbx5* knockout mice suggest that it may play an essential developmental role specifically within the pulmonary veins as well [24]. We observed that a decrement in *Tbx5* in the PPV region decreased *Hcn4* expression, which is not overtly consistent with this transformation model. On the other hand, we observed that many genes with important roles in myocardial automaticity, including *Tbx3, Shox2, Cacna1d, Cacna2d2* and *Hcn4* were more highly expressed in the PPV than the LAA region, and remained more highly expressed in the context of adult-specific *Tbx5* conditional deletion. Furthermore, other *Tbx5*-dependent genes in the PPV region are classical fast conduction genes, including *Scn5a, Gja5,* and *Kcnk3*. Reduction in the expression of these fast conduction genes in the context of retained high expression of automaticity genes such as *Hcn4* could diminish fast conduction physiology and allow the emergence of myocardial automaticity, providing a nidus for the initiation of atrial arrhythmias including AF.

Methods

Animal Experiments and Ethics Statement

All murine experiments were performed under University of Chicago Institutional Animal Care and Use Committee (IACUC) protocol number 71737. Mice harboring the ROSA26 tamoxifen-inducible cre recombinase transgene, *Gt(ROSA)26Sortm1(cre/ERT2)Tyj* (*R26creERT2*), were bred with mice harboring the *Tbx5tm1Jse* allele (*Tbx5flox*) to generate *Tbx5* conditional mutants, *R26creERT2*/*creERT2*; *Tbx5flox*/*flox*, and controls, *R26creERT2*/*creERT2*;*Tbx5*+/+ [23, 54]. Tamoxifen was administered by intraperitoneal

injection for three consecutive days at 6 weeks of age and tissue samples were harvested 1 week after treatment as previously reported [26].

Transcriptional Profiling and Analysis

Transcriptional profiling was performed on the peri-pulmonary vein region of the left atrium (Figure III.1A). Total RNA was extracted from four controls and 6 conditional mutants. Following rRNA removal (Ribo-Zero rRNA Removal Kit, Illumina), libraries were prepared using the TruSeq RNA Sample prep kit v2 (Illumina). Samples were pooled in equimolar ratios and sequenced on the Illumina HiSeq4000 platform. Library preparation and sequencing was performed by the Genomics Core Facility at the University of Chicago.

The peri-pulmonary vein samples were compared with the previously published corresponding left atrial appendage samples [26]. Both the peri-pulmonary vein and left atrial appendage samples were analyzed in parallel as described below. Between 20 and 22 million reads were generated for each replicate (14-19 million reads for left atrial appendage) and aligned to the GRCm38/mm10 build of the *Mus musculus* genome (retrieved from UCSC May 23, 2012) using TopHat v2.1.1 [55, 56]. Reads were filtered to remove poorly aligned (MAPQ<10), duplicated, and unmapped reads using BamTools v2.4.1 [57]. Post-alignment, post-filtered reads were assigned to transcripts using StringTie v1.3.3 [58, 59].

Downstream analysis performed using R v3.4.0 [60]. Differential expression testing was performed using edgeR v3.18.1 [61] and limma v3.32.10 [62]. Low level genes were removed within each condition using median log-transformed counts per gene per million mapped reads (cpm) cutoff of 1. A generalized linear model (GLM) framework was used to test for differential expressed genes between *Tbx5fl/fl*;*R26cre-ERT2* conditional mutants and *Tbx5+/+*;*R26cre-ERT2*

controls for each tissue and between controls of the left atrial appendage and peri-pulmonary vein tissues. Data was plotted using the ggplot2 v2.3.0 [63] and scales v0.5.0 [64] packages. GO term analysis was performed using metascape (http://metascape.org) [27]. The raw and processed data as well as the full outputs were deposited in GEO at accession number GSE128870.

Contributions

The work presented in this chapter was collaborative in nature. Brigitte Laforest (University of Chicago, UC) and Michael T. Broman (UC) helped review and contextualize the work within the literature. Rangarajan D. Nadadur (UC) designed and performed the RNA-seq. Ivan P. Moskowitz (UC) helped design and analyze the work presented in this chapter.

References

1. Nishida K, Datino T, Macle L, & Nattel S (2014) Atrial fibrillation ablation: translating basic mechanistic insights to the patient. *J Am Coll Cardiol* 64(8):823-831.

2. Tucker NR & Ellinor PT (2014) Emerging directions in the genetics of atrial fibrillation. *Circ Res* 114(9):1469-1482.

3. Fatkin D, Santiago CF, Huttner IG, Lubitz SA, & Ellinor PT (2017) Genetics of Atrial Fibrillation: State of the Art in 2017. *Heart Lung Circ* 26(9):894-901.

4. Nielsen JB, *et al.* (2018) Genome-wide Study of Atrial Fibrillation Identifies Seven Risk Loci and Highlights Biological Pathways and Regulatory Elements Involved in Cardiac Development. *Am J Hum Genet* 102(1):103-115.

5. Andrade J, Khairy P, Dobrev D, & Nattel S (2014) The clinical profile and pathophysiology of atrial fibrillation: relationships among clinical features, epidemiology, and mechanisms. *Circ Res* 114(9):1453-1468.

6. Haissaguerre M, *et al.* (1998) Spontaneous initiation of atrial fibrillation by ectopic beats originating in the pulmonary veins. *N Engl J Med* 339(10):659-666.

7. Jais P, *et al.* (1997) A focal source of atrial fibrillation treated by discrete radiofrequency ablation. *Circulation* 95(3):572-576.

8. Chen YJ, Chen SA, Chang MS, & Lin CI (2000) Arrhythmogenic activity of cardiac muscle in pulmonary veins of the dog: implication for the genesis of atrial fibrillation. *Cardiovasc Res* 48(2):265-273.

9. Nguyen BL, Fishbein MC, Chen LS, Chen PS, & Masroor S (2009) Histopathological substrate for chronic atrial fibrillation in humans. *Heart Rhythm* 6(4):454-460.

10. Chen SA, *et al.* (1999) Initiation of atrial fibrillation by ectopic beats originating from the pulmonary veins: electrophysiological characteristics, pharmacological responses, and effects of radiofrequency ablation. *Circulation* 100(18):1879-1886.

11. Boutilier JK, *et al.* (2017) Gene Expression Networks in the Murine Pulmonary Myocardium Provide Insight into the Pathobiology of Atrial Fibrillation. *G3 (Bethesda)* 7(9):2999-3017.

12. Basson CT, *et al.* (1997) Mutations in human TBX5 [corrected] cause limb and cardiac malformation in Holt-Oram syndrome. *Nat Genet* 15(1):30-35.

13. Li QY, *et al.* (1997) Holt-Oram syndrome is caused by mutations in TBX5, a member of the Brachyury (T) gene family. *Nat Genet* 15(1):21-29.

14. McDermott DA, *et al.* (2005) TBX5 genetic testing validates strict clinical criteria for Holt-Oram syndrome. *Pediatr Res* 58(5):981-986.

15. Elek C, Vitez M, & Czeizel E (1991) [Holt-Oram syndrome]. *Orv Hetil* 132(2):73-74, 77-78.

16. Barisic I, *et al.* (2014) Holt Oram syndrome: a registry-based study in Europe. *Orphanet J Rare Dis* 9:156.

17. Holt M & Oram S (1960) Familial heart disease with skeletal malformations. *Br Heart J* 22:236-242.

18. Basson CT, *et al.* (1994) The clinical and genetic spectrum of the Holt-Oram syndrome (heart-hand syndrome). *N Engl J Med* 330(13):885-891.

19. Newbury-Ecob RA, Leanage R, Raeburn JA, & Young ID (1996) Holt-Oram syndrome: a clinical genetic study. *J Med Genet* 33(4):300-307.

20. Huang T, *et al.* (2002) Causes of clinical diversity in human TBX5 mutations. *Cold Spring Harb Symp Quant Biol* 67:115-120.

21. Bohm J, *et al.* (2008) Functional analysis of the novel TBX5 c.1333delC mutation resulting in an extended TBX5 protein. *BMC Med Genet* 9:88.

22. Boogerd CJ, *et al.* (2010) Functional analysis of novel TBX5 T-box mutations associated with Holt-Oram syndrome. *Cardiovasc Res* 88(1):130-139.

23. Bruneau BG, *et al.* (2001) A murine model of Holt-Oram syndrome defines roles of the Tbox transcription factor Tbx5 in cardiogenesis and disease. *Cell* 106(6):709-721.

24. Steimle JD, *et al.* (2018) Evolutionarily conserved Tbx5-Wnt2/2b pathway orchestrates cardiopulmonary development. *Proc Natl Acad Sci U S A* 115(45):E10615-E10624.

25. Peng T, *et al.* (2013) Coordination of heart and lung co-development by a multipotent cardiopulmonary progenitor. *Nature* 500(7464):589-592.

26. Nadadur RD, *et al.* (2016) Pitx2 modulates a Tbx5-dependent gene regulatory network to maintain atrial rhythm. *Sci Transl Med* 8(354):354ra115.

27. Tripathi S, *et al.* (2015) Meta- and Orthogonal Integration of Influenza "OMICs" Data Defines a Role for UBR4 in Virus Budding. *Cell Host Microbe* 18(6):723-735.

28. Bapat A, Anderson CD, Ellinor PT, & Lubitz SA (2018) Genomic basis of atrial fibrillation. *Heart* 104(3):201-206.

29. van Setten J, *et al.* (2018) PR interval genome-wide association meta-analysis identifies 50 loci associated with atrial and atrioventricular electrical activity. *Nat Commun* 9(1):2904.

30. Kramer A, Green J, Pollard J, Jr., & Tugendreich S (2014) Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* 30(4):523-530.

31. Goette A, *et al.* (2000) Increased expression of extracellular signal-regulated kinase and angiotensin-converting enzyme in human atria during atrial fibrillation. *J Am Coll Cardiol* 35(6):1669-1677.

32. Brundel BJ, *et al.* (2001) Alterations in potassium channel gene expression in atria of patients with persistent and paroxysmal atrial fibrillation: differential regulation of protein and mRNA levels for K+ channels. *J Am Coll Cardiol* 37(3):926-932.

33. Barth AS, *et al.* (2005) Reprogramming of the human atrial transcriptome in permanent atrial fibrillation: expression of a ventricular-like genomic signature. *Circ Res* 96(9):1022-1029.

34. Lamirault G, *et al.* (2006) Gene expression profile associated with chronic atrial fibrillation and underlying valvular heart disease in man. *J Mol Cell Cardiol* 40(1):173-184.

35. Cardin S, *et al.* (2007) Contrasting gene expression profiles in two canine models of atrial fibrillation. *Circ Res* 100(3):425-433.

36. De Souza AI, *et al.* (2010) Proteomic and metabolomic analysis of atrial profibrillatory remodelling in congestive heart failure. *J Mol Cell Cardiol* 49(5):851-863.

37. Riley G, Syeda F, Kirchhof P, & Fabritz L (2012) An introduction to murine models of atrial fibrillation. *Front Physiol* 3:296.

38. Deshmukh A, *et al.* (2015) Left atrial transcriptional changes associated with atrial fibrillation susceptibility and persistence. *Circ Arrhythm Electrophysiol* 8(1):32-41.

39. Mahida S (2014) Transcription factors and atrial fibrillation. *Cardiovasc Res* 101(2):194-202.

40. Gudbjartsson DF, *et al.* (2007) Variants conferring risk of atrial fibrillation on chromosome 4q25. *Nature* 448(7151):353-357.

