CHD4 AND THE NURD COMPLEX ORCHESTRATE A TRANSCRIPTIONAL NETWORK TO CONTROL CARDIAC SARCOMERE FORMATION

Caralynn M. Wilczewski

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor in Philosophy in the Curriculum in Genetics and Molecular Biology.

Chapel Hill
2018

Approved by:

Frank L. Conlon

Li Qian

Brian Strahl

Paul A. Wade

Scott Williams

ABSTRACT

Caralynn M. Wilczewski: CHD4 and the NuRD complex orchestrate a transcriptional network to control cardiac sarcomere formation
(Under the direction of Frank L. Conlon)

Cardiac development relies on proper cardiomyocyte differentiation, including expression and assembly of cell-type specific actomyosin subunits into a functional cardiac sarcomere. Control of this process involves not only promoting expression of cardiac sarcomere subunits but also repressing expression of non-cardiac myofibril paralogs. This level of transcriptional control requires broadly expressed multiprotein machines that modify and remodel the chromatin landscape to restrict transcription machinery access. Prominent among these is the Nucleosome Remodeling and Deacetylase (NuRD) complex, which includes the catalytic core subunit CHD4. Here, we demonstrate that direct CHD4-mediated repression of skeletal and smooth muscle myofibril isoforms is required for normal cardiac sarcomere formation, function, and embryonic survival early in gestation. Through transcriptomic and genome-wide analyses of CHD4 localization, we identified novel CHD4 binding sites in smooth muscle myosin heavy chain, fast skeletal α-actin, and the fast skeletal troponin complex genes. We further demonstrate that in the absence of CHD4, cardiomyocytes in the developing heart form a hybrid muscle cell that contains cardiac, skeletal and smooth muscle myofibril components. These misexpressed paralogs intercalate into the nascent cardiac sarcomere to disrupt sarcomere formation and cause impaired cardiac function

in utero. We further identify two new binding partners of CHD4, GATA4 and NKX2-5, which are predicted to recruit the NuRD complex to these regulatory loci to mediate non-cardiac myofibril isoform repression in cardiomyocytes. These results demonstrate the genomic and physiological requirements for CHD4 in mammalian cardiac development and cardiomyocyte differentiation.

To my mom, for her unconditional love and support, and for demonstrating what hard work and service to others looks like every day.

ACKNOWLEDGEMENTS

There are many people without whom this work would not have been possible. The members of the Conlon lab have been my most valuable sounding board, teachers, friends, and coworkers. I would like to personally thank Chris Slagle, Leslie Kennedy, Lauren Kuchenbrod, Zachary Robbe, Haley Davies, Caroline Tarallo, Ross Carroll, Meng Zou, Katie Sampuda, and Tia Andrade. I would especially like to thank Kerry Dorr, Panna Tandon and Lauren Waldron for teaching me everything I know, making themselves available to answer every question I had, and assuring me that I am capable and know more than I think I do. I am so grateful for the lab happy hours that kept things fun, the practice talks that made me a better speaker, the lab meetings that made me a better scientist, and the friends that made me want to come into lab every day.

I would like to thank and acknowledge the collaborators that made this research possible. The Cristea, Wade, and Davis labs were all instrumental in not only completing this research but also patiently explaining the underlying why and how so I could formulate more intelligent questions and design better experiments. I would particularly like to thank Austin Hepperla, Joel Federspiel, Todd Greco, and Takashi Shimbo for their time and expertise.

I would like to thank my committee for their time and advice developing this project and ensuring its success. Brian Strahl, Li Qian, Paul Wade and Scott Williams

have been supportive, inquisitive and interested in this project since its inception, and for that I am very grateful.

I cannot express how grateful I am to Frank Conlon for being my mentor and advisor. His wisdom and guidance have shaped the way I think, communicate, and do science and has made me more articulate, thoughtful, and purposeful. He has always been supportive of my endeavors and given me the space and independence to explore and try new things on my own. I will truly miss our meetings, discussions, and chats from which I always walked away having learned something new.

Finally, I would like to thank my family and friends for their boundless encouragement. My family has always been supportive of my seemingly endless pursuit of knowledge and education, and I thank them for being with me every step of the way. I would like to thank the incredible friends who have made graduate school the best experience with their companionship, commiseration, laughter, fun-loving attitudes and availability to drop everything and go for a run when things got tough. I want to thank James for everything, including but not limited to keeping me fed, happy, and laughing throughout this whole process. His confidence in me kept me going and encouraged me to push beyond what I thought was possible.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1: Introduction

DNA encodes the instructions that permit multicellular life to exist. Remarkably, the sequence of an organism's DNA is identical in all of its somatic cells; however, within one multicellular organism, there are many specialized cell types that express distinct gene sets. The process by which one pluripotent progenitor cell generates a multiplicity of tissue-specific cells in a developing embryo requires exquisite coordination of gene regulation and expression. This process occurs in several stages: As pluripotency is limited, lineage-determining factors are activated, and terminally differentiated cells express the genes necessary to carry out their ultimate cellular functions. This dynamic process involves a balance between the activation and repression of certain genes. This temporal flux in gene expression is exemplified by the differentiation of the first specialized cell type in the earliest developing organ in vertebrates—a cardiomyocyte in the developing heart.

One mechanism by which the cell restricts the differentiation of pluripotent cells into cardiomyocytes is the regulation of gene accessibility. By wrapping DNA around histone proteins to form nucleosomes, the cell is able to regulate which DNA sequences are accessible for transcription and which are hidden and unavailable. Nucleosomal DNA, known as chromatin, can be further organized and packaged into highly ordered structures that affect its accessibility. Structural proteins such as histones can also be modified to encourage or discourage gene expression within these regions.

These organizational processes undergo dynamic changes throughout differentiation, as the cell requires access to different genes during the pluripotency, mesodermal, cardiac progenitor, and cardiomyocyte stages. Chromatin structure can thus carefully guide the cell along its intended path by permitting access to certain genes while restricting access to genes that are either no longer needed or expressed in parallel lineages. This process requires the careful coordination of many aspects of chromatin biology, including topological structure organization, biochemical modification of structural scaffolding proteins such as histones, and dynamic alterations in genome accessibility mediated by chromatin remodeling complexes. The precise control of each of these elements permits the progression from a pluripotent progenitor cell to a functional cardiomyocyte, capable of contraction to drive cardiac function and support continued embryonic growth.

Cardiomyocyte differentiation

Cardiomyocyte differentiation relies on the progressive induction and integration of transcriptional and signaling cues (Brade et al. 2013; Noseda et al. 2011; Evans et al. 2010). The information below will therefore serve as a brief summary of the events required for the progression of pluripotent stem cells through mesodermal lineage commitment, cardiac progenitor specification, and cardiomyocyte differentiation. The molecular mechanisms underlying these transitions have been elucidated using both *in vitro* differentiation and *in vivo* developmental genetics tools in several species, including humans; for the sake of simplicity, mouse gene and protein nomenclature will be used throughout the text below.

The first step in cardiomyocyte differentiation from a pluripotent state is mesoderm specification, which is accomplished via the combination of BMP4 (Kattman et al. 2011), Activin A/Nodal (Conlon et al. 1994; Johansson and Wiles 1995; Kattman et al. 2011), and canonical Wnt3a/β-catenin (Nakamura et al. 2003; Ueno et al. 2007; Flaim et al. 2008) signals (Figure 1.1). The subsequent activation of *Flk-1, Pdgfr-α* (Kattman et al. 2011), and Brachyury or *T* (Conlon et al. 1994; Liu et al. 2007), and later *Eomes* (Aramaki et al. 2013), is demonstrative of successful mesoderm specification. Wnt3a signaling also activates the expression of *Sox17*, which allows the cell to progress to cardiogenic specification by arresting canonical Wnt signaling via noncell-autonomous means (Liu et al. 2007).

In turn, these mesodermal factors (particularly Brachyury) drive the specified cell toward a cardiogenic mesodermal state by inducing *Mesp1*, which is the first definitive cardiogenic marker expressed (David et al. 2011; Kattman et al. 2011). The continuation of cardiogenic specification is further aided by the Mesp1-mediated activation of *Dkk1*, an inhibitor of canonical Wnt/β-catenin signaling (David et al. 2008). Thereafter, the continuation of BMP2 signaling, together with FGF and noncanonical Wnt11 signaling, is required to induce the expression of *Gata4*, a key transcription factor in cardiogenesis (Schlange et al. 2000).

The transcriptional activities of GATA4 and BMP-dependent SMAD induce the expression of the determinant cardiac progenitor markers *Nkx2-5* and *Isl1* (Lien et al. 1999; Lien et al. 2002; Ueno et al. 2007; Cagavi et al. 2014; Quaranta et al. 2018). NKX2-5 reinforces *Gata4* expression and activates *Tbx5* expression to signal the transition into the cardiac progenitor stage (Kennedy et al. 2017; Luna-Zurita et al.

2016). These transcription factors activate the expression of both downstream cardiomyocyte-specific transcriptional mediators and markers of terminal cardiomyocyte differentiation (Luna-Zurita et al. 2016; Quaranta et al. 2018). The induced expression of structural proteins that are characteristic of cardiomyocyte differentiation begins very early during the cardiac progenitor stage, although robust expression and functional cellular differentiation do not occur until later (Kokkinopoulos et al. 2015). Markers of robust cardiomyocyte differentiation include subunits of the cardiac sarcomere, the functional contractile unit of the cardiomyocyte, such as *Myl2, Myl7, Tnnt2, Myh6, Actc1*, and others.

Cardiomyocyte differentiation is distinguished by formation of the contractile apparatus, the cardiac sarcomere. The sarcomere is assembled in several stages in a gradual process that undergoes refinement and maturation over time (Ehler et al. 1999; Hirschy et al. 2006). Molecules of cardiac α-actin and tropomyosin intercalate to form the thin filament, which is then held in alignment by structural proteins such as α-actinin at the Z-disk (Hewett et al. 1994). The thin filament is bound by the cardiac troponin complex composed of TnC1, TnT2 and TnI1 (with a low abundance of skeletal TnI2) (Nishii et al. 2008; Wang et al. 2001; Zhu et al. 1995). The thick filament, composed of myosin-binding proteins and myosin chain isoforms (such as β-myosin heavy chain in embryonic cardiomyocytes), interdigitates between thin filaments and is held in alignment by proteins such as myomesin at the M-line (England and Loughna 2013). At the opposite pole of the thick filament, titin anchors the thick filament to the Z-disc. Formation of these stereotypical sarcomere structures denotes the final stages of cardiomyocyte differentiation and the initiation of specialized cellular function.

Topological chromatin organization

The activation or repression of genes required for cardiomyocyte differentiation is determined by the ability of regulatory elements to access the target gene's promoter to start, stop, or prevent transcription. These regulatory elements can be trans-activating factors, such as transcription factors or transcriptional machinery such as RNA polymerase II. Regulatory elements can also take the form of cis-regulatory components, such as enhancers, by recruiting trans-activating factors to target gene promoters via DNA looping or the relaxation or condensation of the surrounding chromatin. The spatial relationships between these elements and their targets determine their efficacy; thus, the process by which chromatin is organized determines whether a gene program is subsequently executed.

Chromatin compartmentalization

The organization of chromatin into distinct domains within the nucleus is the highest level of genomic organization. These domains form zones of interactions between genomic regulatory regions that control gene activation and repression during cardiomyocyte differentiation. As the cell transitions through differentiation and its transcriptional profile changes, these zones sequester genes that will no longer or never need to be activated. Recent research has found that chromatin exists in a relatively relaxed state after fertilization in embryonic stem cells (ESCs) and gradually moves toward higher levels of compartmentalization and packaging as differentiation proceeds

(Park et al. 2004; Du et al. 2017; Hug et al. 2017; Ke et al. 2017). Compartmentalization begins at the chromosome level, with chromosome regions packed into “shell-like” layers of dense, inactive perinuclear compartments and a less dense, active internal nuclear compartment (Cremer et al. 2015; Szalaj and Plewczynski 2018). The active compartment, or compartment A, contains a higher gene density, transcriptional activity, and greater chromatin accessibility, hallmarks of euchromatin, than the inactive regions (Rao et al. 2014; Lieberman-Aiden et al. 2009). Compartment B contains genomic regions associated with chromatin packaging and formation of heterochromatin; it is often associated with the nuclear lamina. These features manifest in the reduced expression of genes contained within this compartment (Rao et al. 2014; Ryba et al. 2010). Previous research findings suggest that chromatin within these compartments is arranged into disordered fibers of 5–24 nm in length that are uniformly varied across regions of additional compaction (Ou et al. 2017).

Research detailing the process of chromatin compartmentalization during cardiomyocyte differentiation is currently lacking. However, the importance of establishing nuclear compartments during cardiomyocyte differentiation is emphasized by clinical phenotypes exhibited by patients with mutations in nuclear lamin proteins (Fidzianska et al. 2008). These mutations show disproportionate phenotypic effects on cardiac and skeletal muscle, leading to such diagnoses as dilated cardiomyopathy. Myocardial biopsies reveal misshapen nuclear deformation, randomly distributed clumps of heterochromatin, and leakage of chromatin outside of the nuclear envelope. Such cases highlight the importance of normal chromatin organization.

Interactions between topologically associated domains (TADs) and cis-regulatory regions

Within these compartments, chromatin is further organized into loops by proteins such as CTCF and cohesion (Tang et al. 2015; Rubio et al. 2008; Parelho et al. 2008). These loops are organized into 5–20-megabase (Mb) megadomains, which are further arranged into approximately 1-Mb TADs with distinct boundaries between them (Rao et al. 2014). While research into the dynamic properties of TAD formation during cardiomyocyte differentiation is scarce, knowledge can be inferred from extensive studies in other cell types. These studies have indicated that TADs are maintained throughout differentiation, without substantial changes in genomic regions contained within each TAD (Dixon et al. 2015; Rao et al. 2014). Furthermore, TADs appear to be the primary unit for compartment A/B switching during gene expression transitions throughout differentiation (Rao et al. 2014; Dixon et al. 2015; Schoenfelder et al. 2015). These findings indicate that TADs may stabilize the genome during expression profile switching throughout cellular differentiation.

Chromatin loops within TADs function to bring *cis*-regulatory elements into close spatial proximity with one another and with target gene promoters. Conversely, these loops may also keep cis-regulatory regions sequestered from each other and target promoters when the activation of gene expression would be inappropriate (Dixon et al. 2015). These chromatin loops allow distal elements to bypass the nearest promoters and interact with three-dimensionally local but linearly distant promoters, thereby extending their regulatory influence. They may also allow multiple cis-regulatory regions to interact simultaneously to affect gene expression; previous research has

demonstrated that gene activation can be amplified by one promoter binding multiple enhancers. Indeed, highly active promoters have been shown to engage in a greater number of long-range interactions than less active promoters, suggesting an additive effect of enhancer interactions (Schoenfelder et al. 2015).

Paradoxically, increasing the number of promoter interactions with one enhancer does not necessarily correlate with the activation of target gene expression. These interconnected “hubs” of one enhancer bound to multiple promoters may instead reveal a spatial, physical method of coordinating gene expression (Schoenfelder et al. 2015). Fittingly, prior research has found that TADs segregate chromatin into regions of similar expression patterns (Rao et al. 2014). Within these domains, promoter-promoter interactions can also be facilitated, with preferential interactions between promoters at a similar expression level and with similar gene ontological terms (Schoenfelder et al. 2015). Chromatin loops within TADs are thereby important physical mediators of transcriptional regulation via the bringing of cis-regulatory elements into close proximity with one another.

When examining how changes in *cis*-regulatory region interactions contribute to dynamic gene expression changes during cardiomyocyte differentiation, little evidence exists. However, information gleaned from other models of differentiation may provide a provisional blueprint. Differentiation of murine fetal liver cells demonstrates that the number of promoter interactions within TADs increases as differentiation proceeds. Furthermore, during differentiation, there is an extensive rewriting of enhancer-promoter interactions, as very few enhancers interact with the same promoters in both ESCs and differentiated cells (Schoenfelder et al. 2015). Analyzing enhancer-promoter interactions

between different cell types has yielded similar results, with striking differences in cis-regulatory region interactions between nine different cell types (Rao et al. 2014). These dynamic changes in cis-regulatory interactions influence transcriptional activity, which in turn changes chromatin compartmentalization with expansion of the compartment B upon lineage differentiation into mesendoderm, mesenchymal, and neural progenitor cells. As lineage-specific genes are activated, they move into compartment A where there is greater chromatin accessibility and access to transcriptional hubs (Dixon et al. 2015; Schoenfelder et al. 2015). It is likely that a similar process occurs during cardiomyocyte differentiation, as mesoderm is specified, cell fate is determined, cardiomyocyte differentiation proceeds, and other cell fate programs are restricted. Additional experimental evidence is needed to confirm these hypotheses.

Influencing chromatin architecture through histone modifications

How these TADs and cis-regulatory region interactions regulate gene expression during cardiomyocyte differentiation is intricately connected to the histone modifications present at these genetic loci. Other models of cellular differentiation have shown that, in tandem with the regulation of gene expression, histones play an important role in the regulation of interactions between cis-regulatory elements within and between TADs, and create a chromatin signature to mediate intracompartmental shuffling (Schoenfelder et al. 2015; Rao et al. 2014; Dixon et al. 2015). Furthermore, modifications of histones change their affinity to DNA, leading to the opening or closing of genomic regions, as well as their affinity to the trans-activating factors of nucleosomal DNA.

Each subunit of the histone octomer [which comprises two subunits each of histones H2A,H2B, H3, and H4, with potential substitutions such as H2A.Z at transcriptional start sites (TSSs)] (Sutcliffe et al. 2009; Peterson and Laniel 2004) can be reversibly altered by biochemical modifications of amino acids in the N-terminal tail. These modifications serve as distinguishing features for a genomic locus to bind or repel chromatin- or DNA-associated proteins, such as chromatin remodelers, transcription factors, and/or transcriptional machinery, and facilitate interactions between cis-regulatory elements. Common modifications include, but are not limited to, acetylation, methylation, phosphorylation, ubiquitination, or additive modifications of the above.

Considering how dynamic histone modifications regulate chromatin architecture and gene expression, their role in the regulation of highly ordered chromatin is relatively static. Previous research indicates that TAD boundaries are typically marked by trimethylated (me3) H3K36 and H3K4me3 modifications (Rao et al. 2014). Given that TADs do not tend to shift during differentiation but instead more commonly change interactions within or between domains, it can be surmised that a broader regulatory role for histone modifications lies in making smaller changes within these domains to regulate these interactions. Histone modifications may also serve to recruit transcriptional machinery in a dynamic fashion, as gene expression patterns change over the course of cardiomyocyte differentiation (Schoenfelder et al. 2015; Rao et al. 2014).

Effects of the histone code on cardiomyocyte differentiation

While histone modifications are an ongoing topic of research and discovery, a topographic map of their effects on regulatory elements and levels of gene expression has emerged in recent years. Vast knowledge within the chromatin biology field has proven applicable to many models of differentiated cells and has been useful in the establishment of a “histone code” that plays a role in the regulation of gene expression during cardiomyocyte differentiation. As one stage of differentiation concludes and another begins, individual genes or cis-regulatory elements lose and/or gain histone modifications (Papait et al. 2013). In addition, certain histone changes are preserved throughout multiple stages of differentiation, in order to retain potential activity, depending on the receipt of certain differentiation cues. These changes in the histone code influence the cell’s gene expression profile, allowing it to leave the pluripotent stage, activate genes that signify cardiac lineage commitment while decommissioning alternative cell fate genes, and ultimately permit the transcription of markers of cardiomyocyte differentiation. Therefore, understanding the effects of the most common modifications on transcriptional or regulatory activity informs how the histone code influences progression through cardiomyocyte differentiation.

Research has demonstrated that the H3K4me3 modification is present in promoters of actively transcribed genes during discrete stages of gene expression over the course of cardiomyocyte differentiation (Schoenfelder et al. 2015). It is then removed when the transcription of a stage-specific gene is no longer necessary. For example, H3K4me3 is present in promoters of genes such as Nodal in ESCs before mesoderm induction but is removed at the cardiac progenitor stage as differentiation

progresses (Wamstad et al. 2012; Paige et al. 2012). Conversely, H3K4me3 is not present in markers of differentiated cardiomyocytes such as *Tnnt2* during the ESC and mesodermal stages, preventing the ectopic expression of differentiation markers before proper specification and determination stages have taken place. These genes are only expressed during the late cardiac progenitor and cardiomyocyte stages after gaining the H3K4me3 modification during the late cardiac progenitor stage (Paige et al. 2012; Wamstad et al. 2012).

H3K36me3 is another “active” histone modification that marks transcriptional elongation; when it is lost, it signifies that the gene’s expression has been terminated, and the cell is transitioning out of that stage of differentiation (Schoenfelder et al. 2015; Rao et al. 2014). For example, H3K36me3 is present in the genes *Eomes* and Brachyury/*T* in mesoderm cells during cardiac differentiation but is lost as differentiation proceeds into the cardiac progenitor stage (Paige et al. 2012).

Other histone modifications are dynamically regulated within cis-regulatory regions, such as enhancers or proximal promoter regions. For example, H3K27 acetylation (ac) marks the activation of enhancers for genes that are transcribed or will be imminently expressed; it is removed as the enhancer is decommissioned during the transition into another differentiation stage (Wamstad et al. 2012; Creyghton et al. 2010; Schoenfelder et al. 2015; Rao et al. 2014; Dixon et al. 2015). For example, H3K27ac is present in the proximal promoter region of *Isl1* during the mesodermal stage of cardiomyocyte differentiation, which precedes *Isl1* expression in cardiac progenitor cells (Wamstad et al. 2012).

H3K4 mono-methylation (me1) is another modification found in proximal promoter regions or enhancers, and has the slightly more ambiguous role of indicating so-called “bivalent” or “poised” promoters; that is, research has demonstrated that it may mark regions capable of being activated during differentiation following deposition of a more definitive activating marker, such as H3K27ac (Rada-Iglesias et al. 2010; Creyghton et al. 2010; Heintzman et al. 2009; Schoenfelder et al. 2015; Rao et al. 2014; Dixon et al. 2015). Enhancers associated with *Actc1* (cardiac α-actin), which is expressed in differentiated cardiomyocytes, demonstrate the presence of this bivalent state throughout differentiation from the ESC to the cardiac progenitor stage, but *Actc1* expression is only induced after H3K27ac deposition during the mesodermal and cardiac progenitor stages (Wamstad et al. 2012). Without deposition of coactivating markers, regions containing H3K4me1 tend to remain silent and do not activate cell-type-specific gene expression programs (Creyghton et al. 2010; Heintzman et al. 2009). In differentiating cardiomyocytes, these regions tend to be associated with genes in noncardiac lineages, such as genes related to neural development (Wamstad et al. 2012).

Certain histone modifications have been found to correlate with the repression of gene expression. Acquisition of H3K27me3 modifications of enhancers or proximal promoter regulatory regions correlates with the repression of genes that are no longer necessary, as the cell progresses throughout cardiomyocyte differentiation, demonstrating the dynamic control of gene expression via histone modifications. For example, ESC genes such as *NANOG* gain H3K27me3 after the ESC stage; expression

of such genes coordinately drops after this stage and disappears completely after the cardiac progenitor stage of differentiation (Wamstad et al. 2012).

H3K27me3 is also a constant signal of repression throughout cellular differentiation. It has been found in bivalent promoters or enhancers of inactive genes with H3K4me1; research has shown that over the course of differentiation, the cell can resolve these bivalent enhancers or promoters through histone modification removal to guide the cell along the intended differentiation gene expression pattern (Rada-Iglesias et al. 2010; Cui et al. 2009). Interestingly, the repression of noncardiac lineage genes, such as skeletal or smooth muscle-specific transcription factors or structural proteins, during cardiomyocyte differentiation correlates with either the constant presence of H2K27me3 or the absence of detectable H3K27me3 deposition, indicating multiple avenues by which the cell can mediate transcriptional repression (Paige et al. 2012). While their roles regulating gene expression during cardiomyocyte differentiation have not yet been examined, H3K9 dimethylation (me2) and H3K9me3 are two other histone modifications that are demonstrated to correlate with tissue-specific gene silencing in other models of differentiation (Wen et al. 2009). H3K9me2 and H3K9me3 deposition has been shown to correlate with gene repression in a model of cardiac hypertrophy (Papait et al. 2013). Taken together, this research demonstrates that dynamic alterations of certain histone modifications play a significant role in the regulation of gene expression throughout the stages of cardiomyocyte differentiation.

Chromatin remodelers shift chromatin organization

If histone modifications potentiate alterations in chromatin conformation, leading to changes in gene accessibility and activation, the mechanism by which these structural changes occur remains an unanswered question.

Chromatin remodeling complexes interact with cofactors that read the histone code established at these loci and use ATP-dependent helicase subunits to change the local chromatin architecture, either promoting or repressing gene expression. These complexes are responsible for opening or closing the chromatin by moving nucleosomes along the DNA strand to either permit or restrict gene access by trans-activating factors or transcriptional machinery. The chromatin remodeling enzymes in each complex are hypothesized to move nucleosomes along the DNA strand by two mechanisms: remaining in a fixed position on a nucleosome while rotating the DNA around this position, thus changing the DNA bound to the histone, or by binding the histone and turning the enzyme-histone complex around a fixed DNA position like a screw (Peterson 2000).

There are four major chromatin remodeling complexes that are distinguishable by the ATP-dependent remodeling subunit they utilize: vertebrate BRM/BRG1-associated factor (BAF) complexes, which are homologs of the yeast switch/sucrose nonfermentable (SWI/SNF) complexes; chromodomain-helicase-DNA-binding (CHD) complexes, including the nucleosome remodeling and deacetylase (NuRD) complex; imitation SWI (ISWI) complexes; and inositol-requiring 80 (INO80) complexes (Hota and Bruneau 2016). Each of these complexes contains multiple subunits with distinct roles in complex stability, cofactor binding and recruitment, and/or chromatin modification

during cellular differentiation (Chen and Dent 2013; Ho and Crabtree 2010). For the purposes of examining the dynamic alterations in chromatin conformation during cardiomyocyte differentiation, the information presented below will focus on each complex's ATP-dependent chromatin remodeling subunit.

BRM/BRG1 complexes

Vertebrate BAF complexes can contain either the ATP-dependent chromatin remodeling subunit Brahma (BRM) or Brahma-related gene 1 (BRG1). Brahma has been shown to be dispensable for embryonic survival, indicating functional redundancy with BRG1, as BRM-null mice survive to adulthood with only slight proliferation defects that lead to increased body mass, although the fidelity of this mouse model has recently been called into question (Reyes et al. 1998; Hota and Bruneau 2016). BRG1 has been shown to be required for ESC renewal and pluripotency via the activation of pluripotency genes such as *Oct4* and *Sox2*, among others. BRG1 may also bind to bivalent promoters and enhancers containing H3K4me3 and H3K27me3 modifications in genes related to differentiated ESC lineages, demonstrating that BRG1 may also act to transcriptionally repress the premature differentiation of these pluripotent cells (Kidder, Palmer, and Knott 2009). Once mesoderm specification signals are received, BRG1 binds to H3K27ac-modified enhancer regions of *Mesp1* and *Flk1* to contribute to the active expression of mesodermal lineage markers via chromatin remodeling. At this stage, BRG1 also represses transcription factors such as *Tbx5* and *Nkx2-5* as well as nonmesodermal lineage markers by binding H3K27me3-modified enhancers to repress premature differentiation or differentiation down the wrong lineage (Alexander et al.

2015). As cardiomyocyte differentiation proceeds, BRG1 ensures the proper timing of cardiomyocyte maturation via the repression of adult, mature cardiomyocyte sarcomere subunits such as *Myh6* (α-myosin heavy chain) while activating the expression of the developmentally predominant *Myh7* (β-myosin heavy chain) (Hang et al. 2010). Interestingly, BRG1 is not required in differentiated cardiomyocytes to maintain beating and differentiation (Alexander et al. 2015), and is not expressed in adult cardiomyocytes (Hang et al. 2010), indicating that establishment of chromatin architecture for some genes may not require constant maintenance.

CHD complexes

CHD complexes may contain any one of the CHD family members 1–9 that can function either in a complex with other proteins or independently (Marfella and Imbalzano 2007; Hota and Bruneau 2016). Predominant among CHD nucleosome remodeling complexes is the NuRD complex, which utilizes either CHD3 (also known as Mi-2α) or CHD4 (also known as Mi-2β) as the ATP-dependent DNA helicase subunit (Wade et al. 1999; Wade et al. 1998; Zhang et al. 1998). CHD3 and CHD4 have been demonstrated to bind H3K9me2 and H3K9me3, as well as H3K9ac (Mansfield et al. 2011; Musselman et al. 2009; Musselman et al. 2012). CHD4 and the NuRD complex are required in ESCs to repress the early expression of mesoderm genes such as *Eomes,* Brachyury/*T,* and *NANOG* (Beyer et al. 2013). CHD4 also represses the inappropriate early expression of cardiac lineage markers such as *Gata4* and modulates the expression of certain ESC genes such as *Oct4* (O'Shaughnessy-Kirwan et al. 2015). Research has shown that CHD4 is required for normal mesoderm specification via the

repression of the neural fate regulator *Sip1*, which permits *Xbra* (Brachyury/*T*) expression and mesoderm specification (Linder et al. 2007). The role of CHD4 in later stages of cardiac progenitor specification have yet to be elucidated (Hota and Bruneau 2016), whereas the role of CHD4 in cardiomyocyte differentiation is a topic that is currently under investigation.

ISWI complexes

ISWI complexes contain either the ATP-dependent chromatin remodeler SNF2L or SNF2H (Lazzaro and Picketts 2001). SNF2L is expressed in differentiating cells; however, *Snf2l (Smarca1)*-null mice develop normally (Yip et al. 2012), indicating that it may not play a significant role in the dynamic regulation of chromatin architecture during cardiomyocyte differentiation. SNF2H (*Smarca5*) ablation is embryonic lethal due to proliferation defects (Stopka and Skoultchi 2003); however, studies ascertaining its role in mesoderm specification or cardiomyocyte differentiation are lacking, leaving its precise role in these processes unknown (Han et al. 2011; Hota and Bruneau 2016).

INO80 complexes

INO80-containing complexes remodel chromatin by pairing the ATP-dependent catalytic activity of INO80 with the DNA helicase action of RVB1 (*Rvbl1*), also known as Pontin, or RVB2 (*Rvbl2*), also known as Reptin (Huen et al. 2010; Han et al. 2011). INO80 is required for ESC maintenance via chromatin remodeling, specifically nucleosome depletion, of stem cell pluripotency genes (Wang et al. 2014). Its role in mesoderm specification or cardiomyocyte differentiation has yet to be determined,

although studies in zebrafish have identified a requirement for both Pontin and Reptin in the regulation of cardiomyocyte proliferation (Rottbauer et al. 2002).

Dissertation goals

While dynamic changes in chromatin architecture are clearly required for cardiomyocyte differentiation and embryonic heart development, mechanisms underlying these changes remain relatively unknown. The role of chromatin remodeling complexes in this process is a particularly interesting area of research, as they incorporate the dynamic changes in chromatin accessibility and histone modifications with differentiation cues in the form of stage-specific co-factors that change where the complex is targeted. Incorporation of all these signals results in alterations of the local chromatin architecture that lead to modulation of stage-specific gene expression profiles that permit continuation of the cardiomyocyte differentiation program. Execution of this program is required for embryonic heart development and survival, as cardiomyocyte formation is essential to initiate systolic function to circulate oxygen and nutrients throughout the growing embryo.

The role of the NuRD complex in this process has generated considerable interest as it interacts with several cardiac transcription factors mutated in patients with congenital heart defects, such as TBX5, FOG-2 and TBX20 (Waldron et al. 2016; Kaltenbrun et al. 2013; Garnatz et al. 2014; Roche et al. 2008; Basson et al. 1997; Kirk et al. 2007; Tevosian et al. 2000; De Luca et al. 2011). While the NuRD complex's role in ESCs and mesoderm specification has been studied in vitro, its role in cardiomyocyte differentiation and contribution to heart development remain relatively unknown. Some

studies have interrogated its role in a limited context by investigating its interactions with individual co-factors (Kaltenbrun et al. 2013; Waldron et al. 2016; Garnatz et al. 2014; Roche et al. 2008); however, the field lacks an explanation of how the NuRD complex contributes to this process outside these specific interactions.

We therefore sought to investigate the role of the NuRD complex during cardiomyocyte differentiation and cardiac development. We elected to use an in vivo mouse model system to identify the requirement for the NuRD complex within the endogenous cardiac developmental context. Chapter 2 describes a comprehensive study using systems-level genomic, transcriptomic and phenotypic tools that found CHD4 and the NuRD complex are required to repress fast skeletal and smooth muscle myofibril isoforms during cardiac development. Transcriptional repression of these isoforms via the NuRD complex is required for normal cardiac sarcomere formation. In the absence of CHD4-mediated repression, cardiomyocytes form a hybrid cardiac, skeletal, and smooth muscle type that cannot properly initiate systolic function in the developing heart, leading to early embryonic lethality. In Chapter 3, we identify the transcriptional co-factors that recruit CHD4 to these regulatory targets using computational biology and proteomics. We uncover two new co-factors for CHD4, GATA4 and NKX2-5, which have been previously shown to be required for normal cardiac development and cardiomyocyte differentiation. Collectively, this work begins to unravel the requirement of the NuRD complex in regulating cardiomyocyte identity and embryonic heart development.

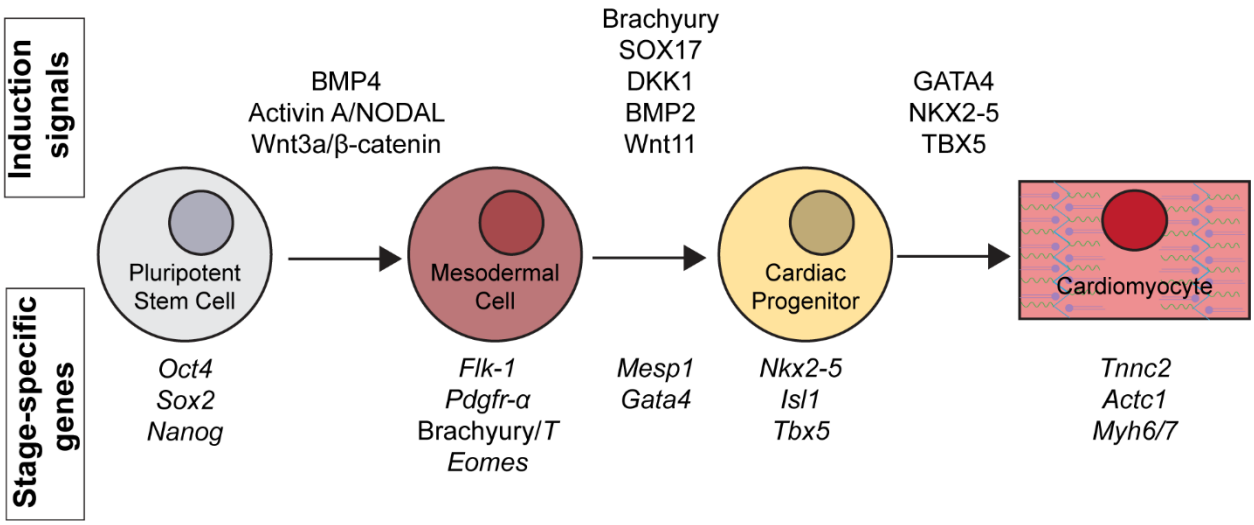


Figure 1.1. Cardiomyocyte differentiation from pluripotent stem cells. Exogenous and endogenous signaling cues induce sequential mesoderm specification, cardiac progenitor specification, and cardiomyocyte differentiation in pluripotent stem cells. As the cell progresses through this process, its gene expression profile will change and unique stage-specific genes will be expressed while genes specific to earlier stages are repressed.

REFERENCES

Alexander, J. M., S. K. Hota, D. He, S. Thomas, L. Ho, L. A. Pennacchio, and B. G. Bruneau. 2015. 'Brg1 modulates enhancer activation in mesoderm lineage commitment', *Development*, 142: 1418-30.

Aramaki, Shinya, Katsuhiko Hayashi, Kazuki Kurimoto, Hiroshi Ohta, Yukihiro Yabuta, Hiroko Iwanari, Yasuhiro Mochizuki, Takao Hamakubo, Yuki Kato, Katsuhiko Shirahige, and Mitinori Saitou. 2013. 'A Mesodermal Factor, T, Specifies Mouse Germ Cell Fate by Directly Activating Germline Determinants', *Dev Cell*, 27: 516-29.

Basson, C. T., D. R. Bachinsky, R. C. Lin, T. Levi, J. A. Elkins, J. Soults, D. Grayzel, E. Kroumpouzou, T. A. Traill, J. Leblanc-Straceski, B. Renault, R. Kucherlapati, J. G. Seidman, and C. E. Seidman. 1997. 'Mutations in human TBX5 [corrected] cause limb and cardiac malformation in Holt-Oram syndrome', *Nat Genet*, 15: 30-5.

Beyer, Tobias A, Alexander Weiss, Yuliya Khomchuk, Kui Huang, Abiodun A Ogunjimi, Xaralabos Varelas, and Jeffrey L Wrana. 2013. 'Switch Enhancers Interpret TGF-β and Hippo Signaling to Control Cell Fate in Human Embryonic Stem Cells', *Cell Reports*, 5: 1611-24.

Brade, Thomas, Luna S. Pane, Alessandra Moretti, Kenneth R. Chien, and Karl-Ludwig Laugwitz. 2013. 'Embryonic Heart Progenitors and Cardiogenesis', *Cold Spring Harbor Perspectives in Medicine*, 3.

Cagavi, Esra, Oscar Bartulos, Carol Y. Suh, Baonan Sun, Zhichao Yue, Zhengxin Jiang, Lixia Yue, and Yibing Qyang. 2014. 'Functional Cardiomyocytes Derived from Isl1 Cardiac Progenitors via Bmp4 Stimulation', *PLoS One*, 9: e110752.

Chen, Taiping, and Sharon Y. R. Dent. 2013. 'Chromatin modifiers and remodellers: regulators of cellular differentiation', *Nature Reviews Genetics*, 15: 93.

Conlon, F. L., K. M. Lyons, N. Takaesu, K. S. Barth, A. Kispert, B. Herrmann, and E. J. Robertson. 1994. 'A primary requirement for nodal in the formation and maintenance of the primitive streak in the mouse', *Development*, 120: 1919-28.

Cremer, T., M. Cremer, B. Hubner, H. Strickfaden, D. Smeets, J. Popken, M. Sterr, Y. Markaki, K. Rippe, and C. Cremer. 2015. 'The 4D nucleome: Evidence for a dynamic nuclear landscape based on co-aligned active and inactive nuclear compartments', *FEBS Lett*, 589: 2931-43.

Creyghton, Menno P., Albert W. Cheng, G. Grant Welstead, Tristan Kooistra, Bryce W. Carey, Eveline J. Steine, Jacob Hanna, Michael A. Lodato, Garrett M. Frampton, Phillip A. Sharp, Laurie A. Boyer, Richard A. Young, and Rudolf Jaenisch. 2010. 'Histone H3K27ac separates active from poised enhancers and predicts developmental state', *Proceedings of the National Academy of Sciences*, 107: 21931-36.

Cui, K., C. Zang, T. Y. Roh, D. E. Schones, R. W. Childs, W. Peng, and K. Zhao. 2009. 'Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation', *Cell Stem Cell*, 4: 80-93.

David, R., C. Brenner, J. Stieber, F. Schwarz, S. Brunner, M. Vollmer, E. Mentele, J. Müller-Höcker, S. Kitajima, H. Lickert, R. Rupp, and W. M. Franz. 2008. 'MesP1 drives vertebrate cardiovascular differentiation through Dkk-1-mediated blockade of Wnt-signalling', *Nat Cell Biol*, 10: 338.

David, Robert, Veronica Barbara Jarsch, Florian Schwarz, Petra Nathan, Moritz Gegg, Heiko Lickert, and Wolfgang-Michael Franz. 2011. 'Induction of MesP1 by Brachyury(T) generates the common multipotent cardiovascular stem cell', *Cardiovasc Res*, 92: 115-22.

De Luca, A., A. Sarkozy, R. Ferese, F. Consoli, F. Lepri, M. L. Dentici, P. Vergara, A. De Zorzi, P. Versacci, M. C. Digilio, B. Marino, and B. Dallapiccola. 2011. 'New mutations in ZFPM2/FOG2 gene in tetralogy of Fallot and double outlet right ventricle', *Clin Genet*, 80: 184-90.

Dixon, Jesse R., Inkyung Jung, Siddarth Selvaraj, Yin Shen, Jessica E. Antosiewicz-Bourget, Ah Young Lee, Zhen Ye, Audrey Kim, Nisha Rajagopal, Wei Xie, Yarui Diao, Jing Liang, Huimin Zhao, Victor V. Lobanenkov, Joseph R. Ecker, James A. Thomson, and Bing Ren. 2015. 'Chromatin architecture reorganization during stem cell differentiation', *Nature*, 518: 331.

Du, Z., H. Zheng, B. Huang, R. Ma, J. Wu, X. Zhang, J. He, Y. Xiang, Q. Wang, Y. Li, J. Ma, X. Zhang, K. Zhang, Y. Wang, M. Q. Zhang, J. Gao, J. R. Dixon, X. Wang, J. Zeng, and W. Xie. 2017. 'Allelic reprogramming of 3D chromatin architecture during early mammalian development', *Nature*, 547: 232-35.

Ehler, E., B. M. Rothen, S. P. Hammerle, M. Komiyama, and J. C. Perriard. 1999. 'Myofibrillogenesis in the developing chicken heart: assembly of Z-disk, M-line and the thick filaments', *J Cell Sci*, 112 (Pt 10): 1529-39.

England, J., and S. Loughna. 2013. 'Heavy and light roles: myosin in the morphogenesis of the heart', *Cell Mol Life Sci*, 70: 1221-39.

Evans, S. M., D. Yelon, F. L. Conlon, and M. L. Kirby. 2010. 'Myocardial lineage development', *Circ Res*, 107: 1428-44.

Fidzianska, A., E. Walczak, Z. Glinka, and G. Religa. 2008. 'Nuclear architecture remodelling in cardiomyocytes with lamin A deficiency', *Folia Neuropathol*, 46: 196-203.

Flaim, Christopher J., Dayu Teng, Shu Chien, and Sangeeta N. Bhatia. 2008. 'Combinatorial Signaling Microenvironments for Studying Stem Cell Fate', *Stem Cells and Development*, 17: 29-40.

Garnatz, A. S., Z. Gao, M. Broman, S. Martens, J. U. Earley, and E. C. Svensson. 2014. 'FOG-2 mediated recruitment of the NuRD complex regulates cardiomyocyte proliferation during heart development', *Dev Biol*, 395: 50-61.

Han, Pei, Calvin T. Hang, Jin Yang, and Ching-Pin Chang. 2011. 'Chromatin Remodeling in Cardiovascular Development and Physiology', *Circ Res*, 108: 378-96.

Hang, Calvin T., Jin Yang, Pei Han, Hsiu-Ling Cheng, Ching Shang, Euan Ashley, Bin Zhou, and Ching-Pin Chang. 2010. 'Chromatin regulation by Brg1 underlies heart muscle development and disease', *Nature*, 466: 62-67.

Heintzman, Nathaniel D., Gary C. Hon, R. David Hawkins, Pouya Kheradpour, Alexander Stark, Lindsey F. Harp, Zhen Ye, Leonard K. Lee, Rhona K. Stuart, Christina W. Ching, Keith A. Ching, Jessica E. Antosiewicz-Bourget, Hui Liu, Xinmin Zhang, Roland D. Green, Victor V. Lobanenkov, Ron Stewart, James A. Thomson, Gregory E. Crawford, Manolis Kellis, and Bing Ren. 2009. 'Histone modifications at human enhancers reflect global cell-type-specific gene expression', *Nature*, 459: 108.

Hewett, T. E., I. L. Grupp, G. Grupp, and J. Robbins. 1994. 'Alpha-skeletal actin is associated with increased contractility in the mouse heart', *Circ Res*, 74: 740-6.

Hirschy, A., F. Schatzmann, E. Ehler, and J. C. Perriard. 2006. 'Establishment of cardiac cytoarchitecture in the developing mouse heart', *Dev Biol*, 289: 430-41.

Ho, Lena, and Gerald R. Crabtree. 2010. 'Chromatin remodelling during development', *Nature*, 463: 474.

Hota, Swetansu K., and Benoit G. Bruneau. 2016. 'ATP-dependent chromatin remodeling during mammalian development', *Development*, 143: 2882-97.

Huen, J., Y. Kakihara, F. Ugwu, K. L. Cheung, J. Ortega, and W. A. Houry. 2010. 'Rvb1-Rvb2: essential ATP-dependent helicases for critical complexes', *Biochem Cell Biol*, 88: 29-40.

Hug, C. B., A. G. Grimaldi, K. Kruse, and J. M. Vaquerizas. 2017. 'Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription', *Cell*, 169: 216-28.e19.

Johansson, B M, and M V Wiles. 1995. 'Evidence for involvement of activin A and bone morphogenetic protein 4 in mammalian mesoderm and hematopoietic development', *Mol Cell Biol*, 15: 141-51.

Kaltenbrun, E., T. M. Greco, C. E. Slagle, L. M. Kennedy, T. Li, I. M. Cristea, and F. L. Conlon. 2013. 'A Gro/TLE-NuRD corepressor complex facilitates Tbx20-dependent transcriptional repression', *J Proteome Res*, 12: 5395-409.

Kattman, S. J., A. D. Witty, M. Gagliardi, N. C. Dubois, M. Niapour, A. Hotta, J. Ellis, and G. Keller. 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: 228-40.

Ke, Y., Y. Xu, X. Chen, S. Feng, Z. Liu, Y. Sun, X. Yao, F. Li, W. Zhu, L. Gao, H. Chen, Z. Du, W. Xie, X. Xu, X. Huang, and J. Liu. 2017. '3D Chromatin Structures of Mature Gametes and Structural Reprogramming during Mammalian Embryogenesis', *Cell*, 170: 367-81.e20.

Kennedy, Leslie, Erin Kaltenbrun, Todd M. Greco, Brenda Temple, Laura E. Herring, Ileana M. Cristea, and Frank L. Conlon. 2017. 'Formation of a TBX20-CASZ1 protein complex is protective against dilated cardiomyopathy and critical for cardiac homeostasis', *PLoS Genet*, 13: e1007011.

Kidder, B. L., S. Palmer, and J. G. Knott. 2009. 'SWI/SNF-Brg1 regulates self-renewal and occupies core pluripotency-related genes in embryonic stem cells', *Stem Cells*, 27: 317-28.

Kirk, E. P., M. Sunde, M. W. Costa, S. A. Rankin, O. Wolstein, M. L. Castro, T. L. Butler, C. Hyun, G. Guo, R. Otway, J. P. Mackay, L. B. Waddell, A. D. Cole, C. Hayward, A. Keogh, P. Macdonald, L. Griffiths, D. Fatkin, G. F. Sholler, A. M. Zorn, M. P. Feneley, D. S. Winlaw, and R. P. Harvey. 2007. 'Mutations in cardiac T-box factor gene TBX20 are associated with diverse cardiac pathologies, including defects of septation and valvulogenesis and cardiomyopathy', *Am J Hum Genet*, 81: 280-91.

Kokkinopoulos, Ioannis, Hidekazu Ishida, Rie Saba, Prashant Ruchaya, Claudia Cabrera, Monika Struebig, Michael Barnes, Anna Terry, Masahiro Kaneko, Yasunori Shintani, Steven Coppen, Hidetaka Shiratori, Torath Ameen, Charles Mein, Hiroshi Hamada, Ken Suzuki, and Kenta Yashiro. 2015. 'Single-Cell Expression Profiling Reveals a Dynamic State of Cardiac Precursor Cells in the Early Mouse Embryo', *PLoS One*, 10: e0140831.

Lazzaro, M. A., and D. J. Picketts. 2001. 'Cloning and characterization of the murine Imitation Switch (ISWI) genes: differential expression patterns suggest distinct developmental roles for Snf2h and Snf2l', *J Neurochem*, 77: 1145-56.

Lieberman-Aiden, E., N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander, and J. Dekker. 2009. 'Comprehensive mapping of long-range interactions reveals folding principles of the human genome', *Science*, 326: 289-93.

Lien, C. L., C. Wu, B. Mercer, R. Webb, J. A. Richardson, and E. N. Olson. 1999. 'Control of early cardiac-specific transcription of Nkx2-5 by a GATA-dependent enhancer', *Development*, 126: 75-84.

Lien, Ching-Ling, John McAnally, James A. Richardson, and Eric N. Olson. 2002. 'Cardiac-Specific Activity of an Nkx2–5 Enhancer Requires an Evolutionarily Conserved Smad Binding Site', *Developmental Biology*, 244: 257-66.

Linder, B., E. Mentele, K. Mansperger, T. Straub, E. Kremmer, and R. A. Rupp. 2007. 'CHD4/Mi-2beta activity is required for the positioning of the mesoderm/neuroectoderm boundary in Xenopus', *Genes Dev*, 21: 973-83.

Liu, Yu, Masanori Asakura, Hironori Inoue, Teruya Nakamura, Motoaki Sano, Zhiyv Niu, Michelle Chen, Robert J. Schwartz, and Michael D. Schneider. 2007. 'Sox17 is essential for the specification of cardiac mesoderm in embryonic stem cells', *Proceedings of the National Academy of Sciences*, 104: 3859-64.

Luna-Zurita, L., C. U. Stirnimann, S. Glatt, B. L. Kaynak, S. Thomas, F. Baudin, M. A. Samee, D. He, E. M. Small, M. Mileikovsky, A. Nagy, A. K. Holloway, K. S. Pollard, C. W. Muller, and B. G. Bruneau. 2016. 'Complex Interdependence Regulates Heterotypic Transcription Factor Distribution and Coordinates Cardiogenesis', *Cell*.

Mansfield, R. E., C. A. Musselman, A. H. Kwan, S. S. Oliver, A. L. Garske, F. Davrazou, J. M. Denu, T. G. Kutateladze, and J. P. Mackay. 2011. 'Plant homeodomain (PHD) fingers of CHD4 are histone H3-binding modules with preference for unmodified H3K4 and methylated H3K9', *J Biol Chem*, 286: 11779-91.

Marfella, C. G., and A. N. Imbalzano. 2007. 'The Chd family of chromatin remodelers', *Mutat Res*, 618: 30-40.

Musselman, C. A., R. E. Mansfield, A. L. Garske, F. Davrazou, A. H. Kwan, S. S. Oliver, H. O'Leary, J. M. Denu, J. P. Mackay, and T. G. Kutateladze. 2009. 'Binding of the CHD4 PHD2 finger to histone H3 is modulated by covalent modifications', *Biochem J*, 423: 179-87.

Musselman, C. A., J. Ramirez, J. K. Sims, R. E. Mansfield, S. S. Oliver, J. M. Denu, J. P. Mackay, P. A. Wade, J. Hagman, and T. G. Kutateladze. 2012. 'Bivalent recognition of nucleosomes by the tandem PHD fingers of the CHD4 ATPase is required for CHD4-mediated repression', *Proc Natl Acad Sci U S A*, 109: 787-92.

Nakamura, Teruya, Motoaki Sano, Zhou Songyang, and Michael D. Schneider. 2003. 'A Wnt- and β-catenin-dependent pathway for mammalian cardiac myogenesis', *Proceedings of the National Academy of Sciences*, 100: 5834-39.

Nishii, K., S. Morimoto, R. Minakami, Y. Miyano, K. Hashizume, M. Ohta, D. Y. Zhan, Q. W. Lu, and Y. Shibata. 2008. 'Targeted disruption of the cardiac troponin T gene causes sarcomere disassembly and defects in heartbeat within the early mouse embryo', *Dev Biol*, 322: 65-73.

Noseda, M., T. Peterkin, F. C. Simoes, R. Patient, and M. D. Schneider. 2011. 'Cardiopoietic factors: extracellular signals for cardiac lineage commitment', *Circ Res*, 108: 129-52.

O'Shaughnessy-Kirwan, A., J. Signolet, I. Costello, S. Gharbi, and B. Hendrich. 2015. 'Constraint of gene expression by the chromatin remodelling protein CHD4 facilitates lineage specification', *Development*, 142: 2586-97.

Ou, H. D., S. Phan, T. J. Deerinck, A. Thor, M. H. Ellisman, and C. C. O'Shea. 2017. 'ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells', *Science*, 357.

Paige, Sharon L, Sean Thomas, Cristi L Stoick-Cooper, Hao Wang, Lisa Maves, Richard Sandstrom, Lil Pabon, Hans Reinecke, Gabriel Pratt, Gordon Keller, Randall T Moon, John Stamatoyannopoulos, and Charles E Murry. 2012. 'A Temporal Chromatin Signature in Human Embryonic Stem Cells Identifies Regulators of Cardiac Development', *Cell*, 151: 221-32.

Papait, Roberto, Paola Cattaneo, Paolo Kunderfranco, Carolina Greco, Pierluigi Carullo, Alessandro Guffanti, Valentina Viganò, Giuliano Giuseppe Stirparo, Michael V. G. Latronico, Gerd Hasenfuss, Ju Chen, and Gianluigi Condorelli. 2013. 'Genome-wide analysis of histone marks identifying an epigenetic signature of promoters and enhancers underlying cardiac hypertrophy', *Proceedings of the National Academy of Sciences*, 110: 20164-69.

Parelho, V., S. Hadjur, M. Spivakov, M. Leleu, S. Sauer, H. C. Gregson, A. Jarmuz, C. Canzonetta, Z. Webster, T. Nesterova, B. S. Cobb, K. Yokomori, N. Dillon, L. Aragon, A. G. Fisher, and M. Merkenschlager. 2008. 'Cohesins functionally associate with CTCF on mammalian chromosome arms', *Cell*, 132: 422-33.

Park, S. H., S. H. Park, M. C. Kook, E. Y. Kim, S. Park, and J. H. Lim. 2004. 'Ultrastructure of human embryonic stem cells and spontaneous and retinoic acid-induced differentiating cells', *Ultrastruct Pathol*, 28: 229-38.

Peterson, C. L. 2000. 'ATP-dependent chromatin remodeling: going mobile', *FEBS Lett*, 476: 68-72.

Peterson, Craig L., and Marc-André Laniel. 2004. 'Histones and histone modifications', *Current Biology*, 14: R546-R51.

Quaranta, Roberto, Jakob Fell, Frank Rühle, Jyoti Rao, Ilaria Piccini, Marcos J. Araúzo-Bravo, Arie O. Verkerk, Monika Stoll, and Boris Greber. 2018. 'Revised roles of ISL1 in a hES cell-based model of human heart chamber specification', *eLife*, 7: e31706.

Rada-Iglesias, Alvaro, Ruchi Bajpai, Tomek Swigut, Samantha A. Brugmann, Ryan A. Flynn, and Joanna Wysocka. 2010. 'A unique chromatin signature uncovers early developmental enhancers in humans', *Nature*, 470: 279.

Rao, Suhas S P., Miriam H Huntley, Neva C Durand, Elena K Stamenova, Ivan D Bochkov, James T Robinson, Adrian L Sanborn, Ido Machol, Arina D Omer, Eric S Lander, and Erez Lieberman Aiden. 2014. 'A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping', *Cell*, 159: 1665-80.

Reyes, J. C., J. Barra, C. Muchardt, A. Camus, C. Babinet, and M. Yaniv. 1998. 'Altered control of cellular proliferation in the absence of mammalian brahma (SNF2alpha)', *Embo j*, 17: 6979-91.

Roche, A. E., B. J. Bassett, S. A. Samant, W. Hong, G. A. Blobel, and E. C. Svensson. 2008. 'The zinc finger and C-terminal domains of MTA proteins are required for FOG-2-mediated transcriptional repression via the NuRD complex', *J Mol Cell Cardiol*, 44: 352-60.

Rottbauer, W., A. J. Saurin, H. Lickert, X. Shen, C. G. Burns, Z. G. Wo, R. Kemler, R. Kingston, C. Wu, and M. Fishman. 2002. 'Reptin and pontin antagonistically regulate heart growth in zebrafish embryos', *Cell*, 111: 661-72.

Rubio, E. D., D. J. Reiss, P. L. Welcsh, C. M. Disteche, G. N. Filippova, N. S. Baliga, R. Aebersold, J. A. Ranish, and A. Krumm. 2008. 'CTCF physically links cohesin to chromatin', *Proc Natl Acad Sci U S A*, 105: 8309-14.