41. Chinchilla A, *et al.* (2011) PITX2 insufficiency leads to atrial electrical and structural remodeling linked to arrhythmogenesis. *Circ Cardiovasc Genet* 4(3):269-279.

42. Kirchhof P, *et al.* (2011) PITX2c is expressed in the adult left atrium, and reducing Pitx2c expression promotes atrial fibrillation inducibility and complex changes in gene expression. *Circ Cardiovasc Genet* 4(2):123-133.

43. Wang J, *et al.* (2010) Pitx2 prevents susceptibility to atrial arrhythmias by inhibiting left-sided pacemaker specification. *Proc Natl Acad Sci U S A* 107(21):9753-9758.

44. Tao Y, *et al.* (2014) Pitx2, an atrial fibrillation predisposition gene, directly regulates ion transport and intercalated disc genes. *Circ Cardiovasc Genet* 7(1):23-32.

45. Mommersteeg MT, *et al.* (2007) Pitx2c and Nkx2-5 are required for the formation and identity of the pulmonary myocardium. *Circ Res* 101(9):902-909.

46. Soufan AT, *et al.* (2004) Reconstruction of the patterns of gene expression in the developing mouse heart reveals an architectural arrangement that facilitates the understanding of atrial malformations and arrhythmias. *Circ Res* 95(12):1207-1215.

47. Huang RT, Xue S, Xu YJ, Zhou M, & Yang YQ (2013) A novel NKX2.5 loss-of-function mutation responsible for familial atrial fibrillation. *Int J Mol Med* 31(5):1119-1126.

48. Xie WH, *et al.* (2013) Prevalence and spectrum of Nkx2.5 mutations associated with idiopathic atrial fibrillation. *Clinics (Sao Paulo)* 68(6):777-784.

49. Qiu XB, *et al.* (2014) PITX2C loss-of-function mutations responsible for idiopathic atrial fibrillation. *Clinics (Sao Paulo)* 69(1):15-22.

50. Wang J, *et al.* (2014) Pitx2-microRNA pathway that delimits sinoatrial node development and inhibits predisposition to atrial fibrillation. *Proc Natl Acad Sci U S A* 111(25):9181-9186.

51. Christoffels VM, *et al.* (2006) Formation of the venous pole of the heart from an Nkx2-5-negative precursor population requires Tbx18. *Circ Res* 98(12):1555-1563.

52. Xie L, *et al.* (2012) Tbx5-hedgehog molecular networks are essential in the second heart field for atrial septation. *Dev Cell* 23(2):280-291.

53. Ye W, *et al.* (2015) A common Shox2-Nkx2-5 antagonistic mechanism primes the pacemaker cell fate in the pulmonary vein myocardium and sinoatrial node. *Development* 142(14):2521-2532.

54. Ventura A, *et al.* (2007) Restoration of p53 function leads to tumour regression *in vivo*. *Nature* 445(7128):661-665.

55. Kim D, *et al.* (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14(4):R36.

56. Trapnell C, Pachter L, & Salzberg SL (2009) TopHat: discovering splice junctions with RNASeq. *Bioinformatics* 25(9):1105-1111.

57. Barnett DW, Garrison EK, Quinlan AR, Stromberg MP, & Marth GT (2011) BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics* 27(12):1691-1692.

58. Pertea M, Kim D, Pertea GM, Leek JT, & Salzberg SL (2016) Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc* 11(9):1650-1667.

59. Pertea M, *et al.* (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* 33(3):290-295.

60. R Core Team (2017) R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).

61. Robinson MD, McCarthy DJ, & Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26(1):139-140.

62. Ritchie ME, *et al.* (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43(7):e47.

63. Wickham H (2009) *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag New York).

64. Wickham H (2017) scales: Scale Functions for Visualization), R package version 0.5.0.

65. Nowak KJ, *et al.* (1999) Mutations in the skeletal muscle alpha-actin gene in patients with actin myopathy and nemaline myopathy. *Nat Genet* 23(2):208-212.

66. Perera D, *et al.* (2007) Bub1 maintains centromeric cohesion by activation of the spindle checkpoint. *Dev Cell* 13(4):566-579.

67. Christodoulou A, Lederer CW, Surrey T, Vernos I, & Santama N (2006) Motor protein KIFC5A interacts with Nubp1 and Nubp2, and is implicated in the regulation of centrosome duplication. *J Cell Sci* 119(Pt 10):2035-2047.

68. Chan GK, Jablonski SA, Starr DA, Goldberg ML, & Yen TJ (2000) Human Zw10 and ROD are mitotic checkpoint proteins that bind to kinetochores. *Nat Cell Biol* 2(12):944-947.

69. Barton PJ, *et al.* (1985) The myosin alkali light chains of mouse ventricular and slow skeletal muscle are indistinguishable and are encoded by the same gene. *J Biol Chem* 260(14):8578-8584.

70. Moldovan GL, *et al.* (2012) Inhibition of homologous recombination by the PCNA-interacting protein PARI. *Mol Cell* 45(1):75-86.

71. Lu X, *et al.* (2011) The cytokine-driven regulation of secretoglobins in normal human upper airway and their expression, particularly that of uteroglobin-related protein 1, in chronic rhinosinusitis. *Respir Res* 12:28.

72. Kitamura T, Takagi S, Naganuma T, & Kihara A (2015) Mouse aldehyde dehydrogenase ALDH3B2 is localized to lipid droplets via two C-terminal tryptophan residues and lipid modification. *Biochem J* 465(1):79-87.

73. Tomsig JL & Creutz CE (2002) Copines: a ubiquitous family of Ca(2+)-dependent phospholipid-binding proteins. *Cell Mol Life Sci* 59(9):1467-1477.

74. Fagerberg L, *et al.* (2014) Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics* 13(2):397-406.

75. Yue F, *et al.* (2014) A comparative encyclopedia of DNA elements in the mouse genome. *Nature* 515(7527):355-364.

76. Yoshida H, *et al.* (2013) KIAA1199, a deafness gene of unknown function, is a new hyaluronan binding protein involved in hyaluronan depolymerization. *Proc Natl Acad Sci U S A* 110(14):5612-5617.

77. Eyre D (2002) Collagen of articular cartilage. *Arthritis Res* 4(1):30-35.

78. Erle DJ, Ruegg C, Sheppard D, & Pytela R (1991) Complete amino acid sequence of an integrin beta subunit (beta 7) identified in leukocytes. *J Biol Chem* 266(17):11009-11016.

79. Pang YL, Poruri K, & Martinis SA (2014) tRNA synthetase: tRNA aminoacylation and beyond. *Wiley Interdiscip Rev RNA* 5(4):461-480.

80. Smith CL, *et al.* (2018) Mouse Genome Database (MGD)-2018: knowledgebase for the laboratory mouse. *Nucleic Acids Res* 46(D1):D836-D842.

81. LopezJimenez N, *et al.* (2010) Examination of FGFRL1 as a candidate gene for diaphragmatic defects at chromosome 4p16.3 shows that Fgfrl1 null mice have reduced expression of Tpm3, sarcomere genes and Lrtm1 in the diaphragm. *Hum Genet* 127(3):325-336.

82. Jisaka M, Kim RB, Boeglin WE, Nanney LB, & Brash AR (1997) Molecular cloning and functional expression of a phorbol ester-inducible 8S-lipoxygenase from mouse skin. *J Biol Chem* 272(39):24410-24416.

83. Lu Z, *et al.* (2016) Molecular Architecture of Contactin-associated Protein-like 2 (CNTNAP2) and Its Interaction with Contactin 2 (CNTN2). *J Biol Chem* 291(46):24133-24147.

84. Stevens DR, *et al.* (2001) Hyperpolarization-activated channels HCN1 and HCN4 mediate responses to sour stimuli. *Nature* 413(6856):631-635.

85. Darzacq X, *et al.* (2002) Cajal body-specific small nuclear RNAs: a novel class of 2'-Omethylation and pseudouridylation guide RNAs. *EMBO J* 21(11):2746-2756.

86. Burgoyne RD (2007) Neuronal calcium sensor proteins: generating diversity in neuronal Ca2+ signalling. *Nat Rev Neurosci* 8(3):182-193.

87. DeChiara TM, Efstratiadis A, & Robertson EJ (1990) A growth-deficiency phenotype in heterozygous mice carrying an insulin-like growth factor II gene disrupted by targeting. *Nature* 345(6270):78-80.

88. Schluter OM, Schmitz F, Jahn R, Rosenmund C, & Sudhof TC (2004) A complete genetic analysis of neuronal Rab3 function. *J Neurosci* 24(29):6629-6637.

89. Steinberg SJ, Wang SJ, McGuinness MC, & Watkins PA (1999) Human liver-specific very-long-chain acyl-coenzyme A synthetase: cDNA cloning and characterization of a second enzymatically active protein. *Mol Genet Metab* 68(1):32-42.

90. Huang W, *et al.* (2017) The long non-coding RNA SNHG3 functions as a competing endogenous RNA to promote malignant development of colorectal cancer. *Oncol Rep* 38(3):1402-1410.

CHAPTER IV: SYNTHESIS, FUTURE DIRECTIONS, AND CONCLUDING REMARKS

Synthesis

Approximately one quarter of infant deaths in the United States are due to congenital heart defects (CHD), which are reported in approximately 1 in 125 live births or roughly 32,000 cases annually [1]. Atrial septal and atrioventricular septal defects (ASDs and AVSDs, respectively) make up one of the most common classes of CHDs [1]. The atrial and atrioventricular septum divides the oxygen-poor right atrium from the oxygen-rich left atrium in the adult [2]. The development of the atrial and atrioventricular septum is a complex process requiring the coordination of multiple tissues and tissue sources. Despite the importance of this structure, we have only begun to understand the molecular processes required for the development of the structure within the last 20 years.

To date, there have been dozens of genes annotated as involved in the formation of the septal structures with the inclusion of many critical cardiogenic transcription factors such as *TBX5*, *NKX2-5*, and *GATA4* [2]. Because of this relationship between the cardiogenic transcription factors and cardiac malformations, the historical view was that atrial and atrioventricular septal defects were thought to be a cardiac problem in origin. However, prior to my work, the field shifted its view on the ontogeny of the atrial septum. A key finding in understanding atrial and atrioventricular septation was the discovery that the second heart field marker, *Isl1*, genetically labels a mesenchymal structure contributing to the mature atrial septum known as the dorsal mesenchymal protrusion (DMP) [3]. The second heart field (SHF) is one of two mesodermally derived tissues that forms the heart and was previously known to give rise to the outflow tract, a portion of the right ventricle, and the atria, but was not known to be involved

in atrial septation [2, 4]. This tissue has also been described as the cardiopulmonary progenitors (CPPs), and gives rise to the mesenchymal portions of the lungs and pulmonary vein [5]. In essence, rather than a problem arising from within the cardiac chambers, atrial and atrioventricular septation is firmly rooted developmentally in aspects of cardiopulmonary development with the CPPs and second heart field.

Prior to my work, our laboratory has contributed to this paradigm shift in understanding the ontogeny of atrial and atrioventricular septation [6-10]. *Shh* expressed in the pulmonary endoderm targets the CPPs, SHF, and the DMP, shown by *Gli1* activation [6] and in its the absence, results in septal defects [6, 11]. More recently, our lab implicated *Tbx5* in DMP formation specifically [9]. Of particular interest, that work demonstrated *Tbx5* conditional KO in *Hh* receiving cells results in a similar defect as the heterozygous *Tbx5* condition, and *Tbx5* and *Gli1* co-regulate a set of genes required in atrial septation, including *Osr1* and *Foxf1* [10, 12], suggesting an interactive role of *Hh* signaling with *Tbx5* [9].