Ryba, T., I. Hiratani, J. Lu, M. Itoh, M. Kulik, J. Zhang, T. C. Schulz, A. J. Robins, S. Dalton, and D. M. Gilbert. 2010. 'Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types', *Genome Res*, 20: 761-70.

Schlange, Thomas, Birgit Andrée, Hans-Henning Arnold, and Thomas Brand. 2000. 'BMP2 is required for early heart development during a distinct time period', *Mechanisms of Development*, 91: 259-70.

Schoenfelder, Stefan, Mayra Furlan-Magaril, Borbala Mifsud, Filipe Tavares-Cadete, Robert Sugar, Biola-Maria Javierre, Takashi Nagano, Yulia Katsman, Moorthy Sakthidevi, Steven W. Wingett, Emilia Dimitrova, Andrew Dimond, Lucas B. Edelman, Sarah Elderkin, Kristina Tabbada, Elodie Darbo, Simon Andrews, Bram Herman, Andy Higgs, Emily LeProust, Cameron S. Osborne, Jennifer A. Mitchell, Nicholas M. Luscombe, and Peter Fraser. 2015. 'The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements', *Genome Res*, 25: 582-97.

Stopka, T., and A. I. Skoultchi. 2003. 'The ISWI ATPase Snf2h is required for early mouse development', *Proc Natl Acad Sci U S A*, 100: 14097-102.

Sutcliffe, E. L., I. A. Parish, Y. Q. He, T. Juelich, M. L. Tierney, D. Rangasamy, P. J. Milburn, C. R. Parish, D. J. Tremethick, and S. Rao. 2009. 'Dynamic histone variant exchange accompanies gene induction in T cells', *Mol Cell Biol*, 29: 1972-86.

Szalaj, P., and D. Plewczynski. 2018. 'Three-dimensional organization and dynamics of the genome', *Cell Biol Toxicol*.

Tang, Z., O. J. Luo, X. Li, M. Zheng, J. J. Zhu, P. Szalaj, P. Trzaskoma, A. Magalska, J. Wlodarczyk, B. Ruszczycki, P. Michalski, E. Piecuch, P. Wang, D. Wang, S. Z. Tian, M. Penrad-Mobayed, L. M. Sachs, X. Ruan, C. L. Wei, E. T. Liu, G. M. Wilczynski, D. Plewczynski, G. Li, and Y. Ruan. 2015. 'CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription', *Cell*, 163: 1611-27.

Tevosian, S. G., A. E. Deconinck, M. Tanaka, M. Schinke, S. H. Litovsky, S. Izumo, Y. Fujiwara, and S. H. Orkin. 2000. 'FOG-2, a cofactor for GATA transcription factors, is

essential for heart morphogenesis and development of coronary vessels from epicardium', *Cell*, 101: 729-39.

Ueno, Shuichi, Gilbert Weidinger, Tomoaki Osugi, Aimee D. Kohn, Jonathan L. Golob, Lil Pabon, Hans Reinecke, Randall T. Moon, and Charles E. Murry. 2007. 'Biphasic role for Wnt/β-catenin signaling in cardiac specification in zebrafish and embryonic stem cells', *Proceedings of the National Academy of Sciences*, 104: 9685-90.

Wade, P. A., P. L. Jones, D. Vermaak, and A. P. Wolffe. 1998. 'A multiple subunit Mi-2 histone deacetylase from Xenopus laevis cofractionates with an associated Snf2 superfamily ATPase', *Curr Biol*, 8: 843-6.

Wade, PA, A Gegonne, PL Jones, E Ballestar, F Aubry, and AP Wolffe. 1999. 'Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation', *Nat Genet*, 23: 62 - 66.

Waldron, L., J. D. Steimle, T. M. Greco, N. C. Gomez, K. M. Dorr, J. Kweon, B. Temple, X. H. Yang, C. M. Wilczewski, I. J. Davis, I. M. Cristea, I. P. Moskowitz, and F. L. Conlon. 2016. 'The Cardiac TBX5 Interactome Reveals a Chromatin Remodeling Network Essential for Cardiac Septation', *Dev Cell*, 36: 262-75.

Wamstad, Joseph A, Jeffrey M Alexander, Rebecca M Truty, Avanti Shrikumar, Fugen Li, Kirsten E Eilertson, Huiming Ding, John N Wylie, Alexander R Pico, John A Capra, Genevieve Erwin, Steven J Kattman, Gordon M Keller, Deepak Srivastava, Stuart S Levine, Katherine S Pollard, Alisha K Holloway, Laurie A Boyer, and Benoit G Bruneau. 2012. 'Dynamic and Coordinated Epigenetic Regulation of Developmental Transitions in the Cardiac Lineage', *Cell*, 151: 206-20.

Wang, Li, Ying Du, James M Ward, Takashi Shimbo, Brad Lackford, Xiaofeng Zheng, Yi-liang Miao, Bingying Zhou, Leng Han, David C Fargo, Raja Jothi, Carmen J Williams, Paul A Wade, and Guang Hu. 2014. 'INO80 Facilitates Pluripotency Gene Activation in Embryonic Stem Cell Self-Renewal, Reprogramming, and Blastocyst Development', *Cell Stem Cell*, 14: 575-91.

Wang, Q., R. S. Reiter, Q. Q. Huang, J. P. Jin, and J. J. Lin. 2001. 'Comparative studies on the expression patterns of three troponin T genes during mouse development', *Anat Rec*, 263: 72-84.

Wen, Bo, Hao Wu, Yoichi Shinkai, Rafael A. Irizarry, and Andrew P. Feinberg. 2009. 'Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells', *Nat Genet*, 41: 246.

Yip, D. J., C. P. Corcoran, M. Alvarez-Saavedra, A. DeMaria, S. Rennick, A. J. Mears, M. A. Rudnicki, C. Messier, and D. J. Picketts. 2012. 'Snf2l regulates Foxg1-dependent progenitor cell expansion in the developing brain', *Dev Cell*, 22: 871-8.

Zhang, Y, G LeRoy, HP Seelig, WS Lane, and D Reinberg. 1998. 'The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities', *Cell*, 95: 279 - 89.

Zhu, L., G. E. Lyons, O. Juhasz, J. E. Joya, E. C. Hardeman, and R. Wade. 1995. 'Developmental regulation of troponin I isoform genes in striated muscles of transgenic mice', *Dev Biol*, 169: 487-503.

CHAPTER 2: CHD4 and the NuRD complex directly control cardiac sarcomere formation[1]

Introduction

Congenital heart disease remains the most common congenital malformation, and as such, attaining a mechanistic understanding of cardiomyocyte formation is crucial for improving outcomes to structural heart disease (Heron et al. 2009; Dolk, Loane, and Garne 2010). Though much emphasis in the last few years has been placed on transcription factor networks that control cardiomyocyte differentiation, these studies have largely focused on transcriptional activation. However, there is growing recognition that alterations in transcriptional repression also lead to congenital heart disease (Waldron et al. 2016; Homsy et al. 2015; Zaidi et al. 2013). Transcriptional repression involves not only cardiac transcription factors but also broadly expressed multiprotein machines that modify and remodel chromatin. Prominent among these is the Nucleosome Remodeling and Deacetylase (NuRD) complex.

The Nucleosome Remodeling and Deacetylase (NuRD) complex is one of the major transcriptional complexes that function to repress gene expression. The

[1]This chapter previously appeared as an article in the journal *Proceedings of the National Academy of Sciences*. The original citation is as follows: Wilczewski CM, Hepperla AJ, Shimbo T, Wasson L, Robbe ZL, Davis IJ, Wade PA, Conlon FL. "CHD4 and the NuRD complex directly control cardiac sarcomere formation," *Proceedings of the National Academy of Science* (June 2018).

NuRD complex has been reported in most instances to function as a chromatin modifying complex and been demonstrated to act by combining histone deacetylase activity with an ATP-dependent chromatin remodeling helicase to modulate chromatin states at target genes (Wade et al. 1999; Zhang et al. 1998; Xue et al. 1998; Wade et al. 1998). The NuRD complex is essential for numerous developmental events, including ensuring proper timing of the switch from stem-cell lineages to differentiated cell types, maintaining cell differentiation, and activating DNA damage response pathways (Zhang 2011; O'Shaughnessy-Kirwan et al. 2015; Hung, Kohnken, and Svaren 2012; Kashiwagi, Morgan, and Georgopoulos 2007; Williams et al. 2004; Yoshida et al. 2008; Luo et al. 2013).

The components of the NuRD complex can vary but it is typically composed of the ATP-dependent chromodomain helicase DNA-binding protein (CHD) 3/4, histone deacetylase (HDAC) 1/2, metastasis-associated protein (MTA) 1/2/3, retinoblastoma binding protein (RBBP) 4/7 (also known as RbAp48/46), GATAD2A/B, and the mCpG-binding domain protein (MBD) 2/3 (Zhang et al. 1998; Xue et al. 1998; Wade et al. 1998; Marhold et al. 2004; Kim et al. 1999). NuRD complex target specificity can be conferred by the association of components of the NuRD complex with tissue-specific co-factors that target the complex to a defined set of loci. Factors include three proteins associated with congenital heart disease: FOG-2, TBX5 and TBX20 (Waldron et al. 2016; Garnatz et al. 2014; Roche et al. 2008; Aguayo-Gomez et al. 2015; Basson et al. 1997; Bruneau et al. 1999; Bruneau et al. 2001; Kirk et al. 2007; Li et al. 1997; Mori and Bruneau 2004; Yoshida et al. 2016; Brown et al. 2005; Kaltenbrun et al. 2013; Yang et al. 2000; Svensson et al.

2000). Consistently, mutations in *CHD4* have been found to be causative to congenital heart disease including atrial and ventricular septal defects (Homsy et al. 2015). Although the functions of the NuRD complex have been studied in a limited context *in vivo,* its role in cardiac development has yet to be defined and no studies to date have directly addressed the requirement or mechanism for CHD4 in cardiac tissue.

Here we report CHD4 is essential for cardiac development as mice cardiac conditionally null for *Chd4* die during mid-gestation. By performing a systems-level analysis of CHD4 target genes combined with temporal transcriptional profiling, we provide evidence CHD4 directly binds proximal-promoter and distal gene regulatory elements to directly repress many fast skeletal and smooth muscle myofibril genes. Moreover, we find mis-expression of fast skeletal and smooth muscle myofibril genes in *Chd4* null embryos is direct and not associated with a reactivation of the embryonic skeletal or embryonic smooth muscle program. We report skeletal and smooth muscle proteins are incorporated into cardiomyocytes forming a hybrid of all three muscle types. Using a non-invasive *in utero* embryonic echocardiography technique, we show expression of all three muscle types impairs cardiomyocyte function leading to a decrease in blood flow and ultimately, embryonic lethality. Collectively these studies define molecular, biochemical, anatomical, and physiological mechanisms for CHD4 and the NuRD complex in repressing the inappropriate expression of the skeletal and smooth muscle programs in the developing heart.

Results

CHD4 is required for cardiac development and myocardial growth

To determine the requirement for CHD4 in heart development, we generated *Chd4* cardiac conditional null mice, *Chd4*$^{\Delta flox/\Delta flox}$, by mating *Chd4*$^{flox/flox}$ female mice to *Chd4*$^{flox/+}$;*Nkx2-5*$^{Cre/+}$ male mice (Moses et al. 2001; Williams et al. 2004). Heterozygote (*Chd4*$^{\Delta flox/+}$) mice were viable, fertile and displayed no obvious phenotypic abnormalities. By contrast, no mice homozygous for *Chd4* ablation were recovered postnatally. Analysis of timed intercrosses to generate *Chd4*$^{\Delta flox/\Delta flox}$ mice failed to identify viable homozygous *Chd4*$^{\Delta flox/\Delta flox}$ embryos subsequent to embryonic day (E)12.5 (Table S2.1).

At E11.5 we observed *Chd4*$^{\Delta flox/\Delta flox}$ mice were viable but displayed pericardial edema, pericardial hemorrhage and stunted growth compared to *Chd4*$^{flox/flox}$ (no Cre recombinase) littermate controls (Figure S1.1A-D). Ultrastructural analysis showed *Chd4*$^{flox/flox}$ and *Chd4*$^{\Delta flox/\Delta flox}$ hearts initiate cardiac chamber formation at E10.5 (Figure 2.1A,B). However, by E11.5 *Chd4*$^{\Delta flox/\Delta flox}$ hearts exhibited hallmarks of cardiac failure including enlarged left and right atria, a reduced right ventricle and an enlarged left ventricle (Figure 2.1C,D) (Conway et al. 2003). Histological analysis on *Chd4*$^{\Delta flox/\Delta flox}$ and *Chd4*$^{flox/flox}$ hearts at E10.5 revealed a decrease in complexity of the trabecular layer of the right and left ventricles and a significant decrease in the thickness of the compact layer by E10.5 (Figure S2.1E-I). This myocardial growth defect was concurrent with a decrease in the mitotic index (Figure S2.1J). However, there was no increase in the levels of apoptosis (Figure S2.1K) or any change in the number of endocardial cells relative to cardiomyocytes (Figure S2.2). Since we confirmed *Chd4*$^{\Delta flox/\Delta flox}$ mice lack

any cardiac CHD4 protein at E9.5 (Figure S2.3), these data imply CHD4 is required for cardiac development at or prior to E10.5.

CHD4 regulates transcription of the skeletal- and smooth muscle-specific programs in the developing heart

To explore the molecular mechanism by which CHD4 functions, we performed transcriptomic analysis (RNA-seq) on E9.5 and E10.5 hearts to reflect states prior to and at the early stage of the observed cardiac defects in *Chd4* null hearts (Slagle and Conlon). Comparing transcript abundances in the presence (*Chd4*$^{flox/flox}$) or absence (*Chd4*$^{\Delta flox/\Delta flox}$) of CHD4 revealed 1285 differentially expressed genes at E9.5 and 1318 differentially expressed genes at E10.5 (adjusted p-value<0.05, $\log_2$(fold change) ≥±0.5) (Figure S2.4A,B). In agreement with the primary role of the NuRD complex as a transcriptional repressor, nearly three times as many genes were upregulated in the absence of CHD4 relative to downregulated genes (913 upregulated versus 372 downregulated at E9.5; 920 upregulated versus 398 downregulated at E10.5, Figure S2.4C,D). Of these genes, 327 were coordinately upregulated, or shared between E9.5 and E10.5, while 65 were coordinately downregulated at both E9.5 and E10.5 (Figure S2.4E,F). These genes tended to be those with the highest degree of change in either dataset. Gene ontology (GO) analyses were performed to investigate the roles and pathways of differentially expressed genes in *Chd4* null hearts. Surprisingly, the most significant over-represented biological process associated with genes transcriptionally regulated by CHD4 was striated muscle contraction (Figure 2.1E) (Mi et al. 2017).

In cardiomyocytes at early embryonic stages (E8.5-E9.5) myofibril subunits become organized and function as a contractile apparatus that will ultimately develop into mature cardiac sarcomeres. These sarcomeres bear contractile stress to drive the heartbeat and thus circulate nutrients and oxygen throughout the growing embryo (Hirschy et al. 2006). Contractility is achieved by cardiomyocyte-specific myofibril subunits, including α-actin (cardiac *Actc1* and to a lesser degree skeletal *Acta1*), β-myosin heavy chain (β-MHC, *Myh7*) and the troponin complex proteins: cardiac/slow skeletal troponin C1 (*Tnnc1*), cardiac troponin TnT2 (*Tnnt2*), and slow skeletal troponin TnI1 (*Tnni1* and to a low degree TnI2, *Tnni2*) (Nishii et al. 2008; Wang et al. 2001; England and Loughna 2013; Saggin et al. 1989; Ilkovski et al. 2005). We found *Chd4* null hearts deviate from this normal developmental gene expression pattern by upregulating the non-cardiac paralogs for all of these genes, including smooth muscle *Myh11* and fast skeletal *Acta1, Tnnc2, Tnnt3* and *Tnni2* (Figure 2.1F). We confirmed misexpression of non-cardiac myofibril isoforms by quantitative PCR (RT-qPCR) (Figure S2.5). Interestingly, aberrant expression of these fast skeletal and smooth muscle paralogs is not accompanied by any significant misexpression of skeletal or smooth muscle transcriptional regulators (Figure 2.1F). CHD4 ablation did not significantly alter expression of α-smooth muscle actin (*Acta2*) or the majority of cardiac myofibril subunits (Figure 2.1F, Figure S1.6) (Clement et al. 2007). We further confirmed increased fast skeletal TnI2 and smooth-muscle myosin heavy chain protein levels by immunohistochemistry (Figure 2.1G-DD) (antibody specificity demonstrated in Figure S2.7).

To address whether the misexpression of non-cardiac myofibril isoforms was due to CHD4-mediated repression in cardiomyocytes, or a stress response to hemodynamic forces, we conditionally ablated *Chd4* with *Tnnt2-cre*. Results phenocopied that of *Nkx2-5*$^{Cre/+}$ in the misexpression of TnI2 and smooth muscle myosin heavy chain (*Myh11)* (Figure S2.8). Taken together, these studies imply in the absence of CHD4, cardiomyocytes form a hybrid of cardiac, skeletal and smooth muscle.

CHD4 binds proximal gene elements to regulate myofibril assembly

To identify genes directly regulated by CHD4, we performed chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) for CHD4 using embryonic hearts collected at E10.0. To our knowledge, this is the first report of CHD4 ChIP-seq using non-cultured *in vivo* tissue. We identified 43,818 regions of signal enrichment and found that CHD4 preferentially localizes to proximal-promoter regions (16.0% over 0.9% baseline genome composition) and introns (40.8% over 33.8% baseline genome composition) (Figure 2.2A, right panel compared to baseline genome composition in left panel).

To elucidate the relationship between CHD4 binding and the previously identified transcriptional changes, CHD4 ChIP-seq peaks were assigned genes by computationally predicted association (McLean et al. 2010). GO analysis of these genes revealed striking concordance with processes predicted to be transcriptionally regulated by CHD4, specifically sarcomere organization, striated muscle contraction and myofibril assembly (Figure 2.2B). This prompted us to compare genes transcriptionally misregulated in *Chd4*$^{\Delta flox/\Delta flox}$ hearts with genes predicted to contain CHD4 ChIP-seq

peaks. Suggestive of direct CHD4-mediated transcriptional regulation, we find a significant enrichment for genes predicted to be directly bound by CHD4 in genes up or downregulated in the absence of CHD4 ($p<0.001$) (Figure S2.9).

For the E9.5 dataset, 71.4% (652 of 913) of upregulated genes and 86.3% (321 of 372) of downregulated genes associated with at least one CHD4 peak; 81.1% (746 of 920) upregulated and 83.4% (333 of 398) downregulated genes were associated with at least one CHD4 peak at E10.5. Genes coordinately up- or downregulated (shared) across E9.5 and E10.5 were also significantly enriched for genes predicted to be directly bound by CHD4; 74.3% (243 of 327) and 84.6% (55 of 65) of genes associating with at least one CHD4 peak respectively (Figure S2.9). For example, CHD4 peaks are present at regions predicted to associate with the cardiac *Mylk3* gene, one of two cardiac sarcomere subunits downregulated in the absence of CHD4 (Figure 2.1F, Figure S2.10). These data are in agreement with studies that have shown CHD4 and the NuRD complex can function in both transcriptional activation and repression (Shimbo et al. 2013; Williams et al. 2004; Yamada et al. 2014; Hung, Kohnken, and Svaren 2012; Miccio et al. 2010; Allen, Wade, and Kutateladze 2013).

As CHD4 appears to directly influence gene regulation, we queried if there was a difference in the magnitude of regulation based on the location of CHD4 binding, specifically between genes with at least one proximal-promoter site (within 1500 base pairs [bp] upstream and 500bp downstream of transcriptional start sites [TSSs]) compared to genes with only distal sites. We observed genes associated with only distal binding sites demonstrated greater differential mis-regulation in the absence of CHD4 relative to genes with at least one proximal-promoter binding site (Figure 2.2C). This

unexpected finding may be due to CHD4 localizing at distal active enhancers, as has been previously reported for the NuRD complex (Shimbo et al. 2013). Collectively, these data demonstrate CHD4 directly transcriptionally regulates cardiac muscle cell development when cardiac function becomes indispensable for continued embryonic growth.

CHD4 coordinates a transcriptional network to repress non-cardiac myofibril gene expression

Gene annotation of CHD4-bound regions and visual inspection of browser tracks identified CHD4 as directly bound to regulatory regions for fast skeletal and smooth muscle myofibril paralogs *Myh11*, *Acta1, Tnnt3*, *Tnnc2,* and *Tnni2*, (Figure 2.3A-F). Combined with the previous transcriptomic analyses, these data suggest loss of CHD4 causes de-repression of the skeletal and smooth muscle program in cardiomyocytes.

Misexpression of non-cardiac myofibril paralogs leads to sarcomere disarray and impaired cardiac function

The observation that CHD4 suppresses non-cardiac myofibril paralogs led us to hypothesize that misexpression of fast skeletal and smooth muscle myofibril paralogs affects sarcomere organization and ultimately, cardiac function in the developing embryo. To test this hypothesis, we analyzed sarcomere formation in the absence of CHD4-mediated transcriptional regulation by transmission electron microscopy. By E10.5 control ventricular cardiomyocytes organize myofibril subunits into discrete sarcomere units in series that are anchored to each other at the Z-disk (Figure 2.4A,

wedge arrows). To provide contractile force, units are arranged in parallel in a higher order structure observed by the alignment of the Z-disks across the muscle (Figure 2.4B, arrows). In contrast, *Chd4*$^{\Delta flox/\Delta flox}$ hearts show a severe reduction in sarcomere organization and formation of Z-disks. This coincides with a decrease in the parallel arrangement of sarcomeres (Figure 2.4C, D). These defects are further associated with a significant decrease in sarcomere organization in *Chd4*$^{\Delta flox/\Delta flox}$ hearts as quantified by the degree of Z-disk and A-band alignment (Figure S2.11).

To determine if the skeletal and smooth muscle proteins are incorporated into the developing cardiac muscle, we analyzed co-localization of a cardiac sarcomere Z-disk protein, α-actinin, with mis-expressed smooth-muscle myosin heavy chain. Fluorescence intensity was measured and quantified over the length of individual sarcomeres. Analysis revealed smooth muscle myosin heavy chain protein organizes into nascent sarcomere in *Chd4*$^{\Delta flox/\Delta flox}$ cardiomyocytes (Figure 2.4E-L). Taken together, our data imply that by intercalating into the nascent cardiac sarcomere, non-cardiac myofibril isoforms displace normal cardiac sarcomere proteins during myofibril assembly.

Co-expression of cardiac, smooth muscle and skeletal muscle paralogs compromises cardiac contractility and function

To determine the physiological consequences of sarcomere disorganization as a result of fast skeletal and smooth muscle myofibril protein mis-expression on cardiac function, we developed non-invasive *in utero* embryonic echocardiography methodologies. We performed ultrasound pulsed-wave (PW) Doppler on E10.5

littermates *in utero* without surgical manipulation of the dam or embryos. This approach enabled us to measure the effect of sarcomere malformation on cardiac function in the context of the developing heart *in situ* while avoiding artificial manipulation of cardiac fluid dynamics or maternal stress. PW Doppler recordings of E10.5 control hearts demonstrate consistently strong atrial and ventricular contractions (Figure 2.4M, Supplemental Video 2.1). In stark contrast, *Chd4*$^{\Delta flox/\Delta flox}$ hearts with significant cardiac sarcomere disarray show severely reduced ventricular contractions with significant decreases in ventricular outflow peak velocity and velocity time integral (Figure 2.4N-P and Supplemental Video 2.2). Collectively, these data demonstrate CHD4 is required to repress expression of non-cardiac myofibril paralogs in the developing heart. In the absence of CHD4, mis-expressed fast skeletal and smooth muscle myofibril paralogs intercalate into the cardiac sarcomere resulting in impaired cardiac function at the point at which cardiac function becomes indispensable for continued embryonic growth.

Discussion

Here we demonstrate CHD4 and the NuRD complex repress both smooth muscle and fast skeletal myofibril paralogs in the developing heart. We find the consequences of activating and incorporating skeletal or smooth muscle myofibrils in cardiomyocytes is a gross disorganization of the sarcomere, a failure of proper contraction and ultimately death of the embryo. There are multiple functional differences between these cell types. For example, cardiac and skeletal muscle have differences in calcium cycling and sensitivity of the troponin complex, as well as cooperativity and activation of the thin filament (Yang et al. 2009). In contrast to cardiac muscle, smooth

muscle cells do not primarily rely on a troponin complex-mediated system of contraction, and instead utilize ATP-mediated phosphorylation of the myosin head to induce contraction. Furthermore, unlike cardiac and skeletal muscle, smooth muscle cell myofibrils do not arrange into a striated sarcomere structure and instead organize into an oblique actomyosin cytoarchitecture (Craig and Woodhead 2006). In striated muscle cells, such as developing cardiomyocytes, myosin heavy chain intercalation is particularly crucial for formation of the thick filament and the associated M-line structure, which is then anchored in alignment between structural Z-disks by titin. Substitution of smooth muscle myosin heavy chain into the nascent cardiac sarcomere may therefore be causing the sarcomere formation defects observed in CHD4-null cardiomyocytes, while incorporation of the fast skeletal sarcomere paralogs may further explain the impaired cardiac function in the absence of CHD4-mediated transcriptional regulation.

While some studies have reported misexpression of skeletal or smooth muscle myofibril components in the developing heart at the transcript level, their presence at the protein level had not yet been confirmed (Gómez-del Arco et al. ; Heidersbach et al. 2013). Furthermore, the mechanism by which misexpression leads to cardiac developmental defects had not yet been explored. In this report, we have demonstrated that in the absence of CHD4-mediated repression, misexpression of smooth muscle myosin heavy chain, fast skeletal α-actin, and the fast skeletal troponin complex leads to sarcomere disarray and impaired cardiac function. Collectively, this study begins to answer how misexpression of skeletal and smooth muscle myofibril subunits during cardiac development negatively affects sarcomere formation and cardiac function.

We further identified the genomic loci targeted by CHD4 to repress non-cardiac myofibril paralogs through an unbiased approach. While some genomic regions have been documented as contributing to repression of these paralogs, these studies have not been performed in a comprehensive fashion using endogenous tissue at the relevant developmental stages of cardiac sarcomere formation (Gómez-del Arco et al. ; Montgomery et al. 2007). Also, the importance of cis-regulatory regions in regulating cardiac development and disease has only recently been emphasized (Postma, Bezzina, and Christoffels 2016). The regulatory regions discovered here thus represent a significant step forward in identifying sites that are crucial for transcriptional regulation and normal heart development.

There is misexpression of many genes in the absence of CHD4-mediated repression in the developing heart. However, at this stage of cardiac development in mouse, continued embryonic growth is dependent on the formation of functional cardiac sarcomere units to initiate systolic function. Our data support a model in which CHD4 loss impedes this process during early heart development. This poses an intriguing question whether in humans impaired cardiac systolic function associated with certain cardiomyopathies or cardiac failure, in the absence of mutations in cardiac sarcomere subunits, may be due to improper expression and intercalation of non-cardiac myofibril paralogs in the cardiac sarcomere. Screening for misexpression of non-cardiac myofibril paralogs in cardiac tissue, or for mutations in these putative regulatory elements that may impair CHD4 recruitment or activity would address this hypothesis and suggest interventions for these patients.

Materials and Methods

Mice

Chd4$^{flox/flox}$ mice were obtained from Dr. Katia Georgeopolos(Williams et al. 2004). *Nkx2-5*$^{Cre/+}$ mice were obtained from Robert Schwartz (Moses et al. 2001). *Tg(Tnnt2-cre)5Blh*/*JiaoJ* (strain 024240) mice were obtained from JAX. *Chd4* conditional knockout mice and control littermates were obtained by breeding female *Chd4*$^{flox/flox}$ mice to male *Chd4*$^{flox/+}$*;Nkx2-5*$^{Cre/+}$ mice. Research was approved by the Institutional Animal Care and Use Committee at the University of North Carolina and conforms to the Guide for the Care and Use of Laboratory Animals.

Histology and Immunohistochemistry

Embryos were fixed in 4% paraformaldehyde/0.1% Tween-20 overnight at 4°C. Embryos were prepared for histology and immunohistochemistry as previously described (Dorr et al. 2015). 10μm paraffin sections were stained with hematoxylin and eosin using standard practices and were imaged on an Olympus BX61 microscope with a Retiga 400R camera. Immunohistochemistry was performed with antigen retrieval on 10μm cryosections as previously described (Dorr et al. 2015) with the following primary antibodies: mouse anti-tropomyosin (DSHB clone CH1), 1:50; rabbit anti-CHD4 (Abcam #ab72418) 1:500; rabbit anti-smooth muscle myosin heavy chain (Abcam ab53219), 1:50; mouse anti-α-actinin (Sigma #A7732), 1:500; rabbit anti-phospho-histone H3 (Millipore #06-570), 1:200. Immunohistochemistry with rabbit anti-troponin I2 (Abcam #ab183508), 1:250; rat anti-PECAM1 (CD31) (BD Biosciences #553370), 1:50 and rabbit anti-cleaved caspase 3 (Cell Signalling #9661), 1:50 did not require antigen

retrieval. Secondary antibodies were Alexa Fluor 488 goat anti-mouse IgG H+L (Thermo #A11001), 1:1000; Alexa Fluor 546 goat anti-mouse IgG_1 (Thermo #A21123), 1:1000; Alexa Fluor 647 goat anti-rabbit IgG H+L (Thermo #A21245), 1:1000. Immunohistochemistry images were captured on a Zeiss LSM 700 laser scanning confocal microscope. Whole-mount images were captured using a Leica MZ 16F dissection microscope with a Retiga 4000RV camera. ImageJ (NIH) was used for image analysis and standard image processing.

Image Quantification

Thickness of the compact layer was quantified using hematoxylin and eosin stained paraffin with ImageJ on three separate regions per ventricle on three sections (anterior, mid, and posterior heart) per embryo, from three embryos per genotype for total N=243 measurements per genotype. Measurements were averaged separately for each ventricle. Sarcomere organization was quantified by classifying sarcomere structure by α-actinin staining in twenty-five 2μm vectors per field, with three fields per ventricle measured in three embryos per genotype for total N=225 vectors per ventricle per genotype. Statistical comparison between genotypes was performed by χ^2 test. Z-stacks of smooth muscle myosin heavy chain and α-actinin double-stained sections were deconvolved using AutoQuantX3 software (Media Cybernetics). Fluorescent signals for each channel were measured on 10 vectors per embryo from three embryos per genotype in ImageJ for total N=30 vectors. Correlations between signals were determined using Spearman's rank correlation. Percentage of pHH3 or cleaved caspase 3 positive cardiomyocytes were assessed by counting pHH3 or cleaved caspase 3

positive cells co-stained with tropomyosin from three separate sections per embryo, three independent embryos per genotype for N=9 whole sections counted per genotype. Percentage of PECAM1+/tropomyosin+ cells was calculated by counting PECAM1 positive cells and dividing by tropomyosin positive cells from three separate sections per embryo, three independent embryos per genotype for N=9 whole sections counted per genotype. Statistical comparison between genotypes was performed by Student's t-test except where noted.

Scanning electron microscopy (SEM)

A standardized procedure for SEM for the heart was used (Dorr et al. 2015). Briefly, the pericardium was removed before embryos were fixed in 2% paraformaldehyde/2.5% glutaraldehyde in 1X PBS (EM grade, Electron Microscopy Services). Embryos were washed in 1X PBS, dehydrated into 100% ethanol and subject to critical point drying. Dried specimens were mounted ventral side up and ion sputtered with gold palladium to 20nm thickness before scanning with Zeiss Supra 25 FESEM microscope with accelerating voltage of 5000V. SEM photomicrographs were taken in standard orientations and magnifications.

RNA-sequencing and Analysis

E10.5 hearts were collected from three *Chd4*$^{flox/flox}$ and three *Chd4*$^{\Delta flox/\Delta flox}$ embryos. E9.5 hearts were collected from five *Chd4*$^{flox/flox}$ and five *Chd4*$^{\Delta flox/\Delta flox}$ embryos. RNA was isolated from individual hearts using standard Trizol extraction. Poly-A selected RNA-seq libraries were prepared at the Vanderbilt Genomic Core using

standard library preparation protocols for mRNA. Samples were run on a HiSeq2500 (Illumina) with 50 base-pair single-end reads.

Fastq files were first filtered for reads containing adapter sequences using TagDust (v1.12) (Lassmann, Hayashizaki, and Daub 2009) with an FDR of 0.001. Reads were then mapped to the mm10 genome using Bowtie (v1.2.0) (Langmead et al. 2009) and options: -m 1, --best, --seed=123, and --nomaqround. Post-alignment, samtools (v1.6) (Li et al. 2009) and BedTools (v2.25.0) (Quinlan 2014) were used to interconvert and manipulate files. HTSeq (v0.6.2) (Anders, Pyl, and Huber 2015), using the refFlat gtf from UCSC Genome Browser (Kent et al. 2002), was run to get read counts over genes and these counts were used in DESeq2 (v1.6.3) (Love, Huber, and Anders 2014) to determine differential transcripts. Genes with an adjusted p-value < 0.05 and a $\log_2$(Fold Change) > 0.5 in either direction. Gene ontology (GO) terms were analyzed for genes with an adjusted p-value<0.05 using PANTHER (v12.0) Biological Process using Fisher's Exact with FDR multiple test correction (Mi et al. 2007). Morpheus was used for heatmap visualization (http://software.broadinstitute.org/morpheus/).

Reverse Transcription and Quantitative PCR

RNA from individual E10.5 hearts (N=3 per genotype) was isolated using Trizol (Invitrogen) and purified on RNeasy columns (Qiagen). cDNA synthesis was performed using random primers and SuperScript II reverse transcriptase (Invitrogen). Quantitative PCR was performed using the SensiFAST SYBR Hi-ROX Kit (Bioline) with the following primers: Myh11F: GCTAATCCACCCCCGGAGTA, Myh11R:

TCGCTGAGCTGCCCTTTCT; Acta1F: GTGACCACAGCTGAACGTG, Acta1R: CCAGGGAGGAGGAAGAGG; Tnnt3F: ACAGATTGGCGGAGGAGAAG, Tnnt3R: CATGGAGGACAGAGCCTTTT; Tnnc2F: GAGATGATCGCTGAGTTCAAG, Tnnc2R: GTCTGCCCTAGCATCCTCAT; Tnni2F: CACCTGAAGAGTGTGATGCTC, Tnni2R: GGGCAGTGTTCTGACAGGTAG; RPS29F: TGGGTCACCAGCAGCTCTAC, RPS29R: GTACTGCCGGAAGCACTGG.

Chromatin Immunoprecipitation (ChIP)-sequencing and Analysis

E10.0 hearts from wild-type CD1 mice (thirty hearts per biological replicate, two biological replicates with three technical replicates total) were fixed for chromatin immunoprecipitation as described previously (Shimbo et al. 2013). Briefly, hearts were dissected in cold PBS and crosslinked with 1.5mM disuccinimidyl glutarate (Thermo) for 45 minutes at room temperature followed by 1% paraformaldehyde (Electron Microscopy Services) for 10 minutes at room temperature. Fixation was quenched with 125mM glycine. Hearts were washed twice with cold PBS, snap frozen and stored at -80°C. Hearts were prepared for ChIP as described (Shimbo, Takaku, and Wade 2016) with some modifications. Hearts were resuspended in TD600 buffer (10mM Tris pH7.4, 2mM $MgCl_2$, 600mM NaCl, 1% Triton-100, 0.1% sodium deoxycholate, protease inhibitor (Roche)), dounced 20 times on ice, and sonicated in a Covaris S220 with the following settings: 145 watts, 20% duty cycle, 200 counts for 1200 seconds at 4°C. Following sonication, samples were centrifuged at 16.1x*g* at 4°C for 5 minutes. The soluble chromatin fraction was diluted 1:4 with TD0 buffer (10mM Tris pH7.4, 2mM $MgCl_2$, 1% Triton-100, 0.1% sodium deoxycholate, 1X protease inhibitor). 10% of the

input was stored at -80°C for input library sequencing. 800µl of the chromatin fraction was mixed with 3µg mouse anti-CHD4 antibody (Abcam #ab70469) and rotated overnight at 4°C. 225µg each Protein A and Protein G Dynabeads (Life Technologies) were blocked overnight at 4°C with 2.6µg/µl bovine serum albumin (Jackson Immuno). The next day, blocked beads were rotated with the chromatin immunoprecipitate for 45 minutes at 4°C, then washed three times with SDS-free RIPA buffer (25mM Tris-HCl pH7.4, 150mM NaCl, 1% NP-40, 1% sodium deoxycholate) followed by three washes with high-salt RIPA buffer (25mM Tris-HCl pH7.4, 600mM NaCl, 1% NP-40, 1% sodium deoxycholate), followed by two washes with TM2 buffer (10mM Tris pH7.4, 2mM $MgCl_2$). The beads were prepared for ChIPmentation using the Nextera XT DNA Library Preparation Kit (Illumina) with the following modifications: 10µl Resuspension Buffer, 20µl Tagment DNA Buffer and 10µl Amplicon Tagment Mix with incubation at 37°C for 15 minutes in a thermal cycler (Schmidl et al. 2015). Beads were washed with SDS-free RIPA buffer, then incubated in 100µl SDS-free RIPA with 40µg Proteinase K at 65°C for 45 minutes. DNA was purified with Zymo DNA Clean & Concentrator-5 columns (Zymo Research). CHD4 ChIP and input libraries were prepared using the Nextera XT DNA Library Kit and purified using the Agencourt AMPure XP PCR Purification Kit (Beckman Coulter). Libraries were sequenced using the NextSeq 500 at the National Institute of Environmental Health Sciences Epigenomics Core Facility with 36 base pair paired-end reads per sample. Fastq files were first filtered for reads containing adapter sequences using TagDust (v1.12) (Lassmann, Hayashizaki, and Daub 2009) with an FDR of 0.001. In-house scripts were used to ensure TagDust filtered reads were maintained in the proper order for alignment. Reads were then mapped to the mm10 genome using

Bowtie (v1.2.0) (Langmead et al. 2009) using options: -m 1, --best, --seed=123, --nomaqround. Aligned fragments were then filtered for any duplicate reads based on alignment position using in-house scripts. Samtools (v1.6) (Li et al. 2009) and BedTools (v2.25.0) (Quinlan 2014) were used to interconvert and manipulate files. Peaks were called with MACS2 (Feng, Liu, and Zhang 2011), and only high confidence peaks (found in at least two of the three replicates) were used in downstream analyses. Cis-regulatory Element Annotation System (CEAS) (v0.9.9.7) (Shin et al. 2009) was used to assign ChIP-seq peaks to genomic regions after making the mm10 annotation. Gene associations and GO terms for CHD4-bound elements were generated using GREAT (McLean et al. 2010). Gene tracks were visualized with UCSC Genome Browser (Kent et al. 2002). Permutations were done using in-house scripts based on the bedtools shuffle tool.

Transmission Electron Microscopy

Embryos were collected and fixed in 2% formaldehyde/2.5% glutaraldehyde in 0.15M sodium phosphate buffer, pH 7.4, overnight at 4°C. Following several washes in sodium phosphate buffer, samples were post-fixed for 1 hour in 1% buffered osmium tetroxide, dehydrated through a graded series of ethanol and embedded in PolyBed 812 epoxy resin (Polysciences). Using a diamond knife, 1µm cross-sections were cut, stained with 1% toluidine blue and examined by light microscopy to isolate the area of interest. Ultrasthin sections were cut with a diamond knife (70-80nm thick), mounted on 200 mesh copper grids and stained with 4% aqueous uranyl acetate for 15 minutes followed by Reynolds' lead citrate for 8 minutes (Reynolds 1963). The sections were

observed using a LEO EM-910 transmission electron microscrope (LEO Electron Microscopy Inc.), accelerating voltage of 80kV and micrographs were taken using a Gatan Orius SC1000 CCD Camera.

Embryonic Echocardiography

A pregnant dam carrying *Chd4*$^{flox/flox}$ and *Chd4*$^{\Delta flox/\Delta flox}$ embryos at E10.5 was anesthetized using 1.5% isoflurane and hair was removed from abdomen using topical hair removal agent. The female was placed under a warming lamp and ultrasound gel was applied to the abdomen. Embryos were imaged according to established protocols (Aristizabal et al. 1998; Srinivasan et al. 1998; Turnbull et al. 1995; Golden et al. 2012). Briefly, embryos were located using the VisualSonics Vevo 2100 Ultrasound machine (Visual Sonics, Inc.) using a 30MHz pediatric probe. B-mode images were used to guide pulsed-wave (PW) Doppler imaging in-line at the point of the atrioventricular canal. Post-imaging, embryos were injected with fluorescent colored microspheres (Cospheric, LLC) diluted in PBS for post-dissection identification and genotyping. Nine cycles of inflow/outflow peaks per recording were measured for peak velocity and velocity time integral (VTI) using VisualSonics image processing software from 5 embryos per genotype for total N=45 measurements per genotype and statistically compared by Student's t-test.

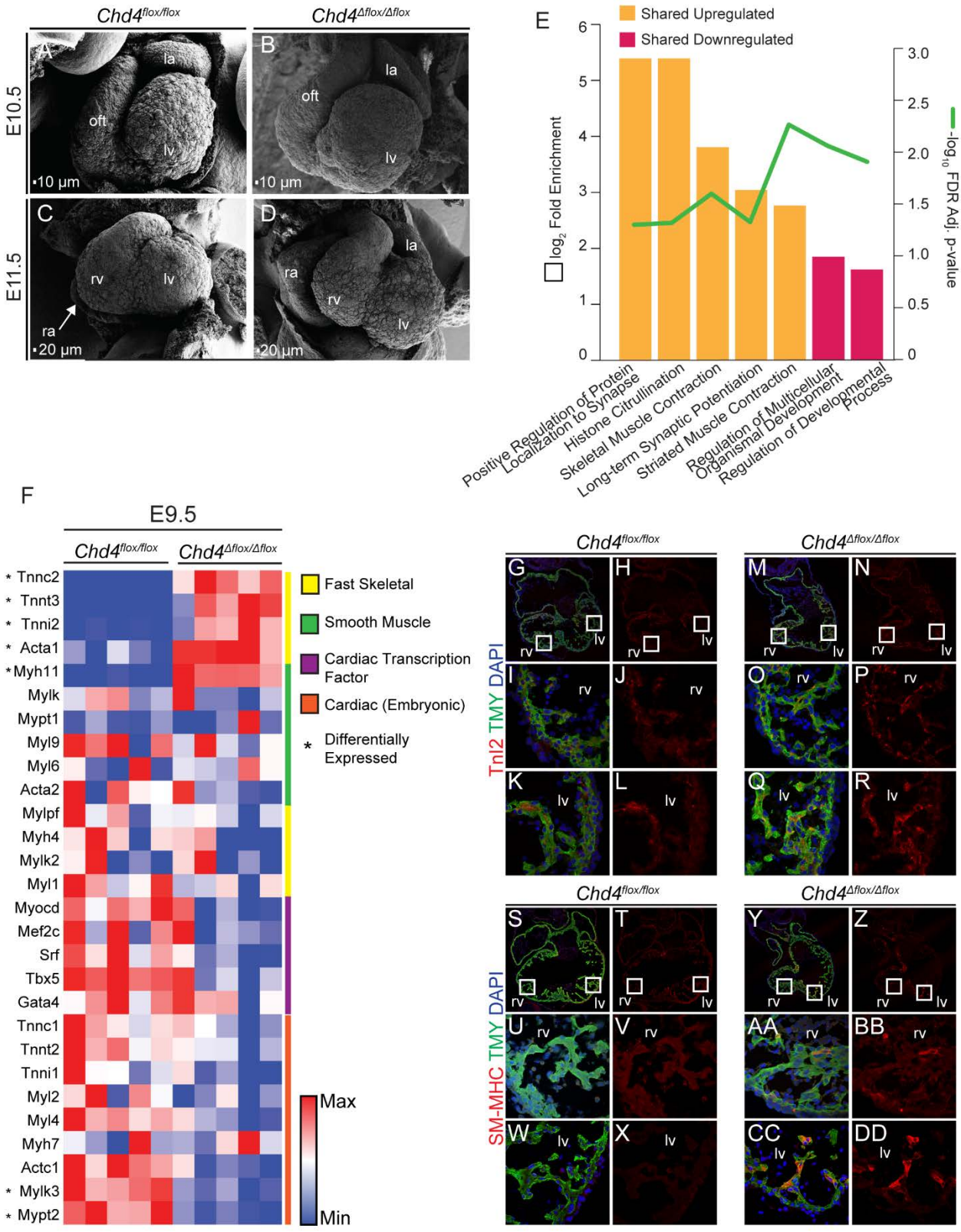

Chd4flox/flox
Chd4Δflox/Δflox
E10.5
E11.5
A
B
C
D
la
oft
lv
rv
ra
10 μm
20 μm
E
Shared Upregulated
Shared Downregulated
log2 Fold Enrichment
-log10 FDR Adj. p-value
Positive Regulation of Protein Localization to Synapse
Histone Citrullination
Skeletal Muscle Contraction
Long-term Synaptic Potentiation
Striated Muscle Contraction
Regulation of Multicellular Organismal Development
Regulation of Developmental Process
F
E9.5
Tnnc2
Tnnt3
Tnni2
Acta1
Myh11
Mylk
Mypt1
Myl9
Myl6
Acta2
Mylpf
Myh4
Mylk2
Myl1
Myocd
Mef2c
Srf
Tbx5
Gata4
Tnnc1
Tnnt2
Tnni1
Myl2
Myl4
Myh7
Actc1
Mylk3
Mypt2
Fast Skeletal
Smooth Muscle
Cardiac Transcription Factor
Cardiac (Embryonic)
Differentially Expressed
Max
Min
Tnl2 TMY DAPI
SM-MHC TMY DAPI
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
W
X
Y
Z
AA
BB
CC
DD

Figure 2.1. CHD4 is required for transcriptional repression of non-cardiac myofibril genes during cardiac development. (A-D) Scanning electron micrographs of *Chd4*$^{\Delta flox/\Delta flox}$ hearts at E10.5 compared to *Chd4*$^{flox/flox}$ controls demonstrate normal initiation of chamber formation (A,B). By E11.5 *Chd4*$^{\Delta flox/\Delta flox}$ hearts exhibit enlarged atria, a smaller right ventricle and an enlarged left ventricle compared to *Chd4*$^{flox/flox}$ hearts (C,D). (E) PANTHER gene ontology (GO) overrepresentation test in biological processes terms for genes up-regulated (orange columns) or down-regulated (pink columns) in *Chd4*$^{\Delta flox/\Delta flox}$ hearts at E9.5 and E10.5. Column height represents $\log_2$(Fold Enrichment) of genes associated with each GO term and green line represents –$\log_{10}$(FDR adjusted p-value) of each GO term. (F) Heatmap of fast skeletal, smooth muscle, and cardiac gene expression in *Chd4*$^{\Delta flox/\Delta flox}$ and *Chd4*$^{flox/flox}$ hearts at E9.5 row-scaled to show relative expression reveals lack of transcriptional repression of a set of fast skeletal and smooth muscle myofibril paralogs in the absence of CHD4. (G-R) Fast skeletal Troponin I2 (TnI2) misexpression in cardiomyocytes costained for tropomyosin (TMY) in E10.5 *Chd4*$^{\Delta flox/\Delta flox}$ hearts compared to *Chd4*$^{flox/flox}$ controls in the right ventricle (I-J compared to O-P) and left ventricle (K-L compared to Q-R). (S-DD) Smooth muscle myosin heavy chain (SM-MHC; *Myh11* gene) misexpression in cardiomyocytes costained for tropomyosin (TMY) in E10.5 *Chd4*$^{\Delta flox/\Delta flox}$ hearts compared to *Chd4*$^{flox/flox}$ controls in the right ventricle (U-V compared to AA-BB) and left ventricle (W-X compared to CC-DD). Boxed regions denote area of higher magnification. la, left atria; lv, left ventricle; oft, outflow tract; ra, right atria; rv, right ventricle.

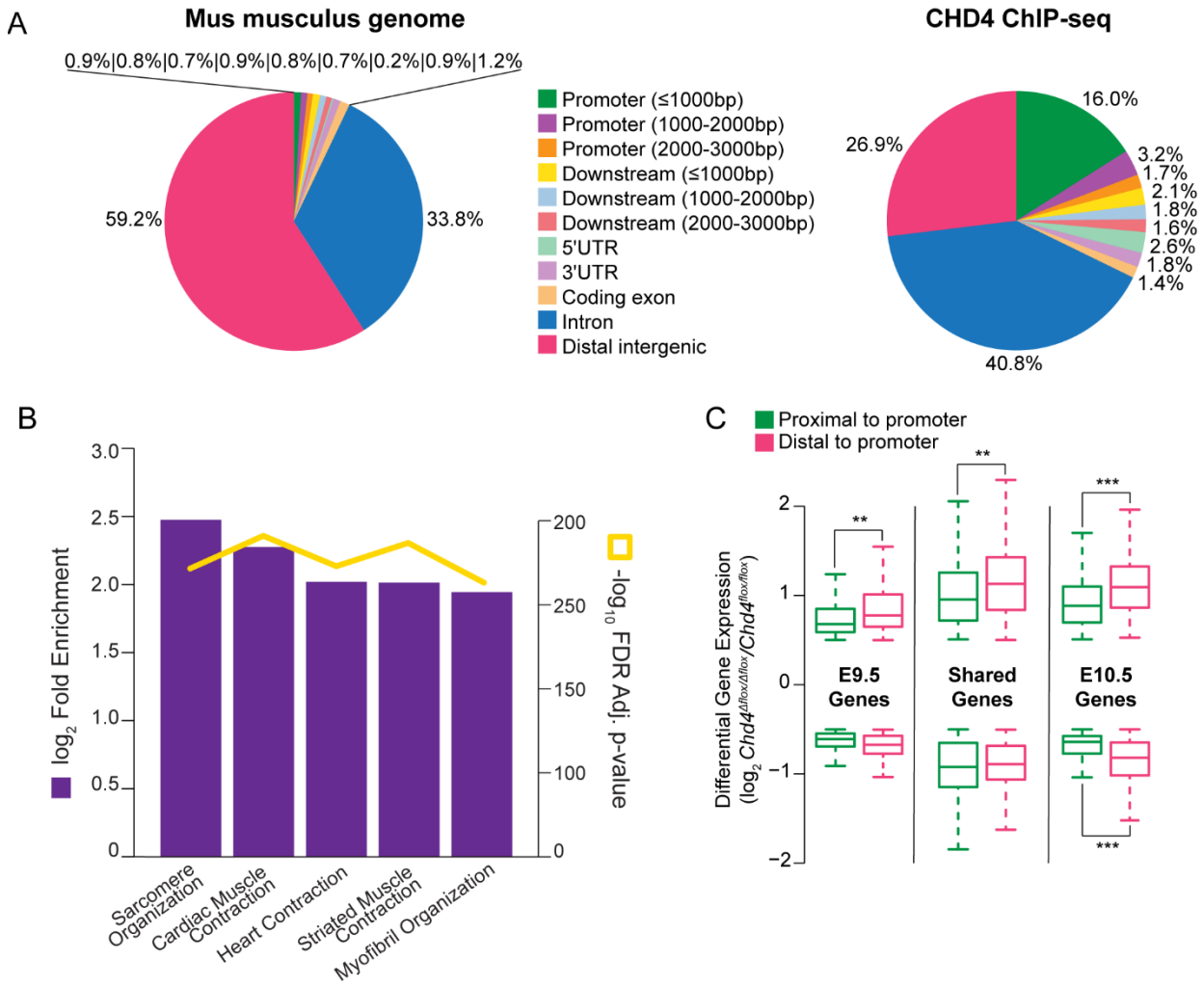


Figure 2.2. CHD4 regulates sarcomere assembly through direct binding to gene regulatory regions in the developing heart. (A) *Left*, composition of *Mus musculus* genome by distance to nearest gene transcription start site (TSS) based on mm10 genome build. *Right*, composition of genomic regions bound by CHD4 in wild-type E10.0 hearts by ChIP-seq demonstrates high enrichment at promoter and intronic regions. (B) GO Biological Processes terms for regions bound by CHD4 ranked by $\log_2$ Fold Enrichment) (purple column) and $-\log_{10}$ (FDR adjusted p-value) (yellow line) indicates CHD4 binds genes required for sarcomere organization and myofibril assembly. (C) Comparing the magnitude of gene expression change in *Chd4*$^{\Delta flox/\Delta flox}$

hearts between genes containing proximal promoter (1500bp upstream≥TSS≤500bp downstream) or distal intergenic CHD4 ChIP-peaks demonstrates a higher degree of change in genes associated with distal regulatory peaks in E9.5, E10.5, and shared upregulated genes and E10.5 downregulated genes. ***=p-value ≤0.001, **=p-value ≤0.01.

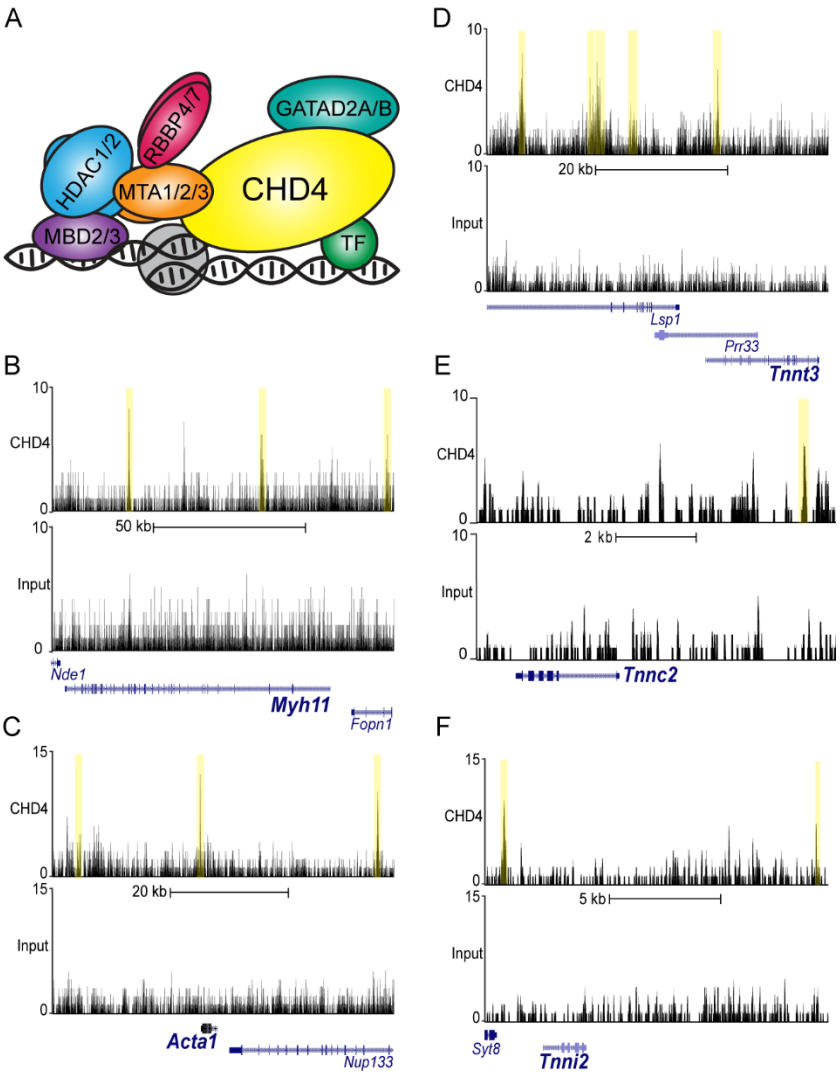


Figure 2.3. CHD4 binds genomic regions linked to *Myh11, Acta1, Tnnc2, Tnnt3 and Tnni2.* (A) Diagram of CHD4 ChIP-seq approach representing methodology. TF, transcription factor. (B-F) CHD4 binding sites identified by ChIP-seq reads enriched over input DNA at non-cardiac myofibril paralog genes are highlighted in yellow.