Through the novel work I present in my dissertation, I have expanded upon the molecular underpinning of TBX5 in the SHF/CPPs for the formation of the atrial and atrioventricular septum. My work not only demonstrates that TBX5 is required for the initiation of lung development, but that the cardiac morphogenesis events of the SHF are dependent on pulmonary development (Chapter II).

Future Directions

The work I have presented here has sparked innovation and hypothesis-driven investigation into two new avenues of interest. The first is a direct follow-up to Chapter II, in which we are working to identify the molecular relationship between *Tbx5*, expressed in the

second heart field and cardiopulmonary progenitors, and *Shh*, expressed in the pulmonary endoderm. I believe this work will close the loop between our lab's previously published work, in which we have presented an epistatic and cooperative relationship between *Tbx5* and *Shh* in the formation of the atrial septum [9-10], with the work presented in Chapter II uncovering the early role of *Tbx5* in lung development for cardiac morphogenesis [13]. In our second new direction, we took the lead from our previous work in the adult atria [14] and Chapter III, and have asked, outside of the developmental context, what is the role of *Tbx5* at homeostasis in the adult lung? This second direction may help address a fundamental, yet unappreciated, role of *TBX5* in the adult lung that has been recently hinted at through genome-wide association studies [15-17]. Although I may not be able to see these either of these works to completion, I believe the work presented in this dissertation has laid the ground-work for exciting future interrogation of *TBX5* in both the heart and lungs.

Tbx5 *maintains Retinoic acid signaling for* Shh *expression during cardiopulmonary development*

As mentioned above, the first area of investigation is a direct follow-up to Chapter II [13]. *Tbx5*, expressed in the cardiopulmonary progenitors (CPPs), and *Shh*, expressed in the foregut-derived pulmonary endoderm, are known to genetically interact for cardiopulmonary morphogenesis [9-10]. Our recent work has demonstrated that *Tbx5* is required for *Shh* expression [13]; however, the mechanism remains unknown. In on-going work, we have evidence that demonstrates *Tbx5* is required to maintain Retinoic acid (RA) signaling from the CPPs, in part through direct regulation of *Aldh1a2* through highly conserved *cis*-regulatory elements (Figure IV.1). In normal development of the CPPs, *Aldh1a2* is required to activate *Tbx5* expression [18]; however, our work demonstrates that *Aldh1a2*/RA/*Tbx5* work in a feed-forward

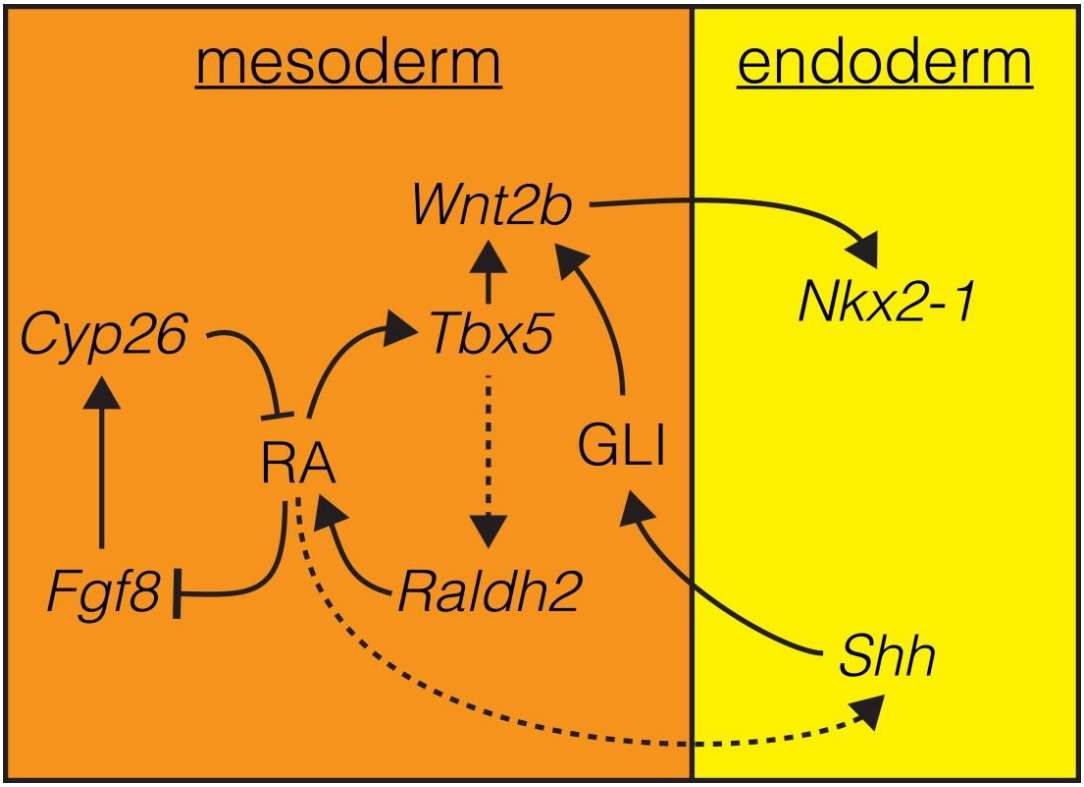


Figure IV.1 *Tbx5* maintains Retinoic acid signaling for *Shh* expression during cardiopulmonary development

Our proposed model for the *Tbx5*-dependence of Shh seen in the pulmonary endoderm relies on the cell non-autonomous interaction between Retinoic Acid (RA) signaling in the cardiopulmonary progenitor mesoderm to the pulmonary endoderm. During normal development, *Aldh1a2*/*Raldh2* expression precedes *Tbx5* expression. *Tbx5* expression drives expression of canonical Wnt signaling (Chapter II), but new evidence demonstrates that TBX5 directly binds and maintains *Aldh1a2*/*Raldh2* expression. RA signaling, normally required for Shh expression in the foregut endoderm prior to Tbx5 expression, becomes depleted in the absence of Tbx5. The depletion of RA is further exacerbated by the antagonistic repression of RA by FGF8/CYP26.

loop to maintain or enhance RA signaling. RA signaling is required for *Shh* expression in the foregut [19], and preliminary work demonstrates ectopic RA treatment can rescue the loss of *Shh* in the absence of *Tbx5*. Furthermore, increased *Fgf8* expression seen in the absence of *Tbx5*, normally repressed by RA signaling in the CPPs, indirectly represses *Shh* via CYP26-mediated metabolism of RA. This indirect effect of FGF8 signaling can be partly rescued by the FGF-receptor inhibitor, SU5402, or the pan-CYP26 inhibitor, ketoconazole. Taken together, our upcoming work supports a model in which *Tbx5* is required to maintain RA for Shh signaling during cardiopulmonary development, and the loss of *Tbx5* results in both decreased production and increased metabolism of RA, which secondarily results in loss of *Shh* expression (Figure IV.1).

Tbx5 *is required in the adult lung to maintain pulmonary vein myocardium gene expression*

As a mentioned previously, recent genome-wide association studies (GWAS) identify *TBX5* in association with adult lung function, including phenotypes associated with fibrosis and airway smooth muscle function [15-17]. As our previous work has demonstrated that adult atrial expression of *Tbx5* is required to maintain atrial rhythm and identity [14], we hypothesize that the maintained expression of *TBX5* in the adult lungs were functionally relevant for lung homeostasis (Figure IV.2A). Preliminary investigation into the expression domain of *Tbx5* in the adult mouse lung identifies localization of TBX5 protein to the nuclei of both presumptive fibroblasts and muscle cells (Figure IV.2B). We investigated the *Tbx5*-dependent transcriptome in the adult lung by performing transcriptional profiling of lungs removed from either $R26^{creERT2/creERT2}$; $Tbx5^{flox/flox}$ (*Tbx5* CKO) over $R26^{creERT2/creERT2}$; $Tbx5^{+/+}$ (*Tbx5* CTRL) mice one week after tamoxifen injection. We identified approximately 650 significantly dysregulated

genes (FDR < 0.01 & |log2FC| > 1) in the absence of *Tbx5* (Figure IV.2C). In order to identify the contribution of cell types to the dysregulated gene expression, we reexamined published scRNA-seq performed in the adult lungs investigating the non-endothelial, non-epithelial, non-immunological portion of the lungs, and identified 10 populations with unique transcriptomic signatures [20]. We overlaid the list of statistically significant genes from the *Tbx5* CKO and CTRL experiment with those genes significantly enriched in each population. Although there was only weak enrichment for upregulated genes in any given population, we did identify a strong enrichment in down regulated genes associated with Cluster 6 (Figure IV.2 E-F). Based on candidate genes and literature searching, Cluster 6 is most likely associated with the muscle cells of the pulmonary venous return of the lungs, which forms a muscular sleeve around the venous return in small mammals and pulmonary vein in all mammals [21-22]. Altogether, our preliminary data suggests that beyond a developmental requirement of *Tbx5* in the lungs, *Tbx5* appears to be required throughout life to maintain the transcriptome of pulmonary vein myocardium [13]. This observation has significant implications for the role *TBX5* may play in diseases such as atrial fibrillation [14].

Concluding Remarks

The T-box transcription factor, *TBX5*, has long held an important role in the formation of the heart (Chapter I). Through the work I have presented (Chapter II and IV), the role of TBX5 has been expanded to include the formation of both the cardiac and pulmonary domains with implications for adult disease (Chapter III and IV). Critically, my work has highlighted the interconnectedness between the heart and the lungs, not just physiologically, but through the *TBX5*-dependent gene regulatory networks throughout development and adult life. Moreover, my

work has suggested that the evolution of air-breathing and septated cardiac chambers utilized this shared *TBX5*-dependent regulatory network during tetrapod evolution with expanded implications in the evolution of new body structures and functions.