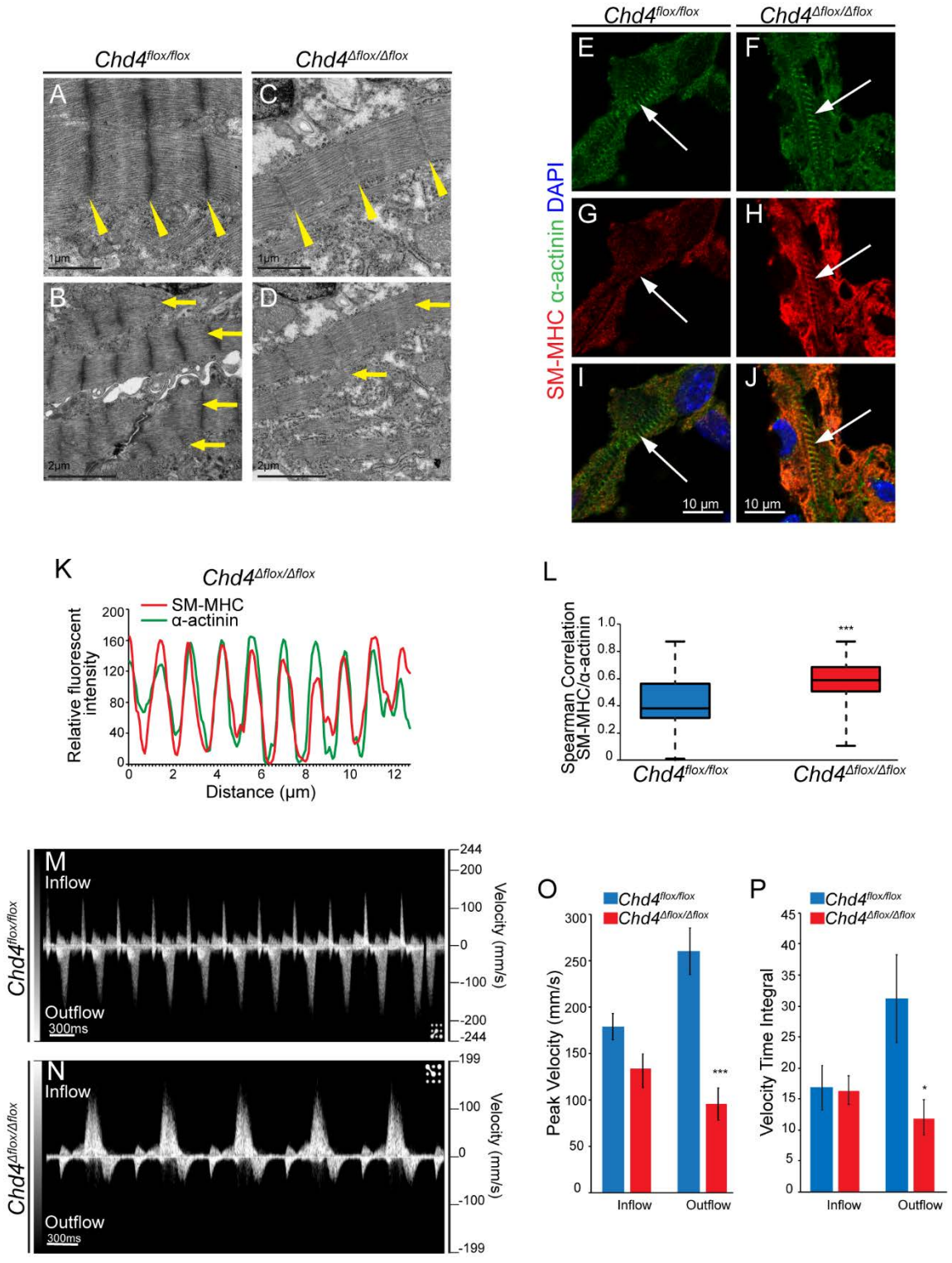


Figure 2.4. Misexpression of non-cardiac myofibril paralogs in the absence of CHD4 leads to sarcomere malformation and altered cardiac function during

development. (A-D) Transmission electron microscopy reveals weak, deficient Z-disk formation (yellow wedges) and decreased sarcomere formation and alignment (yellow arrows) in *Chd4*$^{\Delta flox/\Delta flox}$ hearts. (E-J) Co-staining for α-actinin (E-F) and smooth muscle myosin heavy chain (SM-MHC) (G-H) demonstrates organization of SM-MHC into striated sarcomere structures and integration into the cardiac sarcomere in *Chd4*$^{\Delta flox/\Delta flox}$ cardiomyocytes compared to *Chd4*$^{flox/flox}$ controls (I,J). (K) Relative fluorescent signal of SM-MHC and α-actinin plotted against distance in E10.5 *Chd4*$^{\Delta flox/\Delta flox}$ cardiomyocytes reveals intercalation of SM-MHC into the nascent sarcomere. (L) Spearman correlation between SM-MHC/α-actinin signal in *Chd4*$^{\Delta flox/\Delta flox}$ cardiomyocytes is significant compared to *Chd4*$^{flox/flox}$ controls by Student's t-test, N=27 vectors per genotype. (M-N) Non-invasive *in utero* embryonic echocardiography by pulsed-wave (PW) Doppler on E10.5 embryos show *Chd4*$^{\Delta flox/\Delta flox}$ embryos have pronounced decreases in ventricular (outflow) velocity (N) compared to *Chd4*$^{flox/flox}$ controls (M). (O-P) Quantification of blood flow velocity from PW Doppler shows significant decrease in ventricular cardiac function in *Chd4*$^{\Delta flox/\Delta flox}$ hearts by peak velocity (O) and velocity time integral (VTI) (P) by Student's t-test, N=5 embryos per genotype, SEM ± 14.35, 18.12, 25.12 and 16.93 (O) and 3.57, 2.30, 7.11 and 2.85 (P). *=p-value<0.05, ***=p-value<0.001.

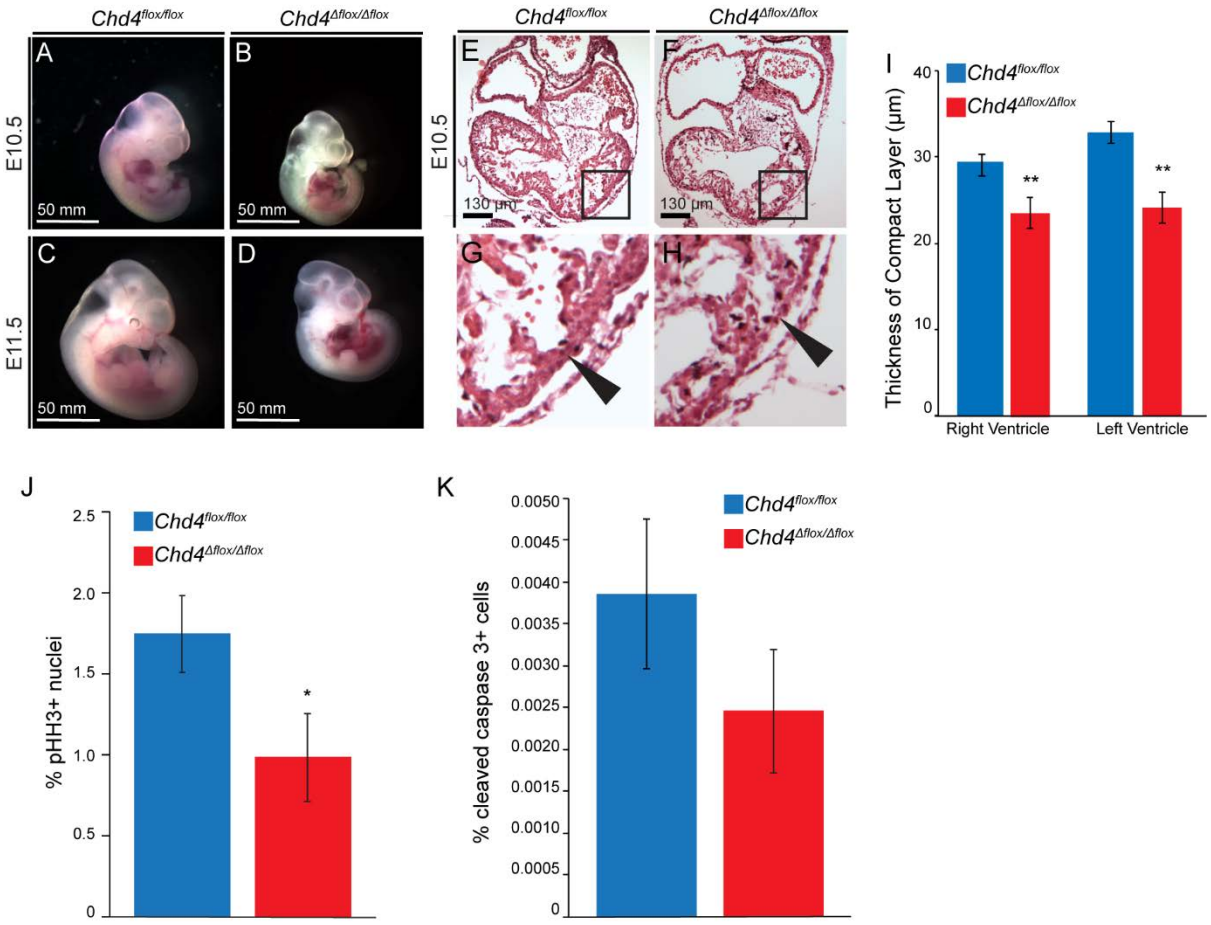


Figure S2.1. CHD4 is required for myocardial growth before E10.5. (A-D) Gross morphology of *Chd4*$^{\Delta flox/\Delta flox}$ and *Chd4*$^{flox/flox}$ embryos at E10.5 (A,B) and E11.5 (C,D) show growth arrest at E10.5 in the absence of CHD4 in the developing heart with concomitant pericardial edema and hemorrhaging. (E-H) Histological analysis at E10.5 reveals decreased trabecular complexity in the ventriclular myocardium in *Chd4*$^{\Delta flox/\Delta flox}$ hearts (E, F) with a thinner myocardial compact layer compared to *Chd4*$^{flox/flox}$ controls (wedge arrow) (G, H). Boxed area denotes region of higher magnification. (I) Quantification demonstrates significant decrease in thickness of the right and left ventricle compact layer in *Chd4*$^{\Delta flox/\Delta flox}$ compared to *Chd4*$^{flox/flox}$ controls by Student's t-test, N=27 vectors per genotype, error bars represent SEM ± 1.31, 1.27, 1.83, 1.77. (J)

Quantification of phosphorylated histone H3+/tropomyosin+ cells indicates a significant decrease in the mitotic index in *Chd4*$^{\Delta flox/\Delta flox}$ hearts by Student's t-test, N=9 sections per genotype, SEM ± 0.238 and 0.271. (J) Quantification of cleaved caspase3+/tropomyosin+ cells indicates no significant change in apoptosis in *Chd4*$^{\Delta flox/\Delta flox}$ hearts by Student's t-test, N=9 sections per genotype, SEM ± 0.0009 and 0.0007. *=p-value<0.05,**=p-value<0.01. oft, outflow tract; lv, left ventricle; la, left atrium; ra, right atrium; rv, right ventricle.

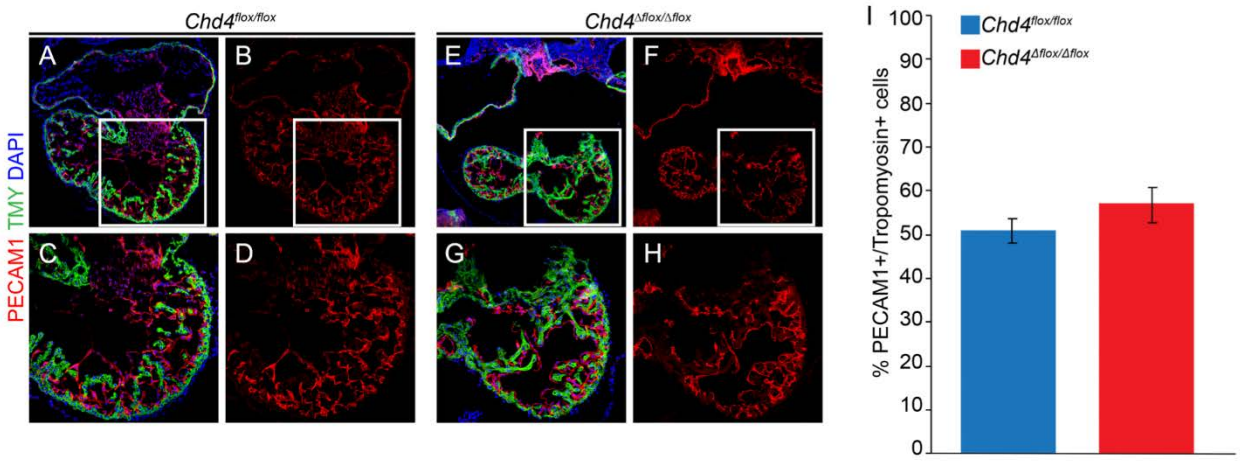


Figure S2.2. CHD4 ablation does not increase non-cardiomyocyte cell types in the developing heart. (A-H) Platelet endothelial cell adhesion molecule 1 (PECAM1) and tropomyosin (TMY) co-staining followed by quantification of PECAM1+ cells and tropomyosin+ cells (I) demonstrate no significant increase in endothelial cells relative to cardiomyocytes in *Chd4*$^{\Delta flox/\Delta flox}$ embryos compared to control *Chd4*$^{flox/flox}$ embryos at E10.5 by Student's t-test, N=29 regions per genotype, error bars represent SEM ± 0.027 and 0.040. Boxed region denotes area of higher magnification.

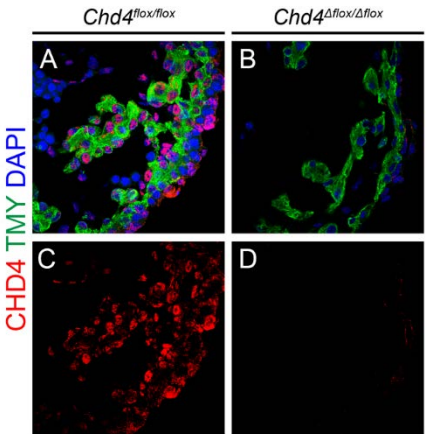


Figure S2.3. CHD4 is depleted from the myocardium at E9.5. (A-D) CHD4 is depleted from cardiomyocytes costained with tropomyosin (TMY) at E9.5 in *Chd4*$^{\Delta flox/\Delta flox}$ embryos (B,D) compared to *Chd4*$^{flox/flox}$ controls (A,C).

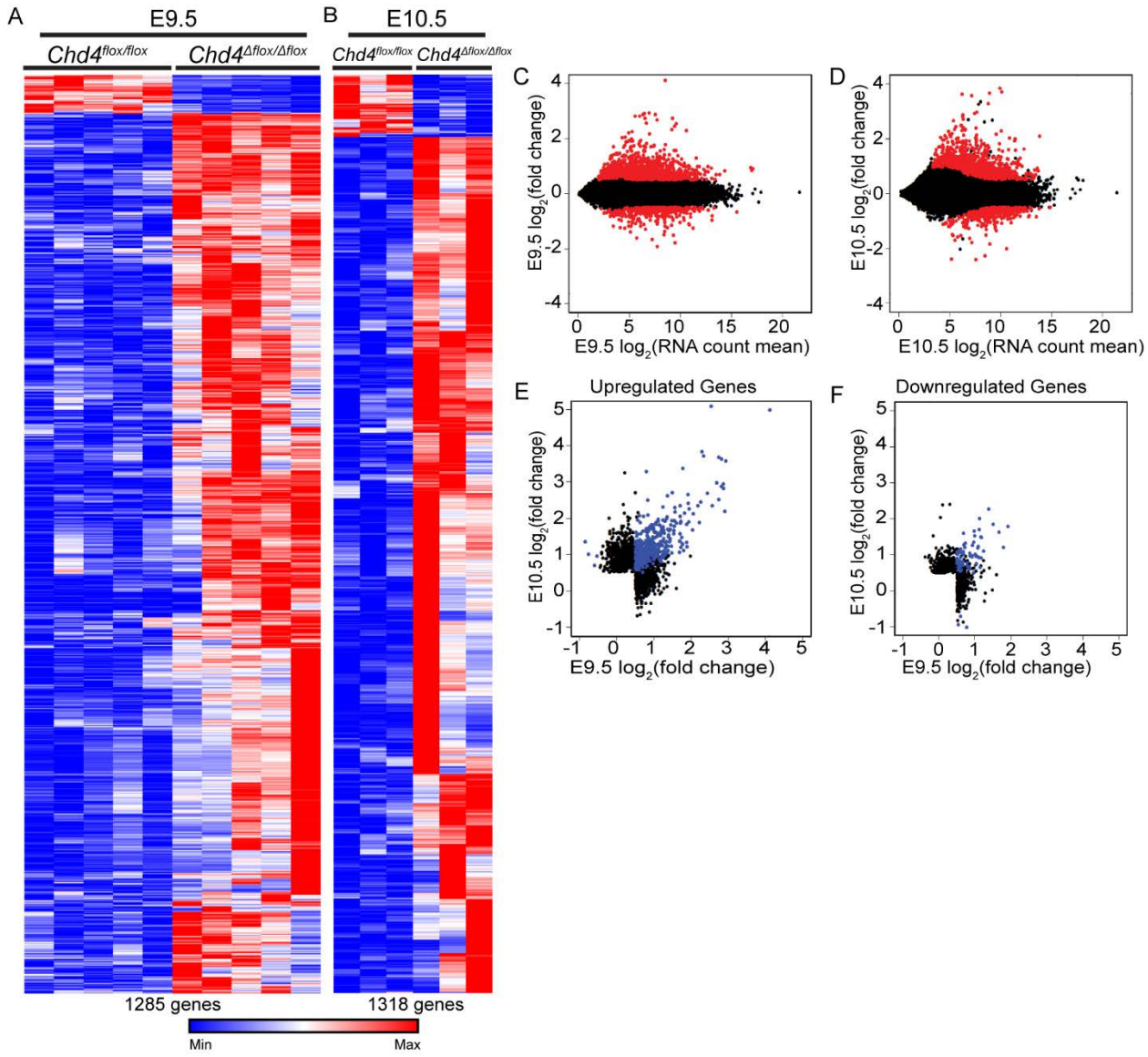


Figure S2.4. CHD4 temporally restricts gene expression in the developing heart.

(A-B) Ward's two-way hierarchical clustering of differential gene expression profiles of E9.5 (A) and E10.5 (B) *Chd4*$^{\Delta flox/\Delta flox}$ and *Chd4*$^{flox/flox}$ hearts show a distinct molecular signature in the absence of CHD4 at both time points. Gene expression is row-scaled to show relative expression. (C-D) MA plots of $\log_2$(fold change) versus $\log_2$(RNA count mean) show a greater number of genes upregulated (913 genes) than downregulated (372 genes) in the absence of CHD4 at E9.5 (C) and a greater number of genes

upregulated (920 genes) than downregulated (398 genes) in the absence of CHD4 at E10.5 (D). Red points indicate adjusted p-value of less than 0.05. (E-F) Plotting E10.5 log_2(fold change) versus E9.5 log_2(fold change), with shared upregulated (E) or downregulated (F) genes colored in blue demonstrate distinct requirements in transcriptional regulation dependent on developmental stage. Black points represent genes misexpressed unique to that stage.

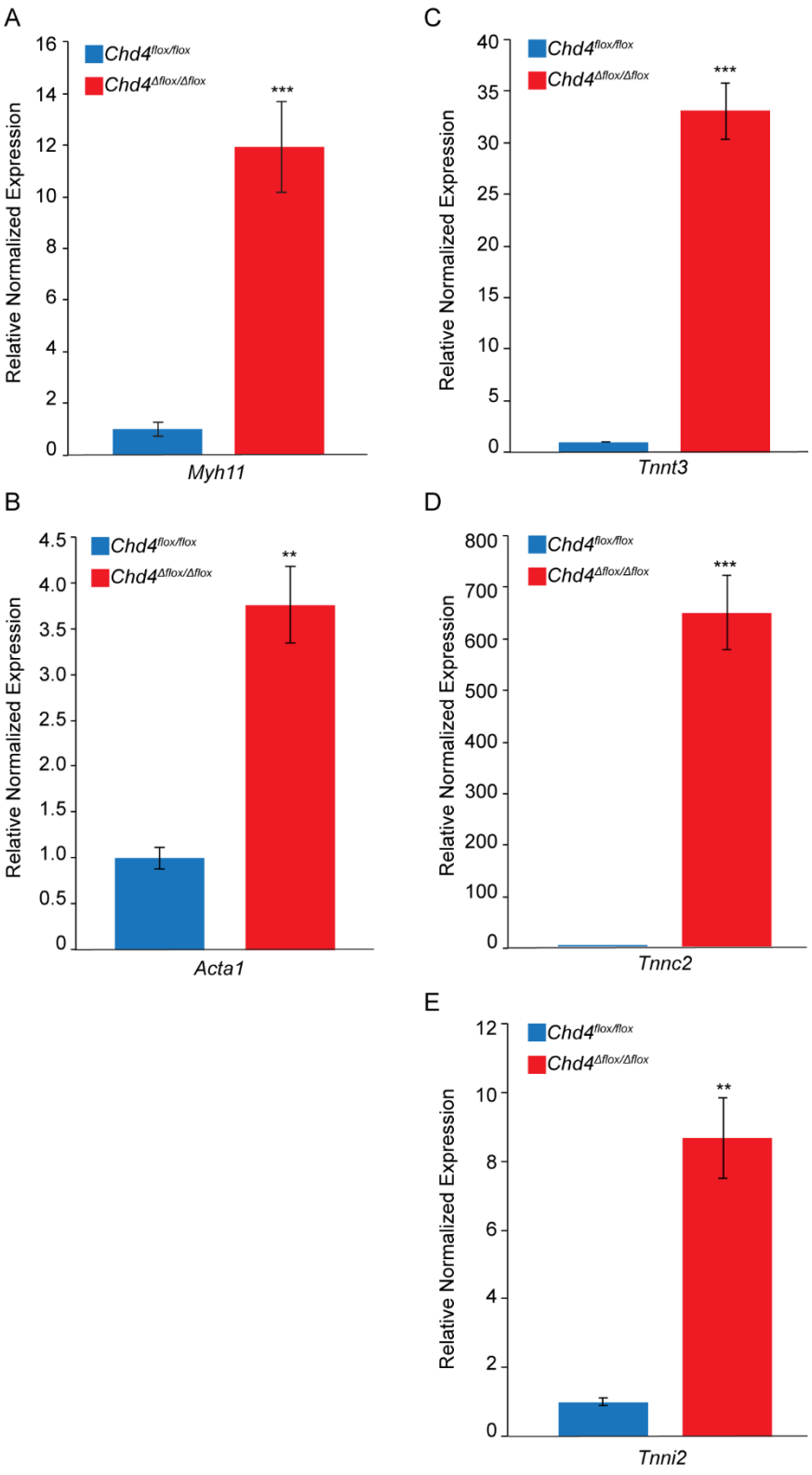


Figure S2.5. CHD4 is required to repress expression of non-cardiac myofibril isoforms in the embryonic heart. (A-E) Reverse transcription of RNA followed by

quantitative real-time PCR (qPCR) demonstrates significantly increased expression of *Myh11* (A), *Acta1* (B), *Tnnt3* (C), *Tnnc2* (D) and *Tnni2* (E) in E10.5 *Chd4*$^{\Delta flox/\Delta flox}$ hearts over *Chd4*$^{flox/flox}$ controls by Student's t-test, ***=p-value<0.001, **=p-value<0.01, error bars represent SEM ± 0.264 and 1.763 (A), 0.122 and 0.418 (B), 0.040 and 2.731 (C), 0.319 and 71.779 (D), 0.101 and 1.172 (E). Data is represented relative to *Rps29* levels and normalized to *Chd4*$^{flox/flox}$ controls.

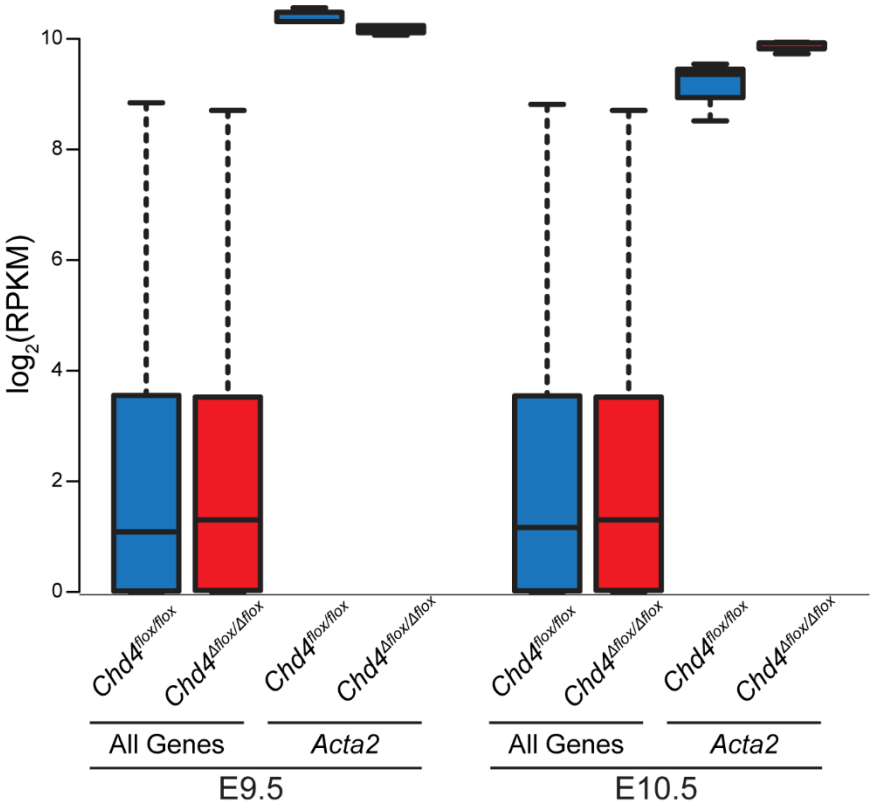


Figure S2.6. Acta2 is not differentially expressed in the absence of CHD4. Plotting log_2(RPKM) values from RNA-seq of *Chd4*$^{flox/flox}$ and *Chd4*$^{\Delta flox/\Delta flox}$ hearts at E9.5 and E10.5 demonstrate that *Acta2* (alpha-smooth muscle actin) is highly expressed compared to median overall gene expression but is not significantly up- or down-regulated in the absence of CHD4.

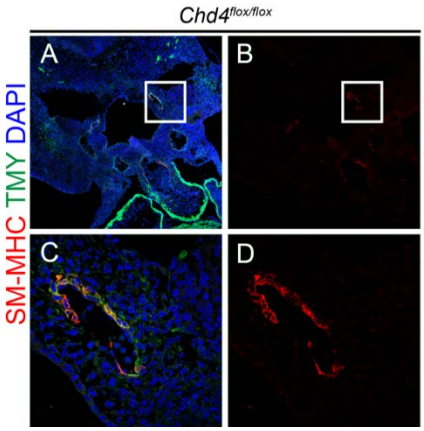


Figure S2.7. Smooth muscle myosin heavy chain antibody specifically stains smooth muscle cells. (A-D) Smooth muscle myosin heavy chain staining of smooth muscle cells surrounding the pharyngeal arch arteries in an E10.5 *Chd4*$^{flox/flox}$ control embryo. Boxed region denotes area of higher magnification.

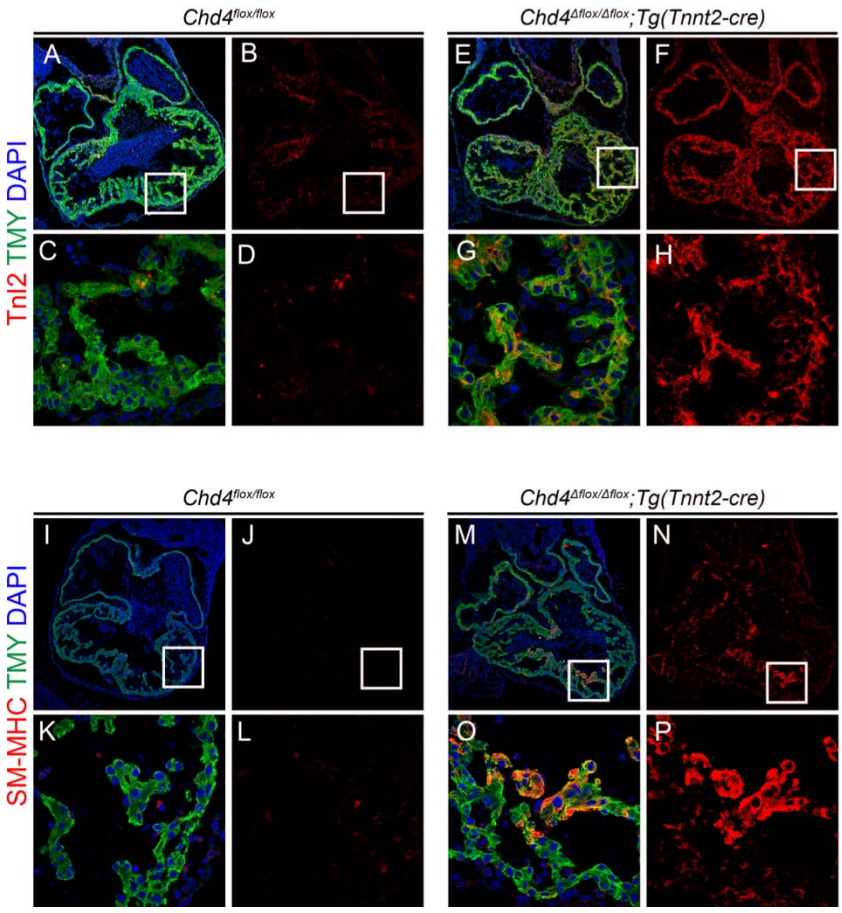


Figure S2.8. Ablation of CHD4 in the myocardium results in misexpression of non-cardiac myofibril paralogs. Immunohistochemistry for fast skeletal TnI2 (A-H) and smooth muscle myosin heavy chain (I-P) reveals increased expression of fast skeletal and smooth muscle myofibril subunits in the myocardium of *Chd4$^{\Delta flox/\Delta flox}$;Tg(Tnnt2-cre)* hearts (E-H, M-P) at E10.5 compared to controls (A-D, I-L). White boxes denote regions of higher magnification in adjacent panels.

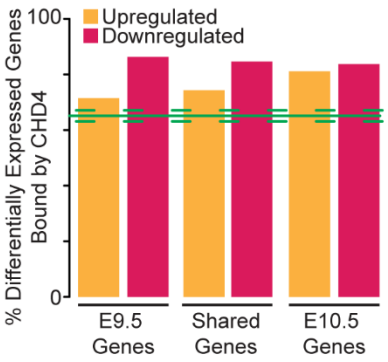


Figure S2.9. CHD4 directly binds genomic elements of transcriptional targets.

Genes differentially expressed in *Chd4*$^{\Delta flox/\Delta flox}$ hearts at E9.5, E10.5 or commonly misregulated (shared) between both time points were cross-referenced to genes containing CHD4 ChIP-seq peaks. Columns represent the significant percentage of genes directly bound by CHD4 and either up- or down-regulated in its absence, p-value<0.001. Green solid line indicates the percent overlap expected by chance (1000 permutations) with dashed lines indicating standard deviation.

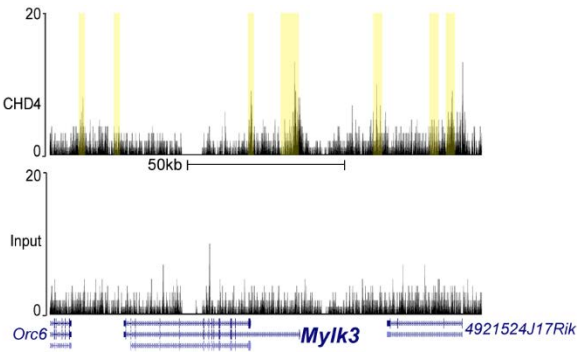


Figure S2.10. CHD4 binds genomic regions linked to cardiac *Mylk3*. CHD4 binding sites identified by ChIP-seq reads enriched over input DNA at cardiac myosin light chain kinase 3 (*Mylk3)* are highlighted in yellow.

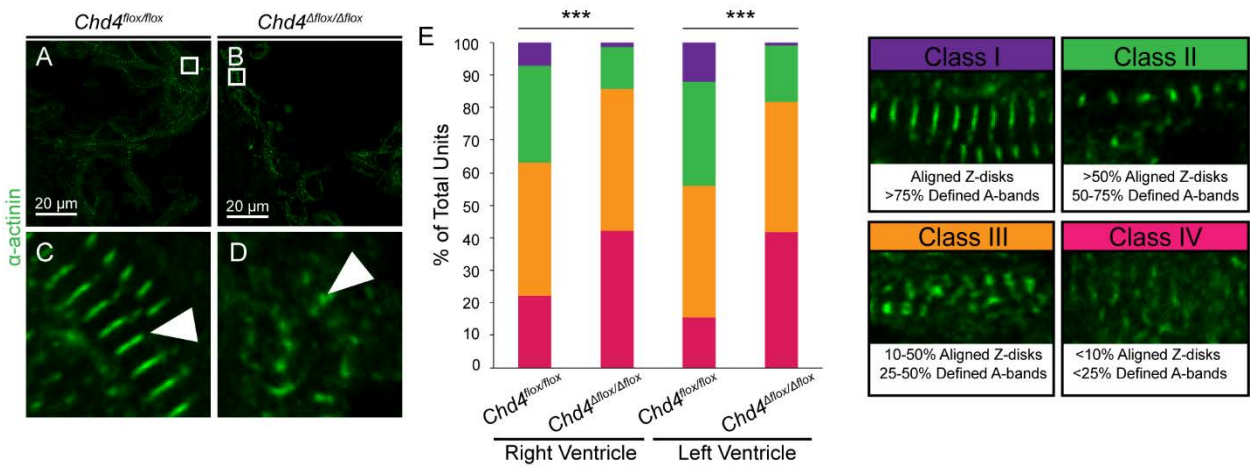


Figure S2.11. Misexpression of non-cardiac myofibril paralogs contribute to sarcomere disarray in the absence of CHD4. (A-D) α-actinin staining reveals disorganized sarcomere structure (wedge arrow) in *Chd4*$^{\Delta flox/\Delta flox}$ cardiomyocytes at E10.5 compared to *Chd4*$^{flox/flox}$ controls. Box denotes region of higher magnification. (E) Measurement of sarcomere organization by classification of Z-disk and A-band formation reveals significant decreases in sarcomere organization in *Chd4*$^{\Delta flox/\Delta flox}$ cardiomyocytes by χ^2 test, N= 225 vectors per ventricle per genotype, ***=p-value<0.001.

REFERENCES

Aguayo-Gomez, A., J. Arteaga-Vazquez, Y. Svyryd, J. Calderon-Colmenero, C. Zamora-Gonzalez, G. Vargas-Alarcon, and O. M. Mutchinick. 2015. 'Identification of Copy Number Variations in Isolated Tetralogy of Fallot', *Pediatr Cardiol*, 36: 1642-6.

Allen, H. F., P. A. Wade, and T. G. Kutateladze. 2013. 'The NuRD architecture', *Cell Mol Life Sci*, 70: 3513-24.

Anders, S., P. T. Pyl, and W. Huber. 2015. 'HTSeq--a Python framework to work with high-throughput sequencing data', *Bioinformatics*, 31: 166-9.

Aristizabal, O., D. A. Christopher, F. S. Foster, and D. H. Turnbull. 1998. '40-MHZ echocardiography scanner for cardiovascular assessment of mouse embryos', *Ultrasound Med Biol*, 24: 1407-17.

Basson, C. T., D. R. Bachinsky, R. C. Lin, T. Levi, J. A. Elkins, J. Soults, D. Grayzel, E. Kroumpouzou, T. A. Traill, J. Leblanc-Straceski, B. Renault, R. Kucherlapati, J. G. Seidman, and C. E. Seidman. 1997. 'Mutations in human TBX5 [corrected] cause limb and cardiac malformation in Holt-Oram syndrome', *Nat Genet*, 15: 30-5.

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

Bruneau, B. G., M. Logan, N. Davis, T. Levi, C. J. Tabin, J. G. Seidman, and C. E. Seidman. 1999. 'Chamber-specific cardiac expression of Tbx5 and heart defects in Holt-Oram syndrome', *Dev Biol*, 211: 100-8.

Bruneau, B.G., G. Nemer, J.P. Schmitt, F. Charron, L. Robitaille, S. Caron, D. A. Conner, M. Gessler, M. Nemer, C. E. Seidman, and J.G. Seidman. 2001. 'A Murine Model of Holt-Oram Syndrome Defines Roles of the T-Box Transcription Factor Tbx5 in Cardiogenesis and Disease', *Cell*, 106: 709-21.

Clement, S., M. Stouffs, E. Bettiol, S. Kampf, K. H. Krause, C. Chaponnier, and M. Jaconi. 2007. 'Expression and function of alpha-smooth muscle actin during embryonic-stem-cell-derived cardiomyocyte differentiation', *J Cell Sci*, 120: 229-38.

Conway, S. J., A. Kruzynska-Frejtag, P. L. Kneer, M. Machnicki, and S. V. Koushik. 2003. 'What cardiovascular defect does my prenatal mouse mutant have, and why?', *Genesis*, 35: 1-21.

Craig, Roger, and John L. Woodhead. 2006. 'Structure and function of myosin filaments', *Current Opinion in Structural Biology*, 16: 204-12.

Dolk, H., M. Loane, and E. Garne. 2010. 'The prevalence of congenital anomalies in Europe', *Adv Exp Med Biol*, 686: 349-64.

Dorr, Kerry M., Nirav M. Amin, Lauren M. Kuchenbrod, Hanna Labiner, Marta S. Charpentier, Larysa H. Pevny, Andy Wessels, and Frank L. Conlon. 2015. 'Casz1 is required for cardiomyocyte G1-to-S phase progression during mammalian cardiac development', *Development*.

England, J., and S. Loughna. 2013. 'Heavy and light roles: myosin in the morphogenesis of the heart', *Cell Mol Life Sci*, 70: 1221-39.

Feng, J., T. Liu, and Y. Zhang. 2011. 'Using MACS to identify peaks from ChIP-Seq data', *Curr Protoc Bioinformatics*, Chapter 2: Unit 2.14.

Garnatz, A. S., Z. Gao, M. Broman, S. Martens, J. U. Earley, and E. C. Svensson. 2014. 'FOG-2 mediated recruitment of the NuRD complex regulates cardiomyocyte proliferation during heart development', *Dev Biol*, 395: 50-61.

Golden, H. B., S. Sunder, Y. Liu, X. Peng, and D. E. Dostal. 2012. 'In utero assessment of cardiovascular function in the embryonic mouse heart using high-resolution ultrasound biomicroscopy', *Methods Mol Biol*, 843: 245-63.

Gómez-del Arco, Pablo, Eusebio Perdiguero, Paula Sofia Yunes-Leites, Rebeca Acín-Pérez, Miriam Zeini, Antonio Garcia-Gomez, Krishnamoorthy Sreenivasan, Miguel Jiménez-Alcázar, Jessica Segalés, Dolores López-Maderuelo, Beatriz Ornés, Luis Jesús Jiménez-Borreguero, Gaetano D'Amato, David Enshell-Seijffers, Bruce Morgan, Katia Georgopoulos, Abul B M. M. K. Islam, Thomas Braun, José Luis de la Pompa, Johnny Kim, José A Enriquez, Esteban Ballestar, Pura Muñoz-Cánoves, and Juan Miguel Redondo. 'The Chromatin Remodeling Complex Chd4/NuRD Controls Striated Muscle Identity and Metabolic Homeostasis', *Cell Metabolism*, 23: 881-92.

Heidersbach, Amy, Chris Saxby, Karen Carver-Moore, Yu Huang, Yen-Sin Ang, Pieter J. de Jong, Kathryn N. Ivey, and Deepak Srivastava. 2013. 'microRNA-1 regulates sarcomere formation and suppresses smooth muscle gene expression in the mammalian heart', *eLife*, 2: e01323.

Heron, M., D. L. Hoyert, S. L. Murphy, J. Xu, K. D. Kochanek, and B. Tejada-Vera. 2009. 'Deaths: final data for 2006', *Natl Vital Stat Rep*, 57: 1-134.

Hirschy, A., F. Schatzmann, E. Ehler, and J. C. Perriard. 2006. 'Establishment of cardiac cytoarchitecture in the developing mouse heart', *Dev Biol*, 289: 430-41.

Homsy, J., S. Zaidi, Y. Shen, J. S. Ware, K. E. Samocha, K. J. Karczewski, S. R. DePalma, D. McKean, H. Wakimoto, J. Gorham, S. C. Jin, J. Deanfield, A. Giardini, G. A. Porter, Jr., R. Kim, K. Bilguvar, F. Lopez-Giraldez, I. Tikhonova, S. Mane, A. Romano-Adesman, H. Qi, B. Vardarajan, L. Ma, M. Daly, A. E. Roberts, M. W. Russell, S. Mital, J. W. Newburger, J. W. Gaynor, R. E. Breitbart, I. Iossifov, M. Ronemus, S. J. Sanders, J. R. Kaltman, J. G. Seidman, M. Brueckner, B. D. Gelb, E. Goldmuntz, R. P.

Lifton, C. E. Seidman, and W. K. Chung. 2015. 'De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies', *Science*, 350: 1262-6.

Hung, H., R. Kohnken, and J. Svaren. 2012. 'The nucleosome remodeling and deacetylase chromatin remodeling (NuRD) complex is required for peripheral nerve myelination', *J Neurosci*, 32: 1517-27.

Ilkovski, B., S. Clement, C. Sewry, K. N. North, and S. T. Cooper. 2005. 'Defining alpha-skeletal and alpha-cardiac actin expression in human heart and skeletal muscle explains the absence of cardiac involvement in ACTA1 nemaline myopathy', *Neuromuscul Disord*, 15: 829-35.

Kaltenbrun, E., T. M. Greco, C. E. Slagle, L. M. Kennedy, T. Li, I. M. Cristea, and F. L. Conlon. 2013. 'A Gro/TLE-NuRD corepressor complex facilitates Tbx20-dependent transcriptional repression', *J Proteome Res*, 12: 5395-409.

Kashiwagi, M., B. A. Morgan, and K. Georgopoulos. 2007. 'The chromatin remodeler Mi-2beta is required for establishment of the basal epidermis and normal differentiation of its progeny', *Development*, 134: 1571-82.

Kent, W. James, Charles W. Sugnet, Terrence S. Furey, Krishna M. Roskin, Tom H. Pringle, Alan M. Zahler, Haussler, and David. 2002. 'The Human Genome Browser at UCSC', *Genome Res*, 12: 996-1006.

Kim, J., S. Sif, B. Jones, A. Jackson, J. Koipally, E. Heller, S. Winandy, A. Viel, A. Sawyer, T. Ikeda, R. Kingston, and K. Georgopoulos. 1999. 'Ikaros DNA-binding proteins direct formation of chromatin remodeling complexes in lymphocytes', *Immunity*, 10: 345-55.

Kirk, E. P., M. Sunde, M. W. Costa, S. A. Rankin, O. Wolstein, M. L. Castro, T. L. Butler, C. Hyun, G. Guo, R. Otway, J. P. Mackay, L. B. Waddell, A. D. Cole, C. Hayward, A. Keogh, P. Macdonald, L. Griffiths, D. Fatkin, G. F. Sholler, A. M. Zorn, M. P. Feneley, D. S. Winlaw, and R. P. Harvey. 2007. 'Mutations in cardiac T-box factor gene TBX20 are associated with diverse cardiac pathologies, including defects of septation and valvulogenesis and cardiomyopathy', *Am J Hum Genet*, 81: 280-91.

Langmead, B., C. Trapnell, M. Pop, and S. L. Salzberg. 2009. 'Ultrafast and memory-efficient alignment of short DNA sequences to the human genome', *Genome Biol*, 10: R25.

Lassmann, T., Y. Hayashizaki, and C. O. Daub. 2009. 'TagDust--a program to eliminate artifacts from next generation sequencing data', *Bioinformatics*, 25: 2839-40.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 2009. 'The Sequence Alignment/Map format and SAMtools', *Bioinformatics*, 25: 2078-9.

Li, Q. Y., R. A. Newbury-Ecob, J. A. Terrett, D. I. Wilson, A. R. Curtis, C. H. Yi, T. Gebuhr, P. J. Bullen, S. C. Robson, T. Strachan, D. Bonnet, S. Lyonnet, I. D. Young, J. A. Raeburn, A. J. Buckler, D. J. Law, and J. D. Brook. 1997. 'Holt-Oram syndrome is caused by mutations in TBX5, a member of the Brachyury (T) gene family', *Nat Genet*, 15: 21-9.

Love, M. I., W. Huber, and S. Anders. 2014. 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biol*, 15: 550.

Luo, M., T. Ling, W. Xie, H. Sun, Y. Zhou, Q. Zhu, M. Shen, L. Zong, G. Lyu, Y. Zhao, T. Ye, J. Gu, W. Tao, Z. Lu, and I. Grummt. 2013. 'NuRD blocks reprogramming of mouse somatic cells into pluripotent stem cells', *Stem Cells*, 31: 1278-86.

Marhold, J., K. Kramer, E. Kremmer, and F. Lyko. 2004. 'The Drosophila MBD2/3 protein mediates interactions between the MI-2 chromatin complex and CpT/A-methylated DNA', *Development*, 131: 6033-9.

McLean, C. Y., D. Bristor, M. Hiller, S. L. Clarke, B. T. Schaar, C. B. Lowe, A. M. Wenger, and G. Bejerano. 2010. 'GREAT improves functional interpretation of cis-regulatory regions', *Nat Biotechnol*, 28: 495-501.

Mi, H, N Guo, A Kejariwal, and PD Thomas. 2007. 'PANTHER version 6: protein sequence and function evolution data with expanded representation of biological pathways', *Nucleic Acids Res*, 35: D247 - D52.

Mi, Huaiyu, Xiaosong Huang, Anushya Muruganujan, Haiming Tang, Caitlin Mills, Diane Kang, and Paul D. Thomas. 2017. 'PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements', *Nucleic Acids Res*, 45: D183-D89.

Miccio, Annarita, Yuhuan Wang, Wei Hong, Gregory D Gregory, Hongxin Wang, Xiang Yu, John K Choi, Suresh Shelat, Wei Tong, Mortimer Poncz, and Gerd A Blobel. 2010. 'NuRD mediates activating and repressive functions of GATA-1 and FOG-1 during blood development', *Embo j*, 29: 442-56.

Montgomery, Rusty L., Christopher A. Davis, Matthew J. Potthoff, Michael Haberland, Jens Fielitz, Xiaoxia Qi, Joseph A. Hill, James A. Richardson, and Eric N. Olson. 2007. 'Histone deacetylases 1 and 2 redundantly regulate cardiac morphogenesis, growth, and contractility', *Genes Dev*, 21: 1790-802.

Mori, A. D., and B. G. Bruneau. 2004. 'TBX5 mutations and congenital heart disease: Holt-Oram syndrome revealed', *Curr Opin Cardiol*, 19: 211-5.

Moses, K. A., F. DeMayo, R. M. Braun, J. L. Reecy, and R. J. Schwartz. 2001. 'Embryonic expression of an Nkx2-5/Cre gene using ROSA26 reporter mice', *Genesis*, 31: 176-80.

Nishii, K., S. Morimoto, R. Minakami, Y. Miyano, K. Hashizume, M. Ohta, D. Y. Zhan, Q. W. Lu, and Y. Shibata. 2008. 'Targeted disruption of the cardiac troponin T gene causes sarcomere disassembly and defects in heartbeat within the early mouse embryo', *Dev Biol*, 322: 65-73.

O'Shaughnessy-Kirwan, A., J. Signolet, I. Costello, S. Gharbi, and B. Hendrich. 2015. 'Constraint of gene expression by the chromatin remodelling protein CHD4 facilitates lineage specification', *Development*, 142: 2586-97.

Postma, Alex V., Connie R. Bezzina, and Vincent M. Christoffels. 2016. 'Genetics of congenital heart disease: the contribution of the noncoding regulatory genome', *J Hum Genet*, 61: 13-19.

Quinlan, A. R. 2014. 'BEDTools: The Swiss-Army Tool for Genome Feature Analysis', *Curr Protoc Bioinformatics*, 47: 11.12.1-34.

Reynolds, E. S. 1963. 'The use of lead citrate at high pH as an electron-opaque stain in electron microscopy', *J Cell Biol*, 17: 208-12.

Roche, A. E., B. J. Bassett, S. A. Samant, W. Hong, G. A. Blobel, and E. C. Svensson. 2008. 'The zinc finger and C-terminal domains of MTA proteins are required for FOG-2-mediated transcriptional repression via the NuRD complex', *J Mol Cell Cardiol*, 44: 352-60.

Saggin, L., L. Gorza, S. Ausoni, and S. Schiaffino. 1989. 'Troponin I switching in the developing heart', *J Biol Chem*, 264: 16299-302.

Schmidl, Christian, Andre F. Rendeiro, Nathan C. Sheffield, and Christoph Bock. 2015. 'ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors', *Nat Meth*, 12: 963-65.

Shimbo, T., Y. Du, S. A. Grimm, A. Dhasarathy, D. Mav, R. R. Shah, H. Shi, and P. A. Wade. 2013. 'MBD3 localizes at promoters, gene bodies and enhancers of active genes', *PLoS Genet*, 9: e1004028.

Shimbo, T., M. Takaku, and P. A. Wade. 2016. 'High-quality ChIP-seq analysis of MBD3 in human breast cancer cells', *Genom Data*, 7: 173-4.

Shin, H., T. Liu, A. K. Manrai, and X. S. Liu. 2009. 'CEAS: cis-regulatory element annotation system', *Bioinformatics*, 25: 2605-6.

Slagle, Christopher E., and Frank L. Conlon. 'Emerging Field of Cardiomics: High-Throughput Investigations into Transcriptional Regulation of Cardiovascular Development and Disease', *Trends in Genetics*, 32: 707-16.

Srinivasan, S., H. S. Baldwin, O. Aristizabal, L. Kwee, M. Labow, M. Artman, and D. H. Turnbull. 1998. 'Noninvasive, in utero imaging of mouse embryonic heart development with 40-MHz echocardiography', *Circulation*, 98: 912-8.

Svensson, Eric C., Gordon S. Huggins, Hua Lin, Cynthia Clendenin, Fang Jiang, Rachel Tufts, Fred B. Dardik, and Jeffrey M. Leiden. 2000. 'A syndrome of tricuspid atresia in mice with a targeted mutation of the gene encoding Fog-2', *Nat Genet*, 25: 353-56.

Turnbull, D. H., T. S. Bloomfield, H. S. Baldwin, F. S. Foster, and A. L. Joyner. 1995. 'Ultrasound backscatter microscope analysis of early mouse embryonic brain development', *Proc Natl Acad Sci U S A*, 92: 2239-43.

Wade, P. A., A. Gegonne, P. L. Jones, E. Ballestar, F. Aubry, and A. P. Wolffe. 1999. 'Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation', *Nat Genet*, 23: 62-6.

Wade, P. A., P. L. Jones, D. Vermaak, and A. P. Wolffe. 1998. 'A multiple subunit Mi-2 histone deacetylase from Xenopus laevis cofractionates with an associated Snf2 superfamily ATPase', *Curr Biol*, 8: 843-6.

Waldron, L., J. D. Steimle, T. M. Greco, N. C. Gomez, K. M. Dorr, J. Kweon, B. Temple, X. H. Yang, C. M. Wilczewski, I. J. Davis, I. M. Cristea, I. P. Moskowitz, and F. L. Conlon. 2016. 'The Cardiac TBX5 Interactome Reveals a Chromatin Remodeling Network Essential for Cardiac Septation', *Dev Cell*, 36: 262-75.

Wang, Q., R. S. Reiter, Q. Q. Huang, J. P. Jin, and J. J. Lin. 2001. 'Comparative studies on the expression patterns of three troponin T genes during mouse development', *Anat Rec*, 263: 72-84.

Williams, C. J., T. Naito, P. G. Arco, J. R. Seavitt, S. M. Cashman, B. De Souza, X. Qi, P. Keables, U. H. Von Andrian, and K. Georgopoulos. 2004. 'The chromatin remodeler Mi-2beta is required for CD4 expression and T cell development', *Immunity*, 20: 719-33.

Xue, Y., J. Wong, G. T. Moreno, M. K. Young, J. Cote, and W. Wang. 1998. 'NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities', *Mol Cell*, 2: 851-61.

Yamada, Tomoko, Yue Yang, Martin Hemberg, Toshimi Yoshida, Ha Young Cho, J. Patrick Murphy, Diasynou Fioravante, Wade G Regehr, Steven P Gygi, Katia Georgopoulos, and Azad Bonni. 2014. 'Promoter Decommissioning by the NuRD Chromatin Remodeling Complex Triggers Synaptic Connectivity in the Mammalian Brain', *Neuron*, 83: 122-34.

Yang, J., D. Hu, J. Xia, Y. Yang, B. Ying, J. Hu, and X. Zhou. 2000. 'Three novel TBX5 mutations in Chinese patients with Holt-Oram syndrome', *Am J Med Genet*, 92: 237-40.

Yang, Zhenyun, Marie Yamazaki, Qingwu W. Shen, and Darl R. Swartz. 2009. 'Differences between Cardiac and Skeletal Troponin Interaction with the Thin Filament Probed by Troponin Exchange in Skeletal Myofibrils', *Biophysical Journal*, 97: 183-94.

Yoshida, A., H. Morisaki, M. Nakaji, M. Kitano, K. S. Kim, K. Sagawa, S. Ishikawa, I. Satokata, Y. Mitani, H. Kato, K. Hamaoka, S. Echigo, I. Shiraishi, and T. Morisaki. 2016.

'Genetic mutation analysis in Japanese patients with non-syndromic congenital heart disease', *J Hum Genet*, 61: 157-62.

Yoshida, T, I Hazan, J Zhang, SY Ng, T Naito, HJ Snippert, EJ Heller, X Qi, LN Lawton, CJ Williams, and K Georgopoulos. 2008. 'The role of the chromatin remodeler Mi-2beta in hematopoietic stem cell self-renewal and multilineage differentiation', *Genes Dev*, 22: 1174 - 89.

Zaidi, S., M. Choi, H. Wakimoto, L. Ma, J. Jiang, J. D. Overton, A. Romano-Adesman, R. D. Bjornson, R. E. Breitbart, K. K. Brown, N. J. Carriero, Y. H. Cheung, J. Deanfield, S. DePalma, K. A. Fakhro, J. Glessner, H. Hakonarson, M. J. Italia, J. R. Kaltman, J. Kaski, R. Kim, J. K. Kline, T. Lee, J. Leipzig, A. Lopez, S. M. Mane, L. E. Mitchell, J. W. Newburger, M. Parfenov, I. Pe'er, G. Porter, A. E. Roberts, R. Sachidanandam, S. J. Sanders, H. S. Seiden, M. W. State, S. Subramanian, I. R. Tikhonova, W. Wang, D. Warburton, P. S. White, I. A. Williams, H. Zhao, J. G. Seidman, M. Brueckner, W. K. Chung, B. D. Gelb, E. Goldmuntz, C. E. Seidman, and R. P. Lifton. 2013. 'De novo mutations in histone-modifying genes in congenital heart disease', *Nature*, 498: 220-3.

Zhang, Y, G LeRoy, HP Seelig, WS Lane, and D Reinberg. 1998. 'The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities', *Cell*, 95: 279 - 89.

Zhang, Y. 2011. 'Biology of the Mi-2/NuRD Complex in SLAC (Stemness, Longevity/Ageing, and Cancer)', *Gene Regul Syst Bio*, 5: 1-26.

CHAPTER 3: CHD4 and the NuRD complex interact with GATA4 and NKX2-5 in the developing heart

Introduction

Congenital heart disease (CHD) is the leading cause of infant mortality in the United States (Joann et al. 2002; Petrini, Damus, and Johnston 1997). Despite extensive research identifying mutations in cardiac transcription factors such as *Nkx2-5, Gata4,* and *Tbx5* as responsible for a number of these cases, there is a persistent lack of understanding how these transcription factors function to instigate normal heart development (Chen et al. 2010; Garg et al. 2003; Hirayama-Yamada et al. 2005; Benson et al. 1999; Basson et al. 1999). Recent research on the roles of chromatin remodeling complexes in cardiac development may provide some explanation (Takeuchi et al. 2011; Bruneau 2010; Homsy et al. 2015; Bevilacqua, Willis, and Bultman 2014; Hang et al. 2010). However, interactions between these complexes and cardiac transcription factors causative to CHD remain incompletely understood. One such complex required for cardiac development is the Nucleosome Remodeling and Deacetylase (NuRD) complex.