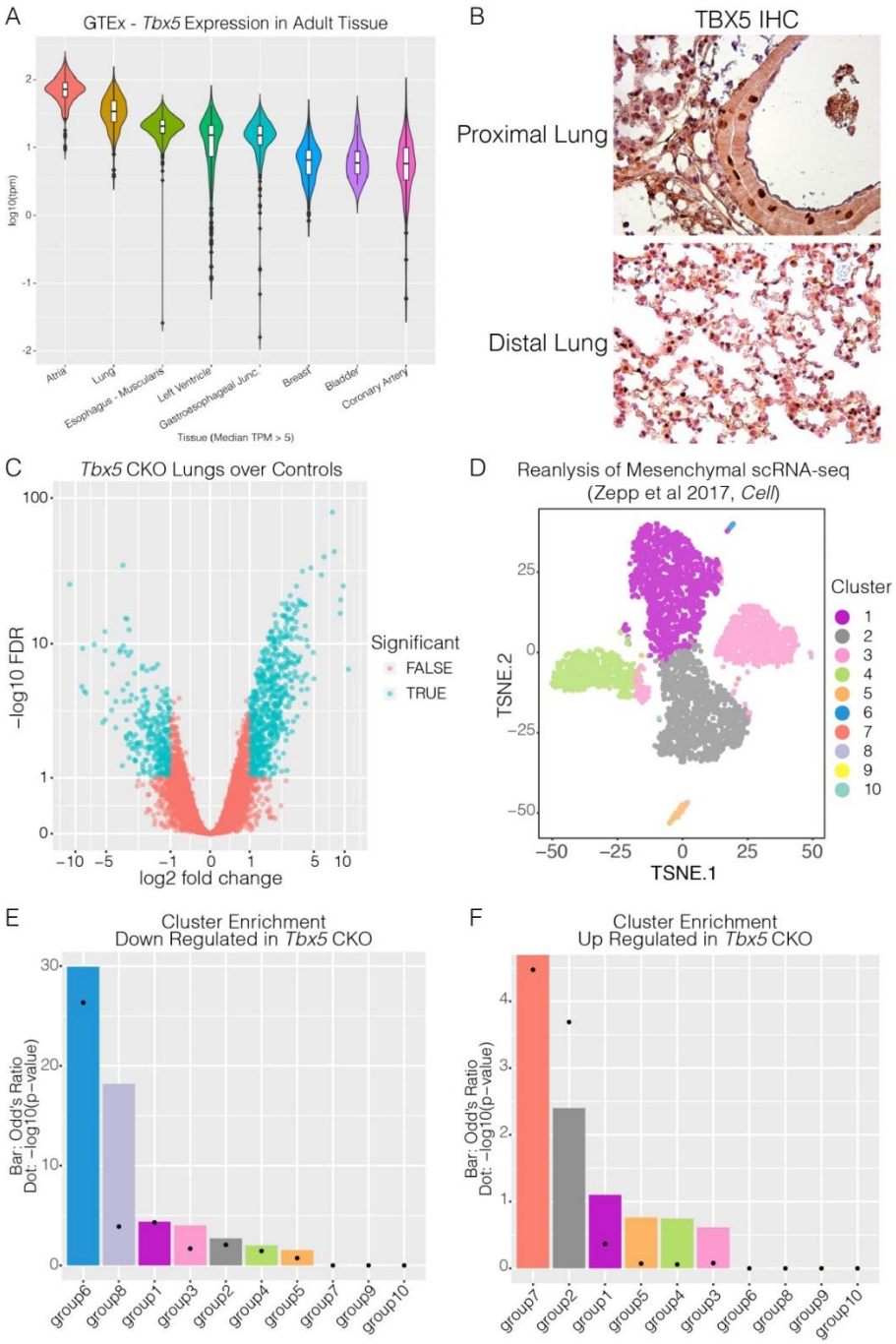


Figure IV.2 Tbx5 is required in the adult lung to maintain pulmonary smooth muscle expression

A. Expression of *TBX5* across all tissues from the GTEx consortium with a transcript per million (TPM) greater than 5. B. Immunohistochemistry (IHC) using anti-TBX5 in the proximal and

(Figure IV.2 continued) distal portions of the lung. C. Volcano plot showing the log2 fold change of R26$^{creERT2/creERT2}$; *Tbx5*$^{flox/flox}$ (*Tbx5* CKO) over *R26*$^{creERT2/creERT2}$; *Tbx5*$^{+/+}$ from whole lung transcriptional profiling from adult mice. Those genes which were statistically significant (FDR < 0.01 & |log2FC| > 1) are highlighted in blue. D. TSNE plot showing 10 statistically identified populations from a reanalysis of scRNA-seq from Zepp et al 2017 (*Cell*) of non-epithelial, non-endothelial, and non-immunological populations isolated from the adult mouse lung. E-F. Statistically significant downregulated genes (E) and upregulated genes (F) from the whole lung RNA-seq were associated with cluster specific gene expression from the scRNA-seq experiment.

References

1. Go, A.S., et al., Heart Disease & Stroke Statistics. Circulation, 2013. 127: p. e6-e245.

2. Briggs, L.E., J. Kakarla and A. Wessels, The pathogenesis of atrial and atrioventricular septal defects with special emphasis on the role of the dorsal mesenchymal protrusion. Differentiation, 2012. 84(1): p. 117-130.

3. Snarr, B.S., et al., Isl1 expression at the venous pole identifies a novel role for the second heart field in cardiac development. Circ Res, 2007. 101(10): p. 971-974.

4. Abu-Issa, R. and M. L. Kirby, Heart field: from mesoderm to heart tube. Annu Rev Cell Dev Biol, 2007. 23: p. 45-68.

5. Peng T, et al. (2013) Coordination of heart and lung co-development by a multipotent cardiopulmonary progenitor. *Nature* 500(7464):589-592.

6. Hoffmann, A.D., et al., sonic hedgehog is required in pulmonary endoderm for atrial septation. Development, 2009. 136(10): p. 1761-70.

7. Kamp, A., et al., Genome-wide identification of mouse congenital heart disease loci. Hum Mol Genet, 2010. 19(16): p. 3105-3113.

8. Friedland-Little, J.M., et al., A novel murine allele of Intraflagellar Transport Protein 172 causes a syndrome including VACTERL-like features with hydrocephalus. Hum Mol Genet, 2011. 20(19): p. 3725-3737.

9. Xie, L., et al., Tbx5-hedgehog molecular networks are essential in the second heart field for atrial septation. Developmental Cell, 2012. 23(2): p. 280-91.

10. Hoffmann A.D., et al. (2014) Foxf genes integrate tbx5 and hedgehog pathways in the second heart field for cardiac septation. PLoS Genet 10:e1004604.

11. Goddeeris, M.M., et al., Intracardiac septation requires hedgehog-dependent cellular contributions from outside the heart. Development, 2008. 135(10): p. 1887-95.

12. Wang, Q., et al., Odd-skipped related 1 (Odd 1) is an essential regulator of heart and urogenital development. Dev Biol, 2005. 288(2): p. 582-594.

13. Steimle JD, *et al.* (2018) Evolutionarily conserved Tbx5-Wnt2/2b pathway orchestrates cardiopulmonary development. *Proc Natl Acad Sci U S A* 115(45):E10615-E10624.

14. Nadadur RD, *et al.* (2016) Pitx2 modulates a Tbx5-dependent gene regulatory network to maintain atrial rhythm. *Sci Transl Med* 8(354):354ra115.

15. Wain, L.V., et al., Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. Lancet Respir Med, 2015. 3(10): p. 769–781.

16. Shrine, N., et al., New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. Nature Genetics, 2019. 51: p. 481-493.

17. Daya, M., et al., Whole-Genome Admixture Mapping Reveals Novel Loci for Bronchodilator Response in African Americans from the Severe Asthma Research Program. Presented at ATS 2018 in A101. Mechanisms and Clinical Features of Severe Asthma, San Diego, 2018.

18. Niederreither, K., et al., Embryonic retinoic acid synthesis is essential for heart morphogenesis in the mouse. Development, 2001. 128(7): p. 1019-1031.

19. Rankin, S.A., et al., A retinoic acid-hedgehog cascade coordinates mesoderm-inducing signals and endoderm competence during lung specification. Cell Rep, 2016. 16(1): p. 66-78.

20. Zepp, J.A., et al., Distinct Mesenchymal Lineages and Niches Promote Epithelial Self-Renewal and Myofibrogenesis in the Lung. Cell, 2017. 170(6): p. 1134-1148.

21. Townsley, M.I., Structure and composition of pulmonary arteries, capillaries and veins. Compr Physiol, 2012. 2: 675-709.

22. Boutilier, J.K., et al., Gene Expression Networks in the Murine Pulmonary Myocardium Provide Insight into the Pathobiology of Atrial Fibrillation. *G3 (Bethesda)*, 2017. 7(9):2999-3017.

APPENDIX: ADDITIONAL WORK CONTRIBUTED

The following appendix lists published or near-published data to which I have contributed significant efforts and time, related to cardiac septation and/or *TBX5* function, but do not directly belong within the greater narrative of my dissertation. Figure A.1, which contributed to Hoffmann et al 2014 (*PLOS Genetics*), demonstrates the high degree of overlap between the Forkhead family of transcription factor, *Shh*-dependent transcription in the second heart field, and GLI-binding. Figure A.2 and Figure A.3 contributed to Waldron et al 2016 (*Developmental Cell*). These two figures suggested that TBX5 may interact with the transcriptional repressive NuRD complex to repress non-cardiac fates in the developing heart. Figure A.4 contributed to Burnicka-Turek et al 2016 (*Human Molecular Genetics*), and helped distinguish a nuanced molecular role of cilia proteins in the background of heterotaxy and congenital heart defects. Figure A.5 contributed to Bersell et al (under review at the time of writing). This work provided *in vivo* evidence that *Tbx5* regulates PDGF-receptor expression, and in its absence, provides a substrate for arrhythmogenic ventricular cardiomyocytes in mice. Figures A.6-8 contributed to Broman et al (in preparation at the time of writing), and together, demonstrate TBX5, GATA4, and FOG2 work in a shared gene regulatory network at common *cis*-regulatory elements in the adult atrial cardiomyocyte, and the misbalance of either TBX5 or FOG2 results in aberrant transcriptional activity of the associated coding genes. Finally, Figures A.9-10 contributed to Burnicka-Turek et al (under review at time the time of writing). Together, these figures demonstrate that TBX5/TBX3 directly act upon *cis*-regulatory elements associated with genes of the fast conduction system of the heart with TBX5 driving positive activity and TBX3 driving repressive activity.

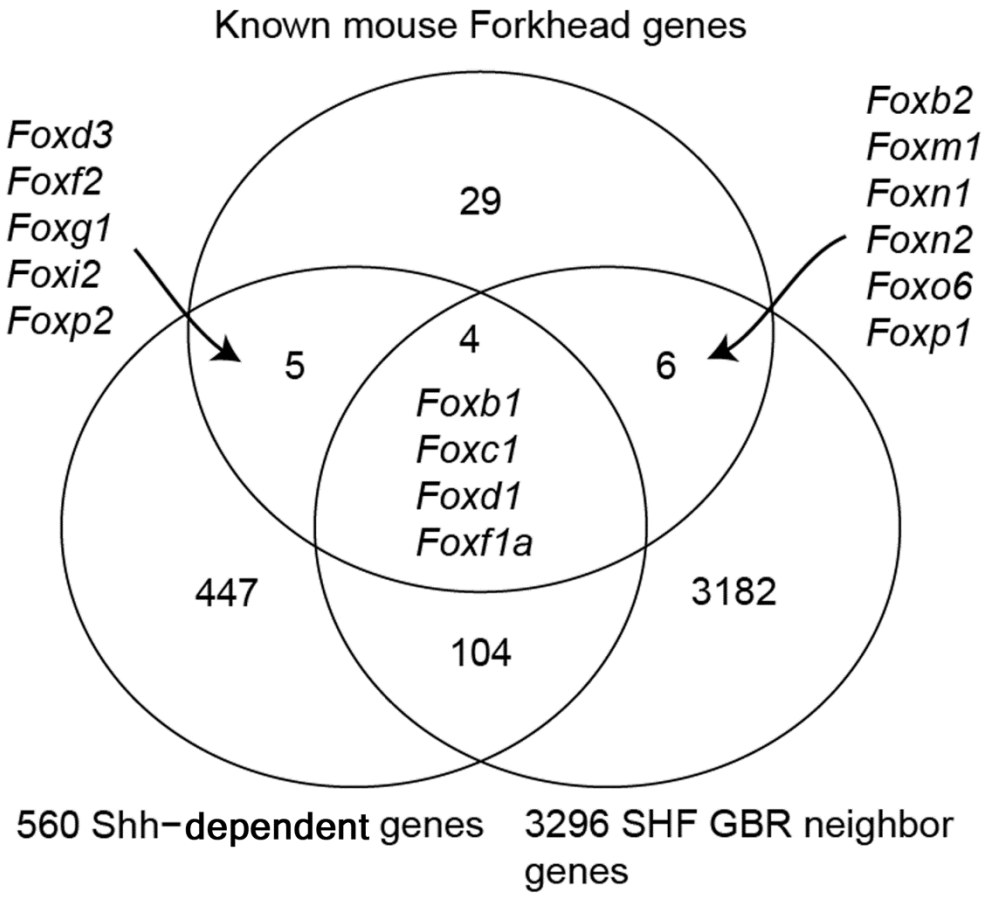
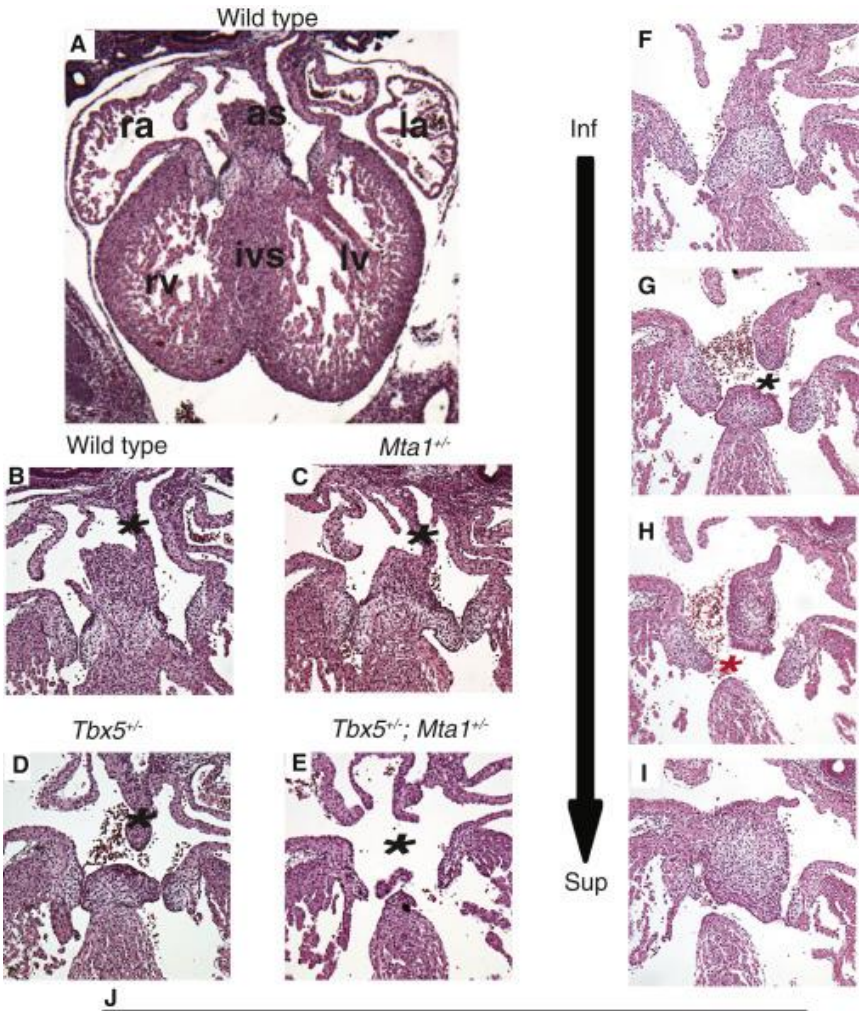


Figure A.1. Known mouse Forkhead-box genes are enriched among the identified *shh*-dependent and Gli3T-bound genes



J

	Wild type	*Mta1*$^{+/-}$	*Tbx5*$^{+/-}$	*Tbx5*$^{+/-}$; *Mta1*$^{+/-}$
Normal	12	14	4	0
ASD	0	0	1	0
VSD	0	0	3	8
CCAVC	0	0	8	11
Total	12	14	16	19
Fisher's exact test (vs wild type)		1	0.0002549	7.09E-9
Fisher's exact test (vs *Tbx5*$^{+/-}$)				0.03549

Figure A.2. TBX5 and the NuRD Complex Genetically Interact

(A–E) Histology (H&E) of embryonic hearts in transverse section at E13.5. (A) Low magnification of a wild-type heart. High magnification of (B) wild-type, (C) *Mta1*$^{+/-}$, (D) *Tbx5*$^{+/-}$, and, (E) *Tbx5*$^{+/-}$; *Mta1*$^{+/}$hearts. *Tbx5*$^{+/-}$; *Mta1*$^{+/}$hearts exhibited cardiac septal defects including ASD, VSD, and CCAVC (asterisk). (F–I) Histology at E13.5 through a single *Tbx5*$^{+/-}$;

(Figure A.2 continued) *Mta1*$^{+/-}$ heart from inferior to superior revealing ASD (G) (black asterisk) and VSD (red asterisk) (H and I) components of CCAVC. (J) Quantification of cardiac defects observed in each genotype. Fisher's exact test showed that *Tbx5*$^{+/-}$; *Mta1*$^{+/-}$ mice demonstrated significantly more septal defects than either wild-type or *Tbx5*$^{+/-}$ mice. as, atrial septum; ivs, interventricular septum; la, left atrium; lv, left ventricle; ra, right atrium; rv, right ventricle.

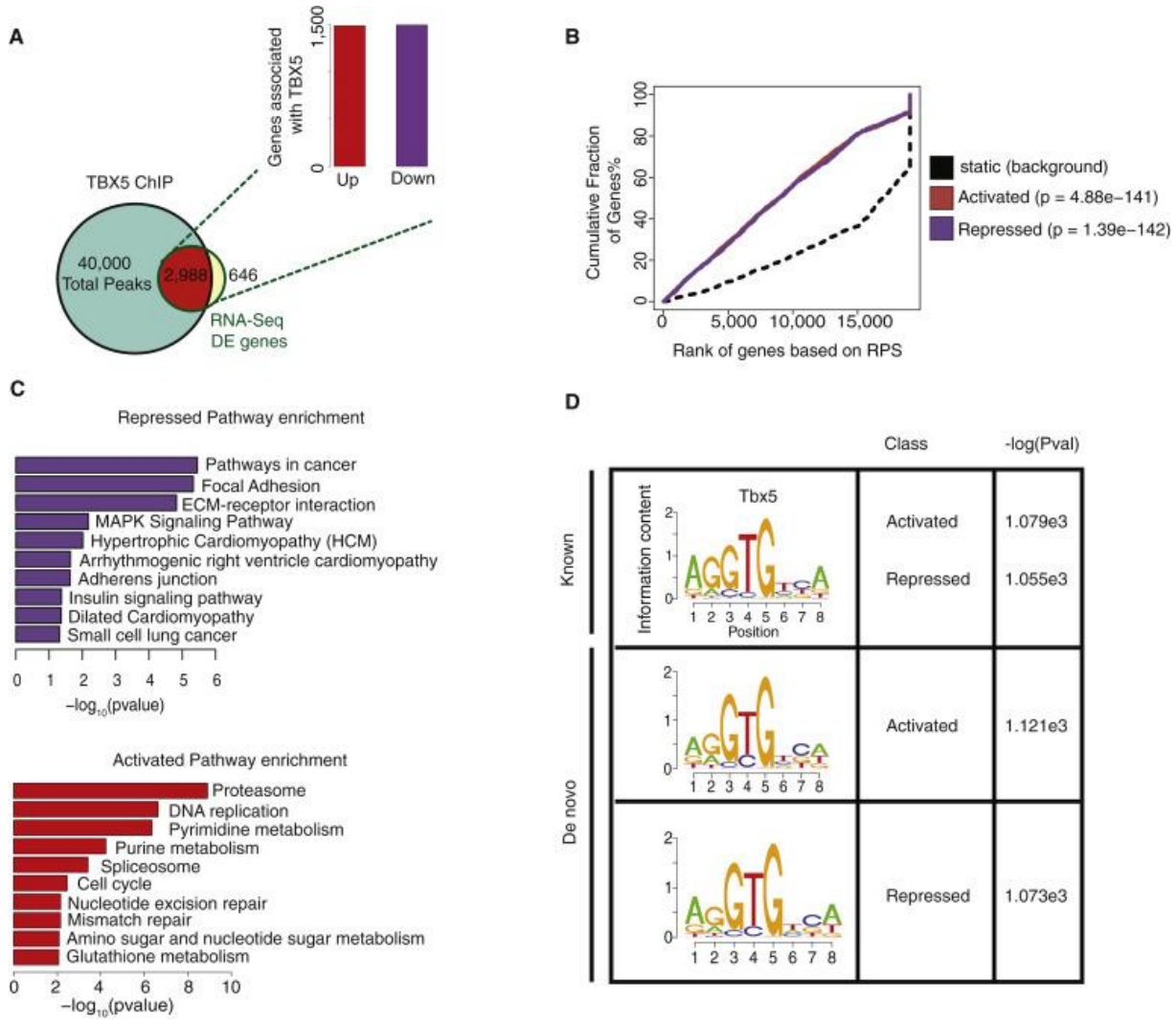


Figure A.3. Analysis of TBX5 Binding Motifs in Activated versus Repressed Genes

(A) Schematic overlay between TBX5 peaks (He et al., 2011) and differentially expressed genes between wild-type and *Tbx5* null heart tissue. 2,988 genes are differentially expressed and associated with a TBX5 peak. Inset shows number of up- and downregulated genes. (B) Activating/repressive function prediction of TBX5 peaks. Red and purple lines represent up-and downregulated genes. Black dashed line represents background (static genes). Genes are ranked from high to low based on their regulatory potential score (RPS) determined by BETA. p values

(Figure A.3 continued) represent significance of up- or downregulated group relative to background (static genes) by Kolmogorov-Smirnov test. (C) KEGG pathway enrichment of downregulated genes associated with TBX5 binding ranked by $-\log_{10}$(p value). (D) TBX5 consensus motif (top) and top motif from de novo identification (bottom), and its presence at ChIP-seq peaks associated with activated or repressed genes.