The NuRD complex is composed of the ATP-dependent chromodomain-helicase-DNA binding protein 3 or 4 (CHD3/4), histone deacetylase 1 or 2 (HDAC1/2), methyl-CpG-binding domain protein 2 or 3 (MBD2/3), retinoblastoma-binding protein 4 or 7 (RBBP4/7), metastasis-associated protein 1, 2 or 3 (MTA1/2/3), GATA zinc-finger

domain containing protein 2A or B (GATAD2A/B) (Wade et al. 1999; Wade et al. 1998; Zhang et al. 1998). The critical nature of CHD4 in cardiac development is highlighted by the observation that mutations in CHD4 are causative to CHDs including atrial and ventricular septal and morphological defects (Homsy et al. 2015). Recent research from our lab demonstrated that CHD4 is required early in cardiac development to directly repress expression of non-cardiac myofibril isoforms, and loss of CHD4-mediated repression leads to misexpression of fast skeletal and smooth muscle myofibril isoforms, cardiac sarcomere malformation and early embryonic lethality (Wilczewski et al. 2018).

Several NuRD subunits have DNA-binding capabilities; however, the NuRD complex must be recruited to its target genes through interactions with sequence-specific co-factors. Recent research has revealed cooperative interactions between cardiac transcription factors causative to CHD and the NuRD complex during heart development (Waldron et al. 2016; Kaltenbrun et al. 2013; Garnatz et al. 2014; Roche et al. 2008). Ablation of these interactions leads to misexpression of target genes and structural defects in the heart. This raises the possibility of discovering other cofactors of this complex which may further inform the genetic etiology of CHD.

In this report, we identify three previously unknown transcriptional cofactors of CHD4 and the NuRD complex in the developing heart. Using an unbiased bioinformatics approach, we found cardiac transcription factor motifs to be highly overrepresented in CHD4-bound regions of the genome. Parallel reaction monitoring mass spectrometry (PRM-MS) in embryonic cardiac tissue uncovered interactions between CHD4 and NKX2-5, GATA4, SMAD3 and TBX5. Further analysis revealed

CHD4 and GATA4 or NKX2-5 converge on regulatory targets for transcriptional repression, including genes for fast skeletal and smooth muscle myofibril isoforms. These results represent a substantial advance in our understanding of how these well-studied transcription factors carry out their regulatory function.

Results

Cardiac transcription factor motifs overrepresented in CHD4-bound genomic elements

Previous work determining the role of CHD4 in cardiac development uncovered the direct regulatory targets of CHD4 during heart development through chromatin immunoprecipication followed by high-throughput sequencing (ChIP-seq) for CHD4 and transcriptional profiling (Wilczewski et al. 2018).To uncover transcriptional co-factors that could guide CHD4 to its regulatory targets through sequence-specific binding, we performed de novo motif discovery with HOMER (Heinz et al. 2010) using CHD4 ChIP-seq data from E10.0 wild-type hearts (Figure 3.1A). This analysis revealed a striking abundance of cardiac transcription factor consensus motifs in CHD4-bound regions (Figure 3.1B).

TBX5, ISL1, NKX2-5, and GATA4 are required for human heart development (Luo et al. 2014; Baban et al. 2014; Basson et al. 1997; Garg et al. 2003; Hirayama-Yamada et al. 2005). SMAD2/3/4 and Brachyury/T play critical roles in cardiac differentiation (Beyer et al. 2013; Johansson and Wiles 1995; Aramaki et al. 2013; David et al. 2011) while MEIS1, CRE-binding protein (CREB), and TEAD4, as a member of the Hippo signaling pathway, have been implicated in regulating cardiac homeostasis (Mahmoud et al. 2013; Xin et al. 2011; Fentzke et al. 1998). MEF2c is

expressed early in the cardiac lineage and is capable of reprograming fibroblasts to induced cardiomyocytes in conjunction with GATA4 and TBX5 (Edmondson et al. 1994; Ieda et al. 2010; Qian et al. 2013; Lin et al. 1997). CTCF is important in establishing chromatin architecture through formation of topologically associated domains (Bell, West, and Felsenfeld 1999; Parelho et al. 2008; Rubio et al. 2008). These motifs indicate CHD4 may interact with many more cardiac-specific transcriptional co-factors than previously thought.

CHD4 physically interacts with GATA4, NKX2-5, SMAD3 and TBX5

We wanted to identify whether these predicted associations physically occur the developing heart. We isolated E10.5 hearts from wild-type mice and performed immunoisolation of CHD4 complexes using anti-GFP immunoprecipitation as a negative control. Since many of these potential co-factors are low-abundance transcription factors, we utilized parallel reaction monitoring mass spectrometry (PRM-MS) to capture these infrequent and transient interactions by monitoring unique peptides for each potential co-factor. This analysis revealed interactions between CHD4 and GATA4, NKX2-5, SMAD3 and TBX5 in the developing heart (Figure 3.2). Interactions between CHD4 and SMAD3 have been previously demonstrated in human embryonic stem cells (Beyer et al. 2013), and we have previously reported an interaction between CHD4 and TBX5 in the developing heart (Waldron et al. 2016), which supports the validity of these results. To our knowledge, this is the first report of an interaction between GATA4 and NKX2-5 and CHD4.

NKX2-5 and GATA4 bind putative regulatory regions for fast skeletal and smooth muscle myofibril isoforms

To identify regulatory targets of these interactions, we analyzed genomic regions directly bound by CHD4 that contained GATA4 or NKX2-5 consensus motifs. We assessed whether these transcription factors bound these regulatory regions in mature cardiomyocytes or cardiac progenitor cells differentiated in culture using previously published ChIP-seq data (Luna-Zurita et al. 2016). Strikingly, we found that GATA4 binds a putative regulatory region for *Myh11* (encoding smooth muscle myosin heavy chain) in mature cardiomyocytes and cardiac precursors (Figure 3.3A). We also found that NKX2-5 binds a putative regulatory region for *Acta1* (encoding fast skeletal α-actin) in mature cardiomyocytes but not cardiac progenitor cells (Figure 3.3B). This data demonstrates that CHD4 likely interacts with these cardiac transcription factors to mediate repression of non-cardiac myofibril isoforms during cardiac development.

Discussion

We have demonstrated that CHD4 interacts with the transcription factors GATA4, NKX2-5, SMAD3 and TBX5 in the developing heart. We also present evidence indicating that CHD4 is recruited to repress fast skeletal and smooth muscle myofibril isoforms through its interactions with GATA4 and NKX2-5.

GATA4 is a well-studied transcription factor in cardiac development and cardiomyocyte differentiation. GATA4 has been demonstrated to be a non-histone target of NuRD complex member HDAC2; however, this interaction did not include a role for direct transcriptional regulation of target genes. In comparison, GATA4

interaction with co-repressor FOG-2 leads to transcriptional repression of some of its target genes (Svensson et al. 1999). FOG-2 has been shown to interact with members of the NuRD complex, particularly MTA1 and RBBP 4/7, to repress gene expression of cell cycle inhibitors; loss of this interaction leads to perinatal lethality with septal defects and a thin myocardium (Garnatz et al. 2014; Roche et al. 2008). The data presented in this report demonstrates that GATA4 can also interact with the NuRD complex through CHD4; however, it remains unclear whether this interaction is dependent on FOG-2 or if GATA4 can interact with NuRD in a FOG-2 independent manner to modulate expression of a different set of target genes. GATA4 binding to the direct CHD4 target *Myh11*, whose upregulation was not reported in FOG-2 mutants, suggests this may be the case.

NKX2-5 is another well-studied transcription factor in cardiac development and differentiation demonstrated to act through transcriptional activation and repression. The mechanism by which NKX2-5 engages in transcriptional repression has not been thoroughly explored. This report therefore demonstrates one potential avenue for NKX2-mediated repression of regulatory targets. Interestingly, a separate study found NKX2-5 acts to directly repress expression of different fast skeletal myofibril subunits (*Tnnt3* and *Tnni2*) in the developing heart but did not mention upregulation of *Acta1* (Dupays et al. 2015). While CHD4 has been demonstrated to bind putative regulatory regions for both *Tnnt3* and *Tnni2* in the developing heart, these regions differ from the proposed NKX2-5 binding site (Wilczewski et al. 2018). Therefore, whether NKX2-5 interacts with the NuRD complex to repress these targets or acts through an alternative, unknown mechanism remains to be determined.

PRM-MS analysis in embryonic cardiac tissue found that the previously discovered interaction between CHD4 and SMAD3 in human embryonic stem cells (hESC) is maintained in the developing heart. In hESCs, the NuRD complex interacts with OCT4, TEAD, and SMAD2/3 to repress mesendoderm genes to maintain pluripotency (Beyer et al. 2013). Induction of mesoderm specification lead to the loss of TEAD binding at these sites and SMAD2/3-induced gene activation. Interestingly, de novo motif discovery also predicted TEAD4 as a co-factor of CHD4; however, this interaction was not detected in embryonic cardiac tissue. This result could indicate that CHD4 and the NuRD complex maintain their interaction with SMAD3 through cardiomyocyte differentiation after mesoderm specification without TEAD to regulate a different set of target genes during cardiac development. Further exploration of CHD4-SMAD3 targets is needed to identify what role this interaction plays during the later stages of differentiation and cardiac development.

This report also validated the interaction we previously reported between CHD4 and TBX5, albeit one day later in cardiac development (Waldron et al. 2016). Further analysis is necessary to determine whether direct targets predicted by CHD4 ChIP-seq at E10.0 uncovers new gene regulatory targets for this interaction as cardiac development proceeds.

This report demonstrates that several cardiac transcription factors engage CHD4 and the NuRD complex to repress gene targets through interactions at cis-regulatory regions. Evaluating other potential targets of these interactions may yield novel targets of these well-studied transcription factors and inform the molecular mechanisms underlying cases of congenital heart disease.

This work was completed in collaboration with Xinlei Sheng, Austin J. Hepperla, Todd Greco, Ian J. Davis and Ileana M. Cristea. This work was funded by R01 HL112618 and R01 HL127640 to F.L.C., and 5T32 HL069768 and 1F31 HL136100 to C.M.W.

Materials and Methods

Motif Discovery

De novo motif discovery on CHD4 chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) regions (Wilczewski et al. 2018) (GSE109012) from embryonic day (E)10.0 cardiac tissue was performed using Hypergeometric Optimization of Motif EnRichment (HOMER) (Heinz et al. 2010).

CHD4 Immunoprecipitation

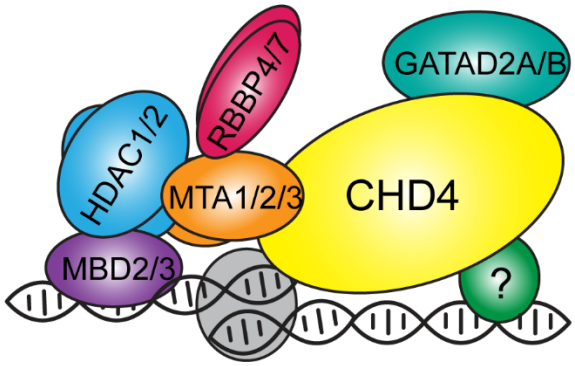
E10.5 wild-type CD1 hearts (minimum twenty-eight hearts per IP, three separate biological replicates) were harvested in cold PBS, snap frozen and cryolysed as previously described (Kaltenbrun et al. 2013). Frozen tissue powder was resuspended in optimized lysis buffer (5ml/1g powder) (20mM K-HEPES pH7.4, 0.11M KOAC, 2mM MgCl2, 0.1% Tween 20, 1μM ZnCl2, 1mM CaCl2, 0.5% Triton-X 100, 150mM NaCl2, protease inhibitor (Sigma), phosphatase inhibitor (Sigma)). CHD4 complexes were solubilized and immunoprecipitated as previously described (Kaltenbrun et al. 2013; Cristea and Chait 2011) using rabbit anti-CHD4 antibody (Abcam #ab72418) or negative control custom rabbit anti-GFP antibody (Cristea et al. 2005) with elution at 95°C for 10 minutes. The immunoisolated proteins were resolved (~ 1 cm) by SDS-

PAGE, and visualized by Coomassie blue. Each lane was subjected to in-gel digestion with trypsin as previously reported (Waldron et al. 2016).

Parallel Reaction Monitoring (PRM) targeted Mass Spectrometry (MS)

Desalted peptides (2ul each) were analyzed by nano-liquid chromatography (nLC)-MS/MS using a Dionex Ultimate 3000 nRSLC system directly coupled to a Q exactive HF orbitrap mass spectrometer (ThermoFisher Scientific) equipped with an EASY-Spray ion source (Thermo, Fisher Scientific). Peptide mixtures were evaporated in vacuo and resuspended in 1% Trifluoroacetic acid/4%acetonitrile/95% water and loaded onto an 50 cm long column with a 75 um internal diameter (PepMap) and separated over a 60 min gradient with a mobile phase containing aqueous acetonitrile and 0.1% formic acid programmed from 1% to 3% acetonitrile over 12 min, then 3 to 35% acetonitrile over 60 min, then 35 to 97% acetonitrile over 1 min, followed by 10 min at 97% acetonitrile and 97 to 70% over 1 min, all at a flow rate of 250 nL/min. The PRM method consisted of targeted MS2 scans at a resolution of 60,000 and with an AGC value of 1x10E6, a max injection time of 500 ms, a 0.8 m/z isolation window, a fixed first mass of 150 m/z, and normalized collision energy of 27, which were recorded as profile data. The targeted MS2 methods were controlled with a timed inclusion list containing the target precursor m/z value, charge, and a retention time window that were determined from shotgun analysis.

A

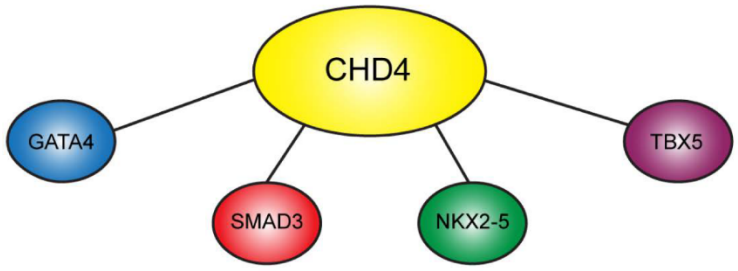


B

Consensus sequence	Motif name	q-value
	CTCF	0
	MEIS1	0
	TBX5	0
	SMAD3	0
	ISL1	0
	SMAD2	0
	SMAD4	0
	NKX2-5	0
	GATA4	0
	Brachyury (T-box)	0
	TEAD4	0.0107
	CRE	0.0062
	MEF2C	0.0178

Figure 3.1. CHD4 bound genomic regions are enriched for cardiac transcription factor motifs. (A) Schematic for de novo motif discovery from CHD4 ChIP-seq data. (B) HOMER motif results from E10.0 CHD4 ChIP-seq peaks. q-value represents Benjamini-Hotchberg correction for multiple comparisons.

A



B

Protein	Monitored Peptide	Average Enrichment Ratio (CHD4 IP/GFP IP)	% Covariance
CHD4	VGGNIEVLGFNAR	**51.4**	92.1
GATA4	EAAAYGSGGGAAGAGLAGR	**321.9**	160.5
SMAD3	AITTQNVNTK	**33.4**	93.2
NKX2-5	SLASGDLSAR	**26.5**	124.9
NKX2-5	VLFSQAQVYELER	**47.9**	105.5
TBX5	SGYPQQQGLSTSYR	**69.4**	137.6

Figure 3.2. CHD4 interacts with a subset of cardiac transcription factors in the developing heart. (A) Schematic showing CHD4 interaction partners detected by parallel reaction monitoring mass spectrometry at E10.5. (B) Table showing average enrichment for each monitored peptide in CHD4 immunoprecipitation over GFP negative control immunoprecipitation. Percent covariance represents the covariance of the average enrichment over three independent biological replicates.

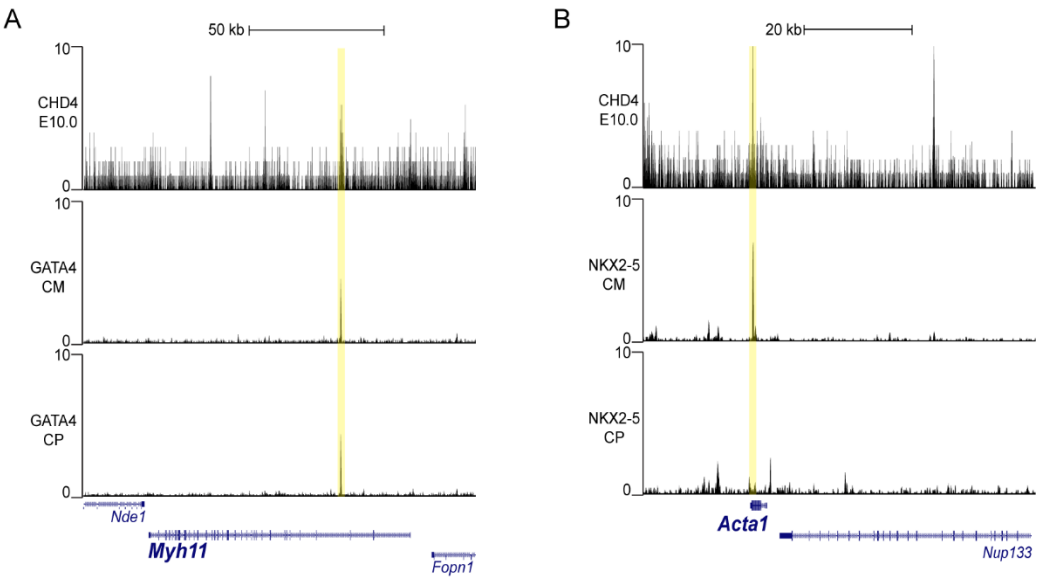


Figure 3.3. CHD4 and cardiac transcription factors are both present at putative regulatory loci. (A) CHD4 bound regulatory region in *Myh11* with GATA4 consensus motif in E10.0 cardiac tissue is also bound by GATA4 in mature cardiomyocytes (CM) and cardiac progenitors (CP). (B) CHD4 bound regulatory region in *Acta1* with NKX2-5 consensus motif in E10.0 cardiac tissue is also bound by NKX2-5 in mature cardiomyocytes (CM) but not cardiac progenitors (CP).

REFERENCES

Aramaki, Shinya, Katsuhiko Hayashi, Kazuki Kurimoto, Hiroshi Ohta, Yukihiro Yabuta, Hiroko Iwanari, Yasuhiro Mochizuki, Takao Hamakubo, Yuki Kato, Katsuhiko Shirahige, and Mitinori Saitou. 2013. 'A Mesodermal Factor, T, Specifies Mouse Germ Cell Fate by Directly Activating Germline Determinants', *Dev Cell*, 27: 516-29.

Baban, A., A. V. Postma, M. Marini, G. Trocchio, A. Santilli, M. Pelegrini, P. Sirleto, M. Lerone, S. B. Albanese, P. Barnett, C. J. Boogerd, B. Dallapiccola, M. C. Digilio, R. Ravazzolo, and G. Pongiglione. 2014. 'Identification of TBX5 mutations in a series of 94 patients with Tetralogy of Fallot', *Am J Med Genet A*, 164a: 3100-7.

Basson, C. T., D. R. Bachinsky, R. C. Lin, T. Levi, J. A. Elkins, J. Soults, D. Grayzel, E. Kroumpouzou, T. A. Traill, J. Leblanc-Straceski, B. Renault, R. Kucherlapati, J. G. Seidman, and C. E. Seidman. 1997. 'Mutations in human TBX5 [corrected] cause limb and cardiac malformation in Holt-Oram syndrome', *Nat Genet*, 15: 30-5.

Basson, C. T., T. Huang, R. C. Lin, D. R. Bachinsky, S. Weremowicz, A. Vaglio, R. Bruzzone, R. Quadrelli, M. Lerone, G. Romeo, M. Silengo, A. Pereira, J. Krieger, S. F. Mesquita, M. Kamisago, C. C. Morton, M. E. Pierpont, C. W. Muller, J. G. Seidman, and C. E. Seidman. 1999. 'Different TBX5 interactions in heart and limb defined by Holt-Oram syndrome mutations', *Proc Natl Acad Sci U S A*, 96: 2919-24.

Bell, A. C., A. G. West, and G. Felsenfeld. 1999. 'The protein CTCF is required for the enhancer blocking activity of vertebrate insulators', *Cell*, 98: 387-96.

Benson, D. W., G. M. Silberbach, A. Kavanaugh-McHugh, C. Cottrill, Y. Zhang, S. Riggs, O. Smalls, M. C. Johnson, M. S. Watson, J. G. Seidman, C. E. Seidman, J. Plowden, and J. D. Kugler. 1999. 'Mutations in the cardiac transcription factor NKX2.5 affect diverse cardiac developmental pathways', *J Clin Invest*, 104: 1567-73.

Bevilacqua, A., M. S. Willis, and S. J. Bultman. 2014. 'SWI/SNF chromatin-remodeling complexes in cardiovascular development and disease', *Cardiovasc Pathol*, 23: 85-91.

Beyer, Tobias A, Alexander Weiss, Yuliya Khomchuk, Kui Huang, Abiodun A Ogunjimi, Xaralabos Varelas, and Jeffrey L Wrana. 2013. 'Switch Enhancers Interpret TGF-β and Hippo Signaling to Control Cell Fate in Human Embryonic Stem Cells', *Cell Reports*, 5: 1611-24.

Bruneau, Benoit G. 2010. 'Chromatin remodeling in heart development', *Current Opinion in Genetics & Development*, 20: 505-11.

Chen, Y., Z. Q. Han, W. D. Yan, C. Z. Tang, J. Y. Xie, H. Chen, and D. Y. Hu. 2010. 'A novel mutation in GATA4 gene associated with dominant inherited familial atrial septal defect', *J Thorac Cardiovasc Surg*, 140: 684-7.

Cristea, I. M., R. Williams, B. T. Chait, and M. P. Rout. 2005. 'Fluorescent proteins as proteomic probes', *Mol Cell Proteomics*, 4: 1933-41.

Cristea, Ileana M., and Brian T. Chait. 2011. 'Conjugation of Magnetic Beads for Immunopurification of Protein Complexes', *Cold Spring Harb Protoc*, 2011: pdb.prot5610.

David, Robert, Veronica Barbara Jarsch, Florian Schwarz, Petra Nathan, Moritz Gegg, Heiko Lickert, and Wolfgang-Michael Franz. 2011. 'Induction of MesP1 by Brachyury(T) generates the common multipotent cardiovascular stem cell', *Cardiovasc Res*, 92: 115-22.

Dupays, Laurent, Catherine Shang, Robert Wilson, Surendra Kotecha, Sophie Wood, Norma Towers, and Timothy Mohun. 2015. 'Sequential Binding of MEIS1 and NKX2-5 on the Popdc2 Gene: A Mechanism for Spatiotemporal Regulation of Enhancers during Cardiogenesis', *Cell Reports*, 13: 183-95.

Edmondson, D. G., G. E. Lyons, J. F. Martin, and E. N. Olson. 1994. 'Mef2 gene expression marks the cardiac and skeletal muscle lineages during mouse embryogenesis', *Development*, 120: 1251-63.

Fentzke, R. C., C. E. Korcarz, R. M. Lang, H. Lin, and J. M. Leiden. 1998. 'Dilated cardiomyopathy in transgenic mice expressing a dominant-negative CREB transcription factor in the heart', *J Clin Invest*, 101: 2415-26.

Garg, V., I. S. Kathiriya, R. Barnes, M. K. Schluterman, I. N. King, C. A. Butler, C. R. Rothrock, R. S. Eapen, K. Hirayama-Yamada, K. Joo, R. Matsuoka, J. C. Cohen, and D. Srivastava. 2003. 'GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5', *Nature*, 424: 443-7.

Garnatz, A. S., Z. Gao, M. Broman, S. Martens, J. U. Earley, and E. C. Svensson. 2014. 'FOG-2 mediated recruitment of the NuRD complex regulates cardiomyocyte proliferation during heart development', *Dev Biol*, 395: 50-61.

Hang, Calvin T., Jin Yang, Pei Han, Hsiu-Ling Cheng, Ching Shang, Euan Ashley, Bin Zhou, and Ching-Pin Chang. 2010. 'Chromatin regulation by Brg1 underlies heart muscle development and disease', *Nature*, 466: 62-67.

Heinz, S., C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh, and C. K. Glass. 2010. 'Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities', *Mol Cell*, 38: 576-89.

Hirayama-Yamada, K., M. Kamisago, K. Akimoto, H. Aotsuka, Y. Nakamura, H. Tomita, M. Furutani, S. Imamura, A. Takao, M. Nakazawa, and R. Matsuoka. 2005. 'Phenotypes with GATA4 or NKX2.5 mutations in familial atrial septal defect', *Am J Med Genet A*, 135: 47-52.

Homsy, J., S. Zaidi, Y. Shen, J. S. Ware, K. E. Samocha, K. J. Karczewski, S. R. DePalma, D. McKean, H. Wakimoto, J. Gorham, S. C. Jin, J. Deanfield, A. Giardini, G. A. Porter, Jr., R. Kim, K. Bilguvar, F. Lopez-Giraldez, I. Tikhonova, S. Mane, A. Romano-Adesman, H. Qi, B. Vardarajan, L. Ma, M. Daly, A. E. Roberts, M. W. Russell, S. Mital, J. W. Newburger, J. W. Gaynor, R. E. Breitbart, I. Iossifov, M. Ronemus, S. J. Sanders, J. R. Kaltman, J. G. Seidman, M. Brueckner, B. D. Gelb, E. Goldmuntz, R. P. Lifton, C. E. Seidman, and W. K. Chung. 2015. 'De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies', *Science*, 350: 1262-6.

Ieda, M., J. D. Fu, P. Delgado-Olguin, V. Vedantham, Y. Hayashi, B. G. Bruneau, and D. Srivastava. 2010. 'Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors', *Cell*, 142: 375-86.

Joann, Petrini, Damus Karla, Russell Rebecca, Poschman Karalee, Davidoff Michael J., and Mattison Donald. 2002. 'Contribution of birth defects to infant mortality in the United States', *Teratology*, 66: S3-S6.

Johansson, B M, and M V Wiles. 1995. 'Evidence for involvement of activin A and bone morphogenetic protein 4 in mammalian mesoderm and hematopoietic development', *Mol Cell Biol*, 15: 141-51.

Kaltenbrun, E., T. M. Greco, C. E. Slagle, L. M. Kennedy, T. Li, I. M. Cristea, and F. L. Conlon. 2013. 'A Gro/TLE-NuRD corepressor complex facilitates Tbx20-dependent transcriptional repression', *J Proteome Res*, 12: 5395-409.

Lin, Q., J. Schwarz, C. Bucana, and E. N. Olson. 1997. 'Control of mouse cardiac morphogenesis and myogenesis by transcription factor MEF2C', *Science*, 276: 1404-7.

Luna-Zurita, L., C. U. Stirnimann, S. Glatt, B. L. Kaynak, S. Thomas, F. Baudin, M. A. Samee, D. He, E. M. Small, M. Mileikovsky, A. Nagy, A. K. Holloway, K. S. Pollard, C. W. Muller, and B. G. Bruneau. 2016. 'Complex Interdependence Regulates Heterotypic Transcription Factor Distribution and Coordinates Cardiogenesis', *Cell*.

Luo, Z. L., H. Sun, Z. Q. Yang, Y. H. Ma, Y. Gu, Y. Q. He, D. Wei, L. B. Xia, B. H. Yang, and T. Guo. 2014. 'Genetic variations of ISL1 associated with human congenital heart disease in Chinese Han people', *Genet Mol Res*, 13: 1329-38.

Mahmoud, A. I., F. Kocabas, S. A. Muralidhar, W. Kimura, A. S. Koura, S. Thet, E. R. Porrello, and H. A. Sadek. 2013. 'Meis1 regulates postnatal cardiomyocyte cell cycle arrest', *Nature*, 497: 249-53.

Parelho, V., S. Hadjur, M. Spivakov, M. Leleu, S. Sauer, H. C. Gregson, A. Jarmuz, C. Canzonetta, Z. Webster, T. Nesterova, B. S. Cobb, K. Yokomori, N. Dillon, L. Aragon, A. G. Fisher, and M. Merkenschlager. 2008. 'Cohesins functionally associate with CTCF on mammalian chromosome arms', *Cell*, 132: 422-33.