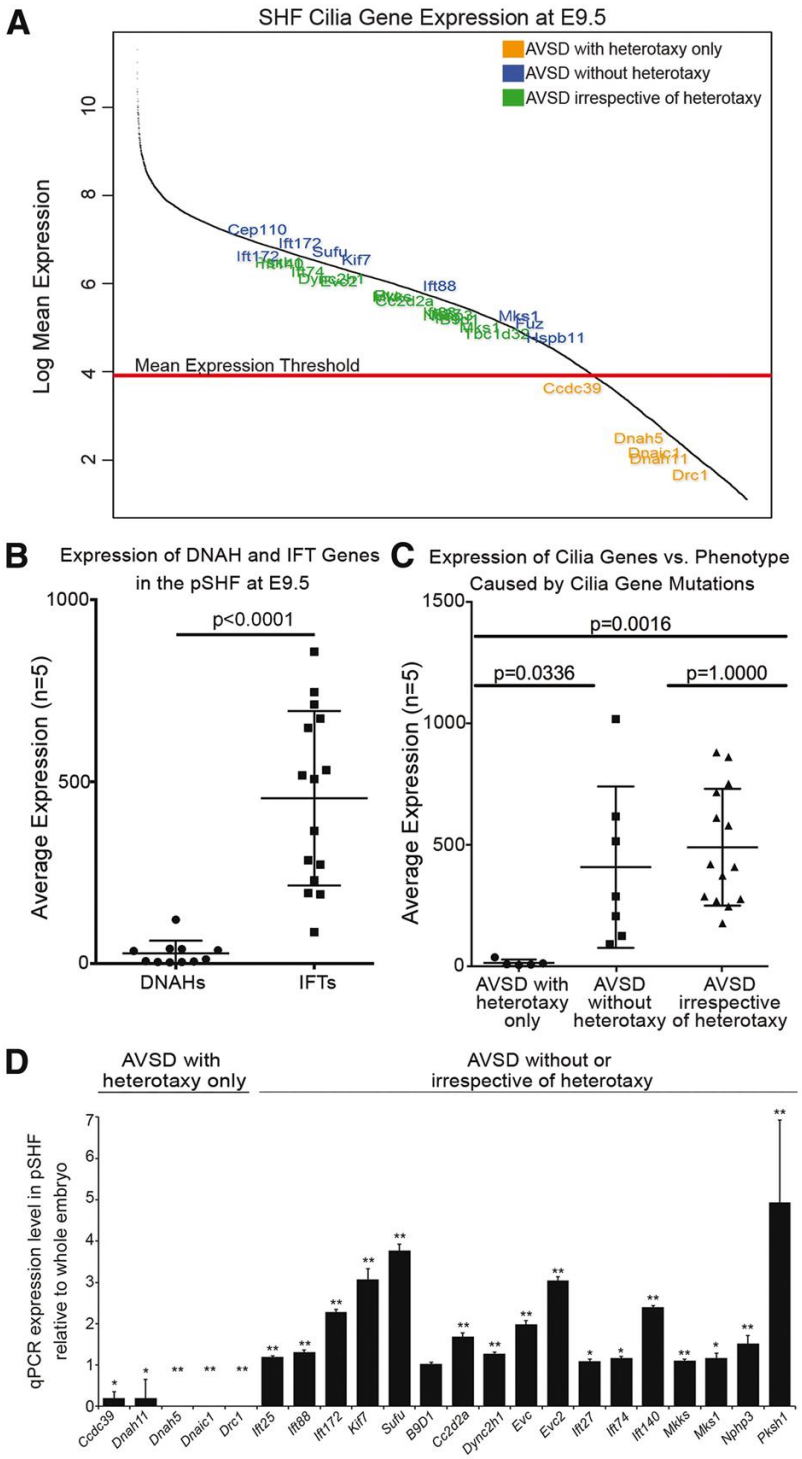


Figure A.4. Cilia structural genes and cilia signaling genes, but not cilia motility genes, are expressed in the SHF

(Figure A.4 continued) (A) All genes detected by RNA-seq plotted by log (mean expression), with genes from Tables 2 and 3 labeled by phenotypic classes (orange, AVSD with heterotaxy only; blue, AVSD without heterotaxy; green, AVSD irrespective of heterotaxy). Genes below a normalized mean expression n of 50 fall within low-level transcriptional noise. (B) Graph of Dnah and Ift cilia genes expressed by respective normalized FPKM. (C) Graph of all cilia genes expression by phenotypic class with respective normalized FPKM. (D) qPCR expression analysis of 22 cilia genes associated with AVSD in mice and humans (Tables 2 and 3) was performed on mRNA isolated from E9.5 wild-type whole mouse embryo and pSHF. Low expression of cilia genes whose mutations caused AVSD with heterotaxy only was observed in contrast to high expression of all of the genes whose mutations caused AVSD without heterotaxy or AVSD irrespective of heterotaxy. qPCR data are mean ± SD, n = 3, *$P < 0.01$, **$P < 0.001$, Student's t test.

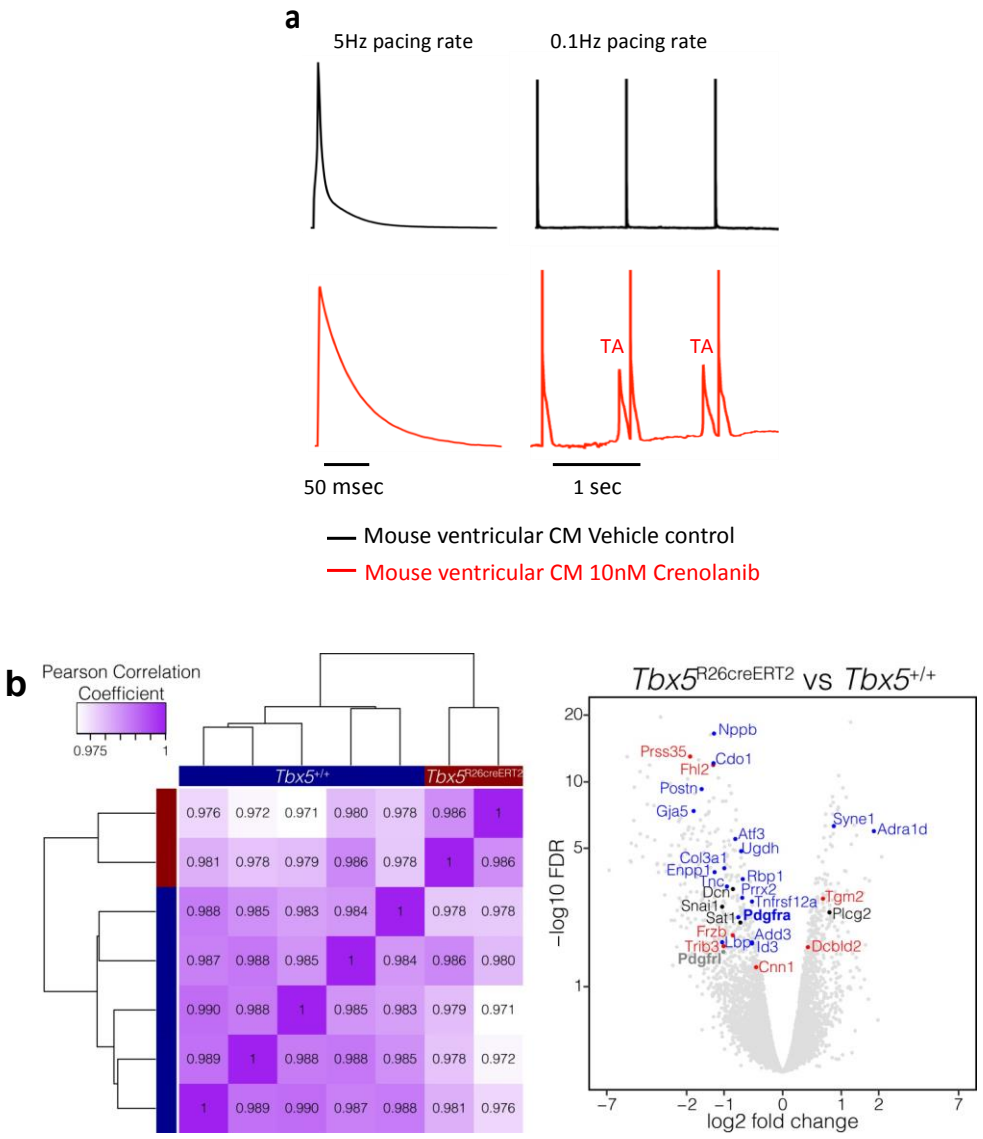
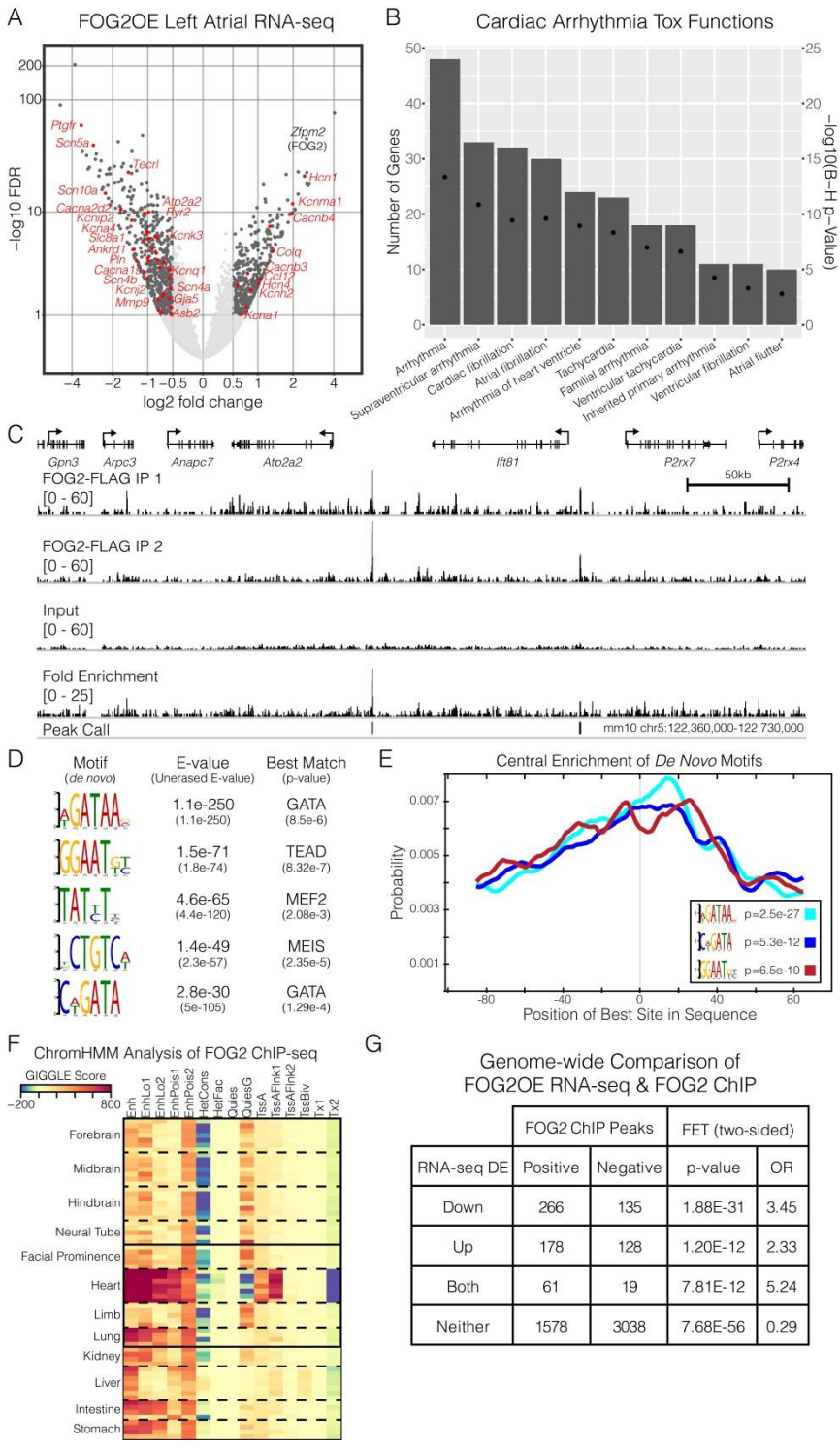


Figure A.5. TBX PDGF phenotypes conserved in mouse cardiomyocytes

(a) Representative mouse ventricular cardiomyocyte (CM) action potentials recorded at pacing rate of 5 Hz and (e) 0.1Hz after 2 hour treatment with vehicle or 100 nM crenolanib. Triggered action potentials (TA) were observed in Crenolanib treated cells paced at 0.1Hz (b) Pearson correlation and hierarchical clustering of biological replicates for the transcriptional profiling of mouse embryonic left ventricle 24-hours after conditional knockout in *Tbx5*$^{flox/flox}$ and controls

(Figure A.5 continued) (left). Differential expression testing of transcriptional profiles with *Pdgfra* and *Pdgfrl* (bold) and members of the PDGF-BB signaling axis that predict inhibition (blue), activation (red), or affected (black) by conditional knockout of *Tbx5* (right).



	FOG2 ChIP Peaks		FET (two-sided)	
RNA-seq DE	Positive	Negative	p-value	OR
Down	266	135	1.88E-31	3.45
Up	178	128	1.20E-12	2.33
Both	61	19	7.81E-12	5.24
Neither	1578	3038	7.68E-56	0.29

Figure A.6. Transcriptional profiling and ChIP define a FOG2 dependent regulatory network

(Figure A.6 continued) A. Volcano plot of relative transcript expression from the left atria of NTG (control) vs DTG (FOG2-overexpressing) mice by RNA-seq. All significantly dysregulated genes ($|\log_2 FC| > 0.5$, FDR < 0.1) are labeled in dark grey, and all nonsignificant transcripts are light grey. Red dots, and a subset of gene names in red, represent those significantly dysregulated genes comprising the term "cardiac arrhythmia," the most significant toxicological (Tox) function (Benjamini–Hochberg corrected p-value: 4.31E-14 to 2.47E-1) by Ingenuity Pathway Analysis (IPA). B. All Tox function terms containing ≥10 genes comprising "cardiac arrhythmia". The bars represent the number of differentially expressed genes for the corresponding Tox function, while the black dots represent the Benjamini-Hochberg corrected p-value. C. A representative genomic view at the *Atp2a2* locus (mm10 chr5: 122,360,000-122,730,000) aligned with the FOG2 ChIP-seq tracks. Both replicates of FOG2-FLAG ChIP are shown (top two tracks) compared to input (third track). Peak calls and fold enrichment found from pooling both FOG2-flag pulldowns against input (bottom tracks). D. Motif analysis showing enriched *de novo* motifs of the FOG2 ChIP peaks. A position weight matrix (PWM) is shown with significance scores, and the protein family that best matches the corresponding binding motif. E. Distribution of centrally enriched *de novo* motifs, centered on FOG2 ChIP-seq summits. Curves represent probability of motif density against distance in base pairs from center. F. Comparison of FOG2 ChIP-seq with mouse ENCODE ChromHMM designations using a GIGGLE score to characterize enrichment across tissue types for 15 genomic states. A higher score indicates more enrichment. G. Published TADs and inter-TADs were categorized by differentially expressed genes and FOG2 ChIP peaks. Fisher's exact test (FET) p-values and odd's ratios (OR) are shown.

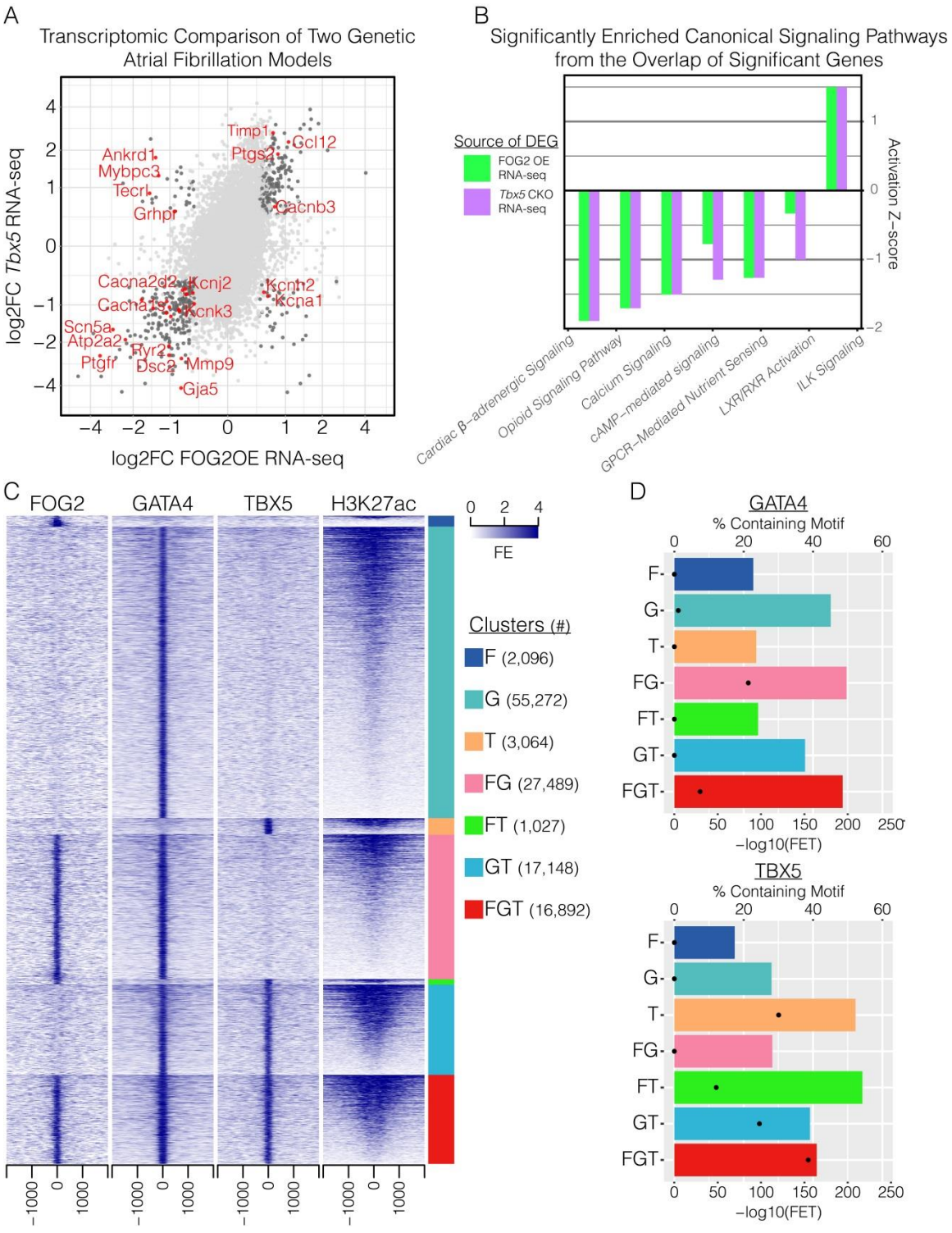


Figure A.7. FOG2 and TBX5 regulate a shared GRN through common GATA4 cis-regulatory elements

(Figure A.7 continued) A. Comparison of FOG2 overexpression (FOG2OE) and *Tbx5* conditional knockout (*Tbx5* CKO) RNA-seq datasets. Dark grey dots represent genes significant in both the FOG2OE ($|\log_2 FC| > 0.5$, FDR < 0.1) and *Tbx5* CKO ($|\log_2 FC| > 0.5$, FDR < 0.05) datasets, while red dots, and a subset of gene names in red, represent those significantly dysregulated genes comprising the term "cardiac arrhythmia" by Ingenuity Pathway Analysis (IPA). B. Activation Z-score, or a prediction on whether a given pathway is activated (positive score) or repressed (negative score), was plotted for the IPA canonical signaling pathways significantly enriched in the FOG2OE and *Tbx5* CKO shared dysregulated genes. C. Heatmap of positional fold enrichment (FE) for a given ChIP-seq dataset ±2000bp of the union of 122,988 summits from FOG2, GATA4, and TBX5 ChIP-seq experiments, clustered by whether they contained signature of FOG2 (F), GATA4 (G), and/or TBX5 (T) and ranked by H3K27ac signal. D. The percentage of peaks (colored bars) for a given F, G, and/or T cluster that contain the published GATA4 (top) or TBX5 (bottom) binding motifs. Dots represent the Fisher's exact test P-value for enrichment.

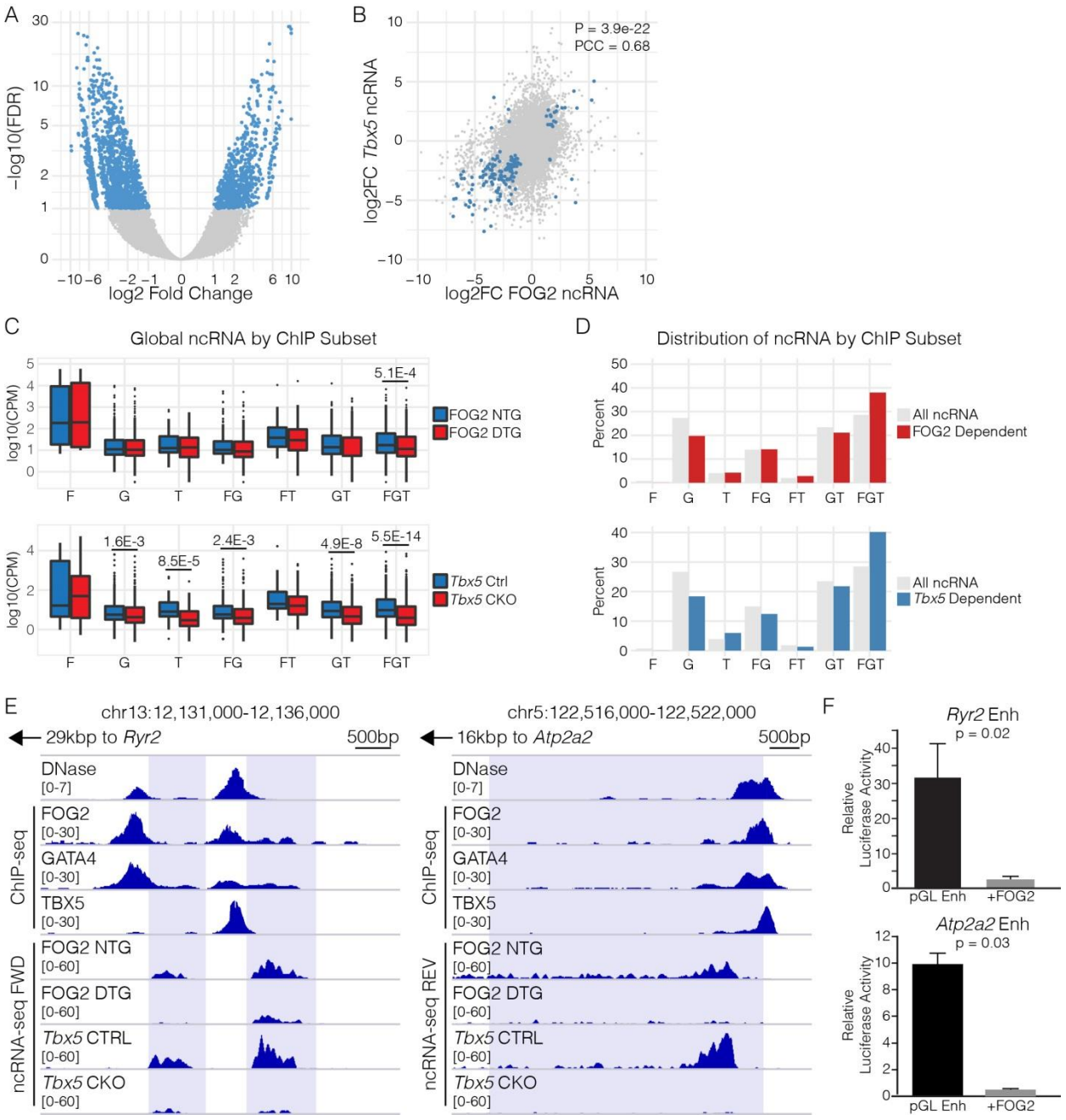


Figure A.8. FOG2-dependent ncRNAs identify shared regulatory elements

A. Volcano plot of relative ncRNA transcript expression from the left atria of NTG (control) vs DTG (FOG2-overexpression) mice by ncRNA-seq. All significantly dysregulated ncRNAs

(Figure A.8 continued) ($|\log_2 FC| > 1$, FDR < 0.1) are labeled in blue, and all nonsignificant transcripts are light grey. B. Comparison of FOG2 overexpression (FOG2OE) and *Tbx5* conditional knockout (*Tbx5* CKO) ncRNA-seq datasets. C. ChIP datasets were overlapped with DNAse hypersensitivity for the following bins: F (FOG only), G (GATA only), T (TBX5 only), FG (FOG and GATA4), FT (FOG and TBX5), GT (GATA4 and TBX5), and FGT (FOG, GATA4, and TBX5). Global ncRNA were then subset into the different bins based on their transcription factor occupancy. D. Comparison of transcription factor occupancy for all ncRNAs and transcription factor dependent ncRNAs. Transcription factor occupancy based on ChIP-seq of FOG2, GATA4, and TBX5 was overlapped with DNAse hypersensitivity for each bin. Transcription factor dependent ncRNAs are defined as ncRNAs within 100kbp of a FOG2 or TBX5 dysregulated gene. E. Genomic view of the *Ryr2* (left, mm10, chr13:12,100,000-12,138,000) and *Atp2a2 (*right, mm10, chr5:122,500,000-122,522,000*)* locus. Tracks show DNAse hypersensitivity (GSE51341), FOG2 ChIP, GATA4 ChIP (He et al 2014), TBX5 ChIP, NTG (control)/DTG (FOG2 overexpression) forward and reverse ncRNA-seq, and wild-type/*Tbx5* CKO forward and reverse ncRNA-seq. The ncRNA transcript is marked with a blue box. F. Relative luciferase activity in HL-1 cardiomyocytes of candidate *Ryr2* and *Atp2a2* enhancers.

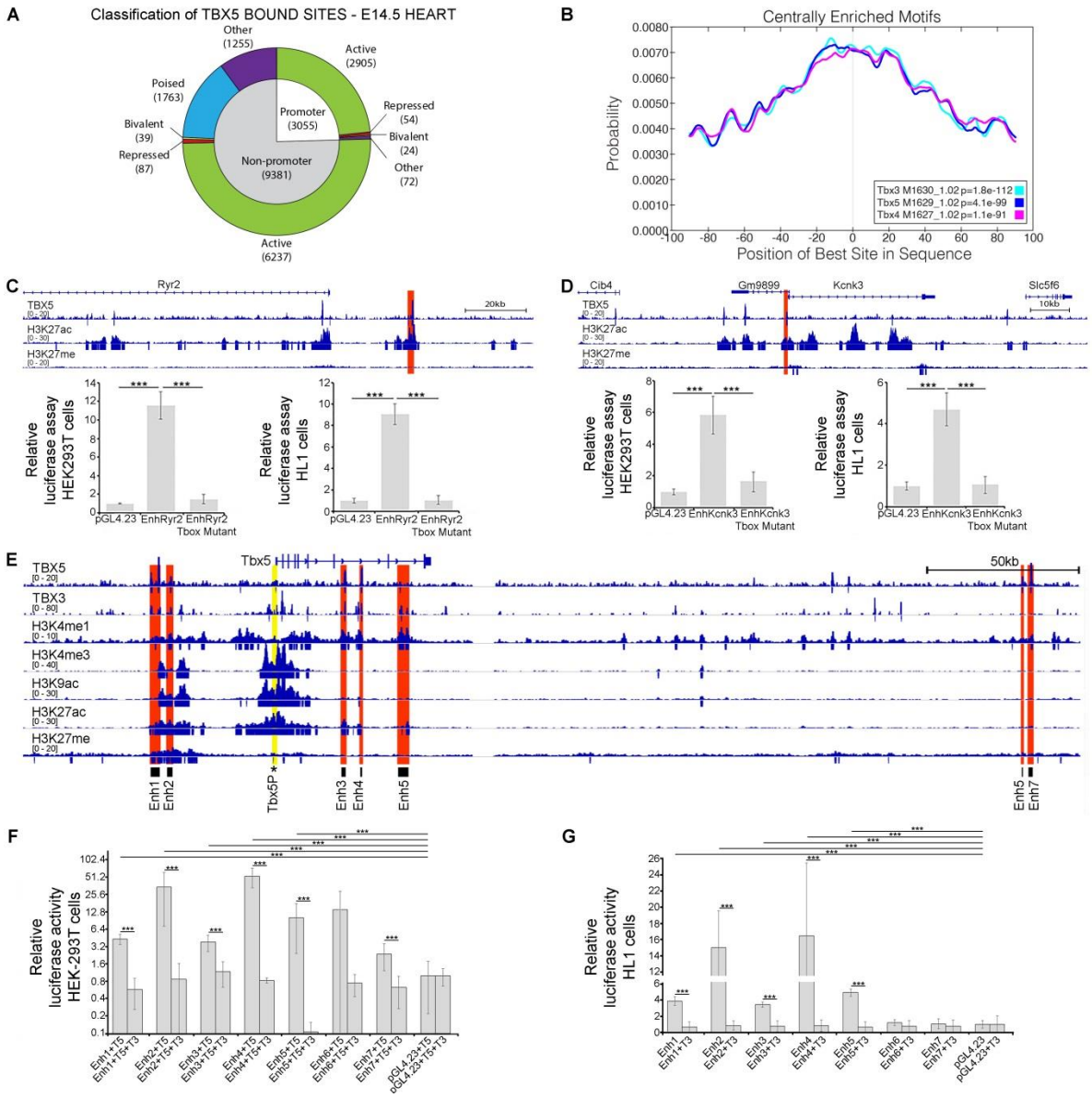


Figure A.9. ***In vivo*** **cardiac localization of TBX5**

A. Classification of the 13,036 peaks identified by TBX5 ChIP-seq from embryonic (E14.5) whole heart based on distance to nearest TSS and relevant histone data. B. Central enrichment plots for the TBX5 ChIP-seq peaks. C-D. Genome browser views for the *Ryr2* (C) and *Kcnk3* (D) loci showing fold-enrichment tracks for TBX5, H3K27ac, and H3K27me3 ChIP-seq (top). Candidate cis-regulatory elements (highlighted red) were cloned, and mutant variants were

(Figure A.9 continued) generated with mutated T-box binding motifs. Luciferase assays were performed in HEK-293T cells (bottom left) provided pcDNA-*Tbx5* expression constructs and HL-1 cardiomyocytes (bottom right). E. Genome browser view for the *Tbx5* locus showing fold-enrichment tracks for TBX5, TBX3, H3K4me1, H3K4me3, H3K9ac, H3K28ac, and H3K27me3. 7 candidate enhancers (highlighted red) were cloned. (F-G) Luciferase assays were performed in HEK-293T cells (F) and HL-1 cells (G) for each of the 7 candidate enhancers at the *Tbx5* locus. In HEK-293T cells, each enhancer was tested for response to pcDNA-*Tbx5* and combinatorial response to pcDNA-*Tbx5* and pcDNA-*Tbx3*. In HL-1 cells, the enhancers were tested in untreated cells or in cells provided pcDNA-*Tbx3*.

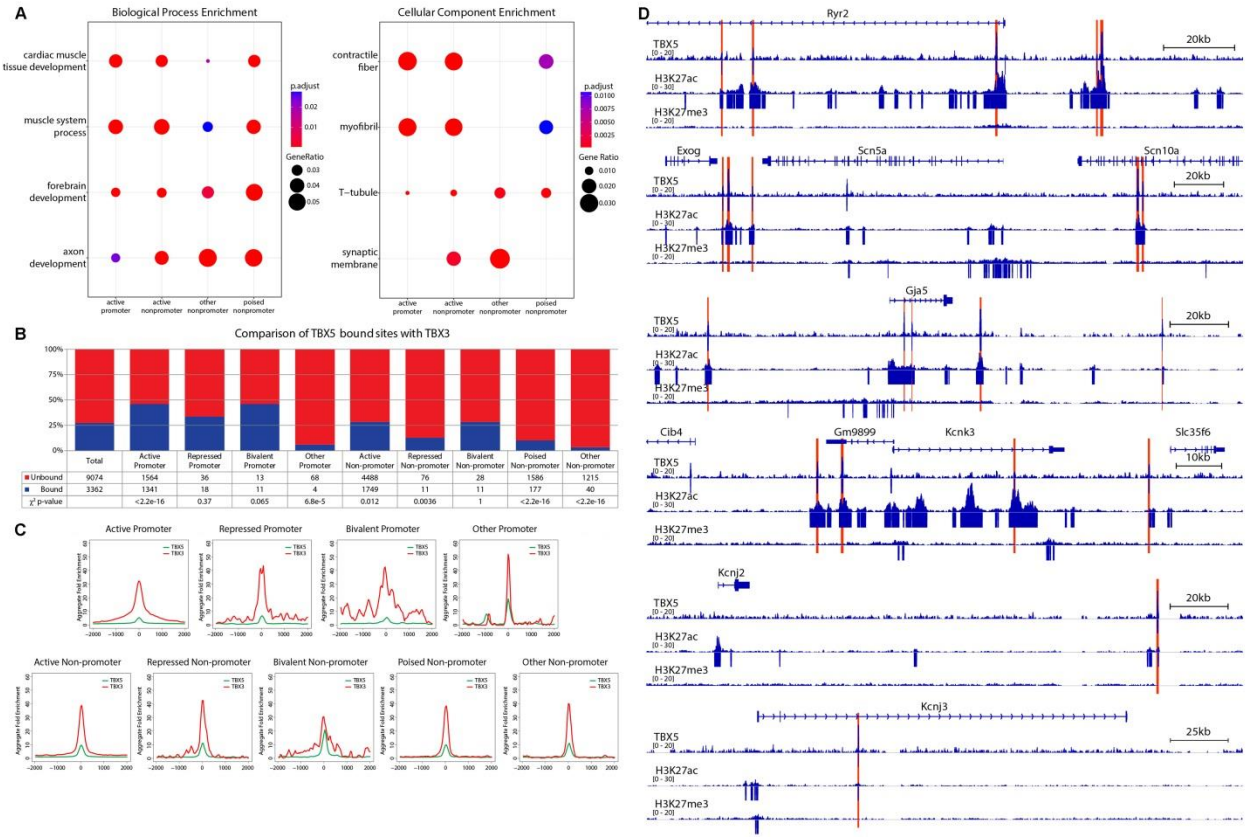


Figure A.10. TBX5 ChIP-seq details and additional genome browser views

A. Top GO terms (Biological Process, left; Cellular Component, right) for TBX5 ChIP-seq associated genes for the active promoter, active nonpromoter, other nonpromoter, and poised nonpromoter sites. B. TBX5 peaks were broken down into sites also identified by TBX3 ChIP-seq (blue) and not found by TBX3 ChIP-seq (red) for each classification of TBX5-bound loci. C. Aggregate fold enrichment for TBX5 (green) and TBX3 (red) for the +/- 2000bp surrounding TBX3-bound TBX5 loci based on prior classification of TBX5-bound loci. D. Genome browser views for the fast conduction genes (*Ryr2*, *Scn5a*, *Gja5*, *Kcnk3*, *Kcnj2*, and *Kcnj3*) associated with an active nonpromoter TBX5 peak (highlighted in red).