Petrini, J., K. Damus, and R. B. Johnston, Jr. 1997. 'An overview of infant mortality and birth defects in the United States', *Teratology*, 56: 8-10.

Qian, L., E. C. Berry, J. D. Fu, M. Ieda, and D. Srivastava. 2013. 'Reprogramming of mouse fibroblasts into cardiomyocyte-like cells in vitro', *Nat Protoc*, 8: 1204-15.

Roche, A. E., B. J. Bassett, S. A. Samant, W. Hong, G. A. Blobel, and E. C. Svensson. 2008. 'The zinc finger and C-terminal domains of MTA proteins are required for FOG-2-mediated transcriptional repression via the NuRD complex', *J Mol Cell Cardiol*, 44: 352-60.

Rubio, E. D., D. J. Reiss, P. L. Welcsh, C. M. Disteche, G. N. Filippova, N. S. Baliga, R. Aebersold, J. A. Ranish, and A. Krumm. 2008. 'CTCF physically links cohesin to chromatin', *Proc Natl Acad Sci U S A*, 105: 8309-14.

Svensson, Eric C., Rachel L. Tufts, Christine E. Polk, and Jeffrey M. Leiden. 1999. 'Molecular cloning of FOG-2: A modulator of transcription factor GATA-4 in cardiomyocytes', *Proceedings of the National Academy of Sciences*, 96: 956-61.

Takeuchi, Jun K., Xin Lou, Jeffrey M. Alexander, Hiroe Sugizaki, Paul Delgado-Olguín, Alisha K. Holloway, Alessandro D. Mori, John N. Wylie, Chantilly Munson, Yonghong Zhu, Yu-Qing Zhou, Ru-Fang Yeh, R. Mark Henkelman, Richard P. Harvey, Daniel Metzger, Pierre Chambon, Didier Y. R. Stainier, Katherine S. Pollard, Ian C. Scott, and Benoit G. Bruneau. 2011. 'Chromatin remodelling complex dosage modulates transcription factor function in heart development', *Nat Commun*, 2: 187.

Wade, P. A., P. L. Jones, D. Vermaak, and A. P. Wolffe. 1998. 'A multiple subunit Mi-2 histone deacetylase from Xenopus laevis cofractionates with an associated Snf2 superfamily ATPase', *Curr Biol*, 8: 843-6.

Wade, PA, A Gegonne, PL Jones, E Ballestar, F Aubry, and AP Wolffe. 1999. 'Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation', *Nat Genet*, 23: 62 - 66.

Waldron, L., J. D. Steimle, T. M. Greco, N. C. Gomez, K. M. Dorr, J. Kweon, B. Temple, X. H. Yang, C. M. Wilczewski, I. J. Davis, I. M. Cristea, I. P. Moskowitz, and F. L. Conlon. 2016. 'The Cardiac TBX5 Interactome Reveals a Chromatin Remodeling Network Essential for Cardiac Septation', *Dev Cell*, 36: 262-75.

Wilczewski, Caralynn M., Austin J. Hepperla, Takashi Shimbo, Lauren Wasson, Zachary L. Robbe, Ian J. Davis, Paul A. Wade, and Frank L. Conlon. 2018. 'CHD4 and the NuRD complex directly control cardiac sarcomere formation', *Proceedings of the National Academy of Sciences*.

Xin, M., Y. Kim, L. B. Sutherland, X. Qi, J. McAnally, R. J. Schwartz, J. A. Richardson, R. Bassel-Duby, and E. N. Olson. 2011. 'Regulation of insulin-like growth factor signaling by Yap governs cardiomyocyte proliferation and embryonic heart size', *Sci Signal*, 4: ra70.

Zhang, Y, G LeRoy, HP Seelig, WS Lane, and D Reinberg. 1998. 'The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities', *Cell*, 95: 279 - 89.

CHAPTER 4: Discussion and Future Directions

Dynamic alterations in chromatin accessibility is one mechanism of gene regulation during cardiomyocyte differentiation and embryonic heart development. CHD4 and the NuRD complex are essential components of this regulatory framework, as previously demonstrated in ESCs and mesoderm specification (Beyer et al. 2013; O'Shaughnessy-Kirwan et al. 2015; Linder et al. 2007). However, its role in later stages of cardiomyocyte differentiation and its functional significance on embryonic heart development remain relatively unknown. This set of studies sought to examine the requirement for the NuRD complex during cardiac development by conditionally ablating its catalytic core subunit, CHD4, and utilized transcriptomic, genomic, proteomic, and phenotypic methodology to uncover the NuRD complex's functional significance during early heart development. We uncovered a requirement for CHD4 in the repression of skeletal and smooth muscle cells in cardiomyocytes that is essential for cardiac function and embryonic development.

CHD4 and the NuRD complex directly control cardiac sarcomere formation

Chapter 2 established a role for CHD4 and the NuRD complex as a negative regulator of skeletal and smooth muscle myofibril subunits. We demonstrated that CHD4 binds to proximal and distal regulatory regions in the developing heart to both positively and negatively regulate gene expression. We found CHD4 directly represses

smooth muscle myosin heavy chain, fast skeletal α-actin, and the fast skeletal troponin complex in cardiomyocytes in the developing heart. In the absence of CHD4-mediated repression, these non-cardiac myofibril subunits are expressed in cardiomyocytes, intercalate into the nascent cardiac sarcomere and interfere with sarcomere formation. Ultimately, this leads to a failure to properly initiate systolic function in the developing heart and embryonic lethality.

CHD4 directly modulates transcription at active and repressed target genes

The NuRD complex has in many cases been demonstrated to act as a transcriptional repressor (Musselman et al. 2012; O'Shaughnessy-Kirwan et al. 2015; Pfefferli et al. 2014; Yamada et al. 2014). However, its role modulating the expression of actively transcribed genes has recently become more appreciated (Shimbo et al. 2013; Allen, Wade, and Kutateladze 2013; Hung, Kohnken, and Svaren 2012; Ingram et al. 2013). While our findings suggest that the cardiac defects and embryonic lethality in the absence of CHD4 are due to lost repression of non-cardiac myofibril genes, we also found a significant number of genes that are directly bound by CHD4 and are downregulated upon its ablation. While these directly regulated targets are not represented by any significant gene ontology terms, it is an interesting finding that fits with the emerging model of the NuRD complex as a transcriptional modulator, rather than solely a transcriptional repressor.

One proposed hypothesis is that the NuRD complex is responsible for participating in the rapid repositioning of nucleosomes and fast turnover of histone modifications in the active regions of the genome, which balances the activating

properties of the BAF complex and histone acetyltransferases (Allen, Wade, and Kutateladze 2013; Hu and Wade 2012; Shimono et al. 2003). An example of this interplay between activation and repression at actively expressed genes has been demonstrated in vascular development (Curtis and Griffin 2012). It would be intriguing to assess the state of histone modifications at these directly regulated and actively expressed targets of CHD4 using previously published datasets in cardiomyocytes (Paige et al. 2012; Wamstad et al. 2012). This analysis could clarify whether these sites contain signatures of actively expressed, repressed or bivalent histone modifications and provide context for when CHD4 will act as a transcriptional activator or repressor in cardiomyocytes. Furthermore, assessing whether the BAF complex is present by analyzing BRG1 ChIP-seq data from embryonic cardiac tissue (Hang et al. 2010) would inform potential mechanisms of transcriptional modulation at these target genes.

Another potential explanation for decreased expression of these direct targets could be a decrease in recruitment of shared co-factors upon CHD4 depletion, such as TBX5, GATA4, or NKX2-5, which have all been demonstrated to have dual roles in transcriptional activation and repression (Dupays et al. 2015; Svensson et al. 1999; Dai and Markham 2001; Waldron et al. 2016). This could be evaluated by performing ChIP-seq for each of these factors in the presence and absence of CHD4 in the developing heart.

CHD4 and the NuRD complex repress alternative cell fates in cardiomyocyte differentiation

With our conditional knockout mouse model system, we were able to ascertain the role of CHD4 in cardiac development and cardiomyocyte differentiation in an endogenous developmental context. Utilizing *Nkx2-5* as the conditional Cre driver enabled us to ablate CHD4 early in this process, as *Cre* recombinase expression is activated at cardiac precursor specification upon induction of *Nkx2-5* expression (Moses et al. 2001; Lien et al. 1999).

This approach allowed us to circumvent the early embryonic requirements for CHD4 in gastrulation and mesoderm specification (Linder et al. 2007; O'Shaughnessy-Kirwan et al. 2015). However, due to the timing of *Cre* activation and *Chd4* ablation, we cannot draw conclusions regarding the role of CHD4 and the NuRD complex in cardiac precursor specification. To investigate the role of CHD4 during this step of cardiomyocyte differentiation, the *Chd4* conditional allele could be crossed to a *Mesp1* Cre driver (Saga et al. 1999). Classification of the effects of this conditional ablation on cardiac precursor specification would provide information regarding the role of CHD4 during these earlier stages of cardiac progenitor specification in the endogenous context of cardiac development.

This study demonstrates a critical role for CHD4/NuRD in cardiomyocyte differentiation through the repression of other muscle type genes. Previous research has demonstrated the existence of a bipotential myocardial and smooth muscle cell precursor in murine development as late as E9.5 (Wu et al. 2006). Our research implicates CHD4 and the NuRD complex as a determining factor for cardiomyocyte cell

fate in this process as we demonstrated that CHD4 is required for repression of the smooth muscle gene *Myh11*. Furthermore, the NuRD subunits HDAC1/2 have been shown to be directly responsible for repression of fast skeletal *Tnni2* and repression of fast skeletal *Tnnt3* in the adult heart (Montgomery et al. 2007). These findings were recapitulated by ablating CHD4 in adult mice, which led to severe cardiomyopathy, impaired cardiac function and sudden death (Gómez-del Arco et al.). These findings indicate that CHD4 and the NuRD complex are required to repress markers of alternative cell fates, specifically smooth and skeletal muscle, to maintain cardiomyocyte differentiation, which complement our studies demonstrating its requirement for the establishment of differentiated cardiomyocytes.

Cardiac function defects observed in the presence of skeletal and smooth muscle myofibril proteins

Cardiomyocyte differentiation is required to form the first functional organ in developing vertebrates capable of circulating oxygen and nutrients throughout the rapidly growing embryo (Hirschy et al. 2006). Differentiation is characterized by expression of cardiac-specific myofibril proteins that arrange in a stereotypical manner to form a functional cardiac sarcomere capable of contracting in response to calcium (Ehler et al. 1999; Nishii et al. 2008; Sehnert et al. 2002; Takimoto and Kass 2012; van Eldik and Passier 2013). We have demonstrated that in the absence of CHD4, differentiation is impaired by skeletal and smooth muscle myofibril subunits disrupting sarcomere formation and cardiomyocyte function.

While other studies have found misexpression of skeletal and smooth muscle genes in genetic models of heart development and disease, these studies failed to address the functional significance of non-cardiac myofibril protein expression (Heidersbach et al. 2013; Montgomery et al. 2007; Gómez-del Arco et al. 2016; Petchey et al. 2014; Risebro et al. 2009). Through ultramicroscopy, quantitative immunofluorescence and embryonic echocardiography, we established the functional consequences of alternate muscle myofibril misexpression. An interesting question these studies did not address was how each different myofibril type contributes to cardiomyocyte function defects To address this question, we could parse out the effect of fast skeletal troponin complex misexpression on calcium cycling rates from contractility defects by performing calcium imaging on CHD4-null cardiomyotyes using commercially available imaging techniques or patch-clamp experiments (Qian et al. 2012; Golebiewska and Scarlata 2011; Guatimosim, Guatimosim, and Song 2011). This would address whether hybrid cardiac and fast skeletal troponin complexes altered cardiac function in CHD4 null cardiomyocytes, or whether substituting smooth muscle myosin heavy chain and skeletal α-actin, or a combination of both mechanisms, resulted in cardiac function defects.

Implications for human cases of cardiomyopathy

For patients affected by cardiac defects such as dilated cardiomyopathy, left ventricular noncompaction, right ventricular volume overload, and hypertrophic cardiomyopathy, this study may provide mechanistic insight to the development of their disease. Several studies have identified mutations in cardiac sarcomere genes as

causative to these types of conditions (Kamisago et al. 2000; Morita et al. 2008; Gerull et al. 2002; Klaassen et al. 2008; Kaski et al. 2008). However, the results of our research demonstrate that misexpression of skeletal or smooth muscle myofibril isoforms is equally capable of causing cardiac sarcomere defects that lead to impaired cardiac function. To this point, research on human cases of cardiac hypertrophy and cardiomyopathy has found upregulation of skeletal α-actin in myocardial tissue (Suurmeijer et al. 2003). This evidence suggests a compelling hypothesis that previously unexamined genetic causes, such as mutations in transcriptional repressors such as CHD4, may be at the root of some of these cases.

These studies therefore provide a potential mechanism for congenital heart disease patients with mutations in *CHD4* (Homsy et al. 2015; Sifrim et al. 2016; Weiss et al. 2016). A subset of these patients show defects such as dilated right atrium with right ventricular volume overload, which could be attributed to defects in cardiac sarcomere contractility. If it were possible to obtain samples of myocardial tissue from these patients, assessing the expression of skeletal and/or smooth muscle myofibril proteins may provide a mechanistic insight to the pathogenesis of their disease. In lieu of these difficult to obtain tissue samples, we have initiated a collaboration with Jonathan and Christine Seidman's laboratories at Harvard University to engineer these patient missense mutations into *CHD4* in human ESCs. By differentiating these cells into cardiomyocytes, then assaying direct transcriptional control by CHD4 ChIP-seq and transcriptomic analysis we aim to address whether CHD4 engages in the same gene regulatory mechanisms during human cardiomyocyte differentiation as in murine cardiac development. In the future, it may be possible to engineer gene therapies to treat

human cardiac disease by targeting these misexpressed gene products for degradation before they cause detrimental effects to cardiac sarcomere formation and cardiac function.

CHD4 and the NuRD complex interact with GATA4 and NKX2-5 in the developing heart

Considering CHD4 and the NuRD complex require sequence-specific co-factors to localize to regulatory targets, we were interested in identifying co-factors responsible for guiding CHD4 to its regulatory targets during cardiomyocyte differentiation. In Chapter 3, we used proteomic and computational bioinformatics to establish an interaction between CHD4 and the cardiac transcription factors GATA4 and NKX2-5 in the developing heart. We found that in differentiating cardiomyocytes, these transcription factors localize to putative regulatory regions for skeletal and smooth muscle myofibril isoforms which are directly repressed by CHD4 in the developing heart. We found that CHD4 maintains its interaction with TBX5 in the developing heart from E9.5 to E10.5 (Waldron et al. 2016). Furthermore, we found that CHD4 interacts with SMAD3, a finding previously reported in ESCs but reported for the first time here in the heart (Beyer et al. 2013).

Predicted cardiac co-factors of CHD4 and the NuRD complex

De novo motif discovery of CHD4-bound genomic regions revealed DNA binding motifs for several transcription factors with demonstrated roles in cardiac development and disase (Luo et al. 2014; Baban et al. 2014; Basson et al. 1997; Garg et al. 2003;

Hirayama-Yamada et al. 2005; Beyer et al. 2013; Johansson and Wiles 1995; Aramaki et al. 2013; David et al. 2011; Mahmoud et al. 2013; Xin et al. 2011; Fentzke et al. 1998; Edmondson et al. 1994; Ieda et al. 2010; Qian et al. 2012; Lin et al. 1997). While proteomic analysis of E10.5 hearts demonstrated CHD4 interacts with at least four of these predicted co-factors, several predicted co-factors such as ISL1, Brachyury/T, TEAD4, CREB1, CTCF, MEIS1, and SMAD2 were not evident in CHD4 immunoisolated complexes by parallel reaction monitoring mass spectrometry. It is important not to rule out interaction of these potential co-factors with CHD4 in the developing heart based on this incidence of absence of evidence; while PRM-MS is an advanced technology capable of detecting specific proteins from small amounts of input material, it remains a possibility that the candidate peptides selected for monitoring were not ideally suited for detection by the mass spectrometer. Furthermore, the stringent criteria for candidate peptide selection meant detection in most cases relied on one single candidate peptide per protein, increasing the likelihood that some potential co-factors may miss detection. Increasing the input material could circumvent some of these technical issues and result in increased detection of these predicted co-factors.

Alternatively, it remains a possibility that while these predicted co-factors have DNA binding motifs within CHD4-bound regions, they do not interact with CHD4 in the developing heart to regulate those targets. For example, ISL1 DNA binding motifs were found in CHD4 ChIP-seq peaks; however, research suggests that this transcription factor is expressed in cardiac progenitors but its expression is extinguished in cardiomyocytes (Cai et al. 2003). Two possible explanations exist: 1) CHD4 is recruited to these regulatory targets by ISL1 early in cardiac progenitor stages and maintains its

association at that locus after ISL1 exits, or 2) CHD4 is guided to these regulatory targets by other transcriptional co-factors and the presence of ISL1 motifs at the same locus represent the demonstrated likelihood of finding different cardiac transcription factor motifs in close proximity to one another in genomic regulatory elements (Luna-Zurita et al. 2016). To address the first hypothesis, CHD4 complexes could be isolated in cardiac progenitor cells to identify whether an interaction exists at these stages. We have begun to address the second hypothesis by beginning a computational analysis of CHD4-bound elements to identify all transcription factor motifs present within each ChIP-seq peak. This will address whether other transcription factor motifs may be responsible for recruiting CHD4 to loci that also contain an ISL1 motif, which could indicate other binding partners responsible for CHD4 recruitment at these loci.

Transcriptional repression through CHD4/GATA4 and CHD4/NKX2-5 interactions

Chapter 2 demonstrated that CHD4 directly represses non-cardiac myofibril isoforms in the developing heart; Chapter 3 showed that repression of two of these isoforms in cardiomyocytes may be due to GATA4 and NKX2-5 recruitment of CHD4 and the NuRD complex. Previous research demonstrated that NKX2-5 may be responsible for repression of fast skeletal *Tnni2* and *Tnnt3* (Dupays et al. 2015); however, this study did not show how NKX2-5 was acting as a transcriptional repressor. This research begins to address that important question. An intriguing question raised by the present study is whether human congenital heart disease patients with mutations in *GATA4* and *NKX2-5* also misexpress skeletal and smooth muscle myofibril paralogs, and whether this plays into the pathogenesis of their disease.

Mutations in *CHD4*, *GATA4* and *NKX2-5* result in similar congenital heart defects in some cases, particularly Tetrology of Fallot and atrial and ventricular septal defects (Luo et al. 2014; Baban et al. 2014; Basson et al. 1997; Garg et al. 2003; Hirayama-Yamada et al. 2005; Weiss et al. 2016; Homsy et al. 2015; Sifrim et al. 2016). However, our knowledge of the molecular mechanisms underlying these defects remains underwhelming. It is possible that mutations in these genes disrupt the interaction between these transcription factors and CHD4, akin to our recent findings with *TBX5* Holt-Oram syndrome patient mutations (Waldron et al. 2016). To address this hypothesis, future work will map the regions of each protein required for these interactions. After mapping the congenital heart disease patient mutations that lie within these domains, one could test whether these mutations abolish interactions between CHD4 and GATA4 or CHD4 and NKX2-5. Using the CHD4 ChIP-seq data generated in this study, as well as publicly available GATA4 and NKX2-5 ChIP-seq datasets in differentiated cardiomyocytes (Luna-Zurita et al. 2016) it would be possible to identify genes directly regulated by each of these interactions. This research would address the requirement for these interactions in cardiac differentiation and cardiac development within the context of normal cardiac development and disease.

REFERENCES

Allen, H. F., P. A. Wade, and T. G. Kutateladze. 2013. 'The NuRD architecture', *Cell Mol Life Sci*, 70: 3513-24.

Aramaki, Shinya, Katsuhiko Hayashi, Kazuki Kurimoto, Hiroshi Ohta, Yukihiro Yabuta, Hiroko Iwanari, Yasuhiro Mochizuki, Takao Hamakubo, Yuki Kato, Katsuhiko Shirahige, and Mitinori Saitou. 2013. 'A Mesodermal Factor, T, Specifies Mouse Germ Cell Fate by Directly Activating Germline Determinants', *Dev Cell*, 27: 516-29.

Baban, A., A. V. Postma, M. Marini, G. Trocchio, A. Santilli, M. Pelegrini, P. Sirleto, M. Lerone, S. B. Albanese, P. Barnett, C. J. Boogerd, B. Dallapiccola, M. C. Digilio, R. Ravazzolo, and G. Pongiglione. 2014. 'Identification of TBX5 mutations in a series of 94 patients with Tetralogy of Fallot', *Am J Med Genet A*, 164a: 3100-7.

Basson, C. T., D. R. Bachinsky, R. C. Lin, T. Levi, J. A. Elkins, J. Soults, D. Grayzel, E. Kroumpouzou, T. A. Traill, J. Leblanc-Straceski, B. Renault, R. Kucherlapati, J. G. Seidman, and C. E. Seidman. 1997. 'Mutations in human TBX5 [corrected] cause limb and cardiac malformation in Holt-Oram syndrome', *Nat Genet*, 15: 30-5.

Beyer, Tobias A, Alexander Weiss, Yuliya Khomchuk, Kui Huang, Abiodun A Ogunjimi, Xaralabos Varelas, and Jeffrey L Wrana. 2013. 'Switch Enhancers Interpret TGF-β and Hippo Signaling to Control Cell Fate in Human Embryonic Stem Cells', *Cell Reports*, 5: 1611-24.

Cai, Chen-Leng, Xingqun Liang, Yunqing Shi, Po-Hsien Chu, Samuel L. Pfaff, Ju Chen, and Sylvia Evans. 2003. 'Isl1 Identifies a Cardiac Progenitor Population that Proliferates Prior to Differentiation and Contributes a Majority of Cells to the Heart', *Dev Cell*, 5: 877-89.

Curtis, C. D., and C. T. Griffin. 2012. 'The chromatin-remodeling enzymes BRG1 and CHD4 antagonistically regulate vascular Wnt signaling', *Mol Cell Biol*, 32: 1312-20.

Dai, Y. S., and B. E. Markham. 2001. 'p300 Functions as a coactivator of transcription factor GATA-4', *J Biol Chem*, 276: 37178-85.

David, Robert, Veronica Barbara Jarsch, Florian Schwarz, Petra Nathan, Moritz Gegg, Heiko Lickert, and Wolfgang-Michael Franz. 2011. 'Induction of MesP1 by Brachyury(T) generates the common multipotent cardiovascular stem cell', *Cardiovasc Res*, 92: 115-22.

Dupays, Laurent, Catherine Shang, Robert Wilson, Surendra Kotecha, Sophie Wood, Norma Towers, and Timothy Mohun. 2015. 'Sequential Binding of MEIS1 and NKX2-5 on the Popdc2 Gene: A Mechanism for Spatiotemporal Regulation of Enhancers during Cardiogenesis', *Cell Reports*, 13: 183-95.

Edmondson, D. G., G. E. Lyons, J. F. Martin, and E. N. Olson. 1994. 'Mef2 gene expression marks the cardiac and skeletal muscle lineages during mouse embryogenesis', *Development*, 120: 1251-63.

Ehler, E., B. M. Rothen, S. P. Hammerle, M. Komiyama, and J. C. Perriard. 1999. 'Myofibrillogenesis in the developing chicken heart: assembly of Z-disk, M-line and the thick filaments', *J Cell Sci*, 112 (Pt 10): 1529-39.

Fentzke, R. C., C. E. Korcarz, R. M. Lang, H. Lin, and J. M. Leiden. 1998. 'Dilated cardiomyopathy in transgenic mice expressing a dominant-negative CREB transcription factor in the heart', *J Clin Invest*, 101: 2415-26.

Garg, V., I. S. Kathiriya, R. Barnes, M. K. Schluterman, I. N. King, C. A. Butler, C. R. Rothrock, R. S. Eapen, K. Hirayama-Yamada, K. Joo, R. Matsuoka, J. C. Cohen, and D. Srivastava. 2003. 'GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5', *Nature*, 424: 443-7.

Gerull, B., M. Gramlich, J. Atherton, M. McNabb, K. Trombitas, and S. Sasse-Klaassen. 2002. 'Mutations of TTN, encoding the giant muscle filament titin, cause familial dilated cardiomyopathy', *Nat Genet*, 30.

Golebiewska, Urszula, and Suzanne Scarlata. 2011. 'Measuring Fast Calcium Fluxes in Cardiomyocytes', *Journal of Visualized Experiments : JoVE*: 3505.

Gómez-del Arco, Pablo, Eusebio Perdiguero, Paula Sofia Yunes-Leites, Rebeca Acín-Pérez, Miriam Zeini, Antonio Garcia-Gomez, Krishnamoorthy Sreenivasan, Miguel Jiménez-Alcázar, Jessica Segalés, Dolores López-Maderuelo, Beatriz Ornés, Luis Jesús Jiménez-Borreguero, Gaetano D'Amato, David Enshell-Seijffers, Bruce Morgan, Katia Georgopoulos, Abul B M. M. K. Islam, Thomas Braun, José Luis de la Pompa, Johnny Kim, José A Enriquez, Esteban Ballestar, Pura Muñoz-Cánoves, and Juan Miguel Redondo. 'The Chromatin Remodeling Complex Chd4/NuRD Controls Striated Muscle Identity and Metabolic Homeostasis', *Cell Metabolism*, 23: 881-92.

Guatimosim, Silvia, Cristina Guatimosim, and Long-Sheng Song. 2011. 'Imaging Calcium Sparks in Cardiac Myocytes', *Methods Mol Biol*, 689: 205-14.

Hang, Calvin T., Jin Yang, Pei Han, Hsiu-Ling Cheng, Ching Shang, Euan Ashley, Bin Zhou, and Ching-Pin Chang. 2010. 'Chromatin regulation by Brg1 underlies heart muscle development and disease', *Nature*, 466: 62-67.

Heidersbach, Amy, Chris Saxby, Karen Carver-Moore, Yu Huang, Yen-Sin Ang, Pieter J. de Jong, Kathryn N. Ivey, and Deepak Srivastava. 2013. 'microRNA-1 regulates sarcomere formation and suppresses smooth muscle gene expression in the mammalian heart', *eLife*, 2: e01323.

Hirayama-Yamada, K., M. Kamisago, K. Akimoto, H. Aotsuka, Y. Nakamura, H. Tomita, M. Furutani, S. Imamura, A. Takao, M. Nakazawa, and R. Matsuoka. 2005. 'Phenotypes

with GATA4 or NKX2.5 mutations in familial atrial septal defect', *Am J Med Genet A*, 135: 47-52.

Hirschy, A., F. Schatzmann, E. Ehler, and J. C. Perriard. 2006. 'Establishment of cardiac cytoarchitecture in the developing mouse heart', *Dev Biol*, 289: 430-41.

Homsy, J., S. Zaidi, Y. Shen, J. S. Ware, K. E. Samocha, K. J. Karczewski, S. R. DePalma, D. McKean, H. Wakimoto, J. Gorham, S. C. Jin, J. Deanfield, A. Giardini, G. A. Porter, Jr., R. Kim, K. Bilguvar, F. Lopez-Giraldez, I. Tikhonova, S. Mane, A. Romano-Adesman, H. Qi, B. Vardarajan, L. Ma, M. Daly, A. E. Roberts, M. W. Russell, S. Mital, J. W. Newburger, J. W. Gaynor, R. E. Breitbart, I. Iossifov, M. Ronemus, S. J. Sanders, J. R. Kaltman, J. G. Seidman, M. Brueckner, B. D. Gelb, E. Goldmuntz, R. P. Lifton, C. E. Seidman, and W. K. Chung. 2015. 'De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies', *Science*, 350: 1262-6.

Hu, Guang, and Paul A Wade. 2012. 'NuRD and Pluripotency: A Complex Balancing Act', *Cell Stem Cell*, 10: 497-503.

Hung, H., R. Kohnken, and J. Svaren. 2012. 'The nucleosome remodeling and deacetylase chromatin remodeling (NuRD) complex is required for peripheral nerve myelination', *J Neurosci*, 32: 1517-27.

Ieda, M., J. D. Fu, P. Delgado-Olguin, V. Vedantham, Y. Hayashi, B. G. Bruneau, and D. Srivastava. 2010. 'Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors', *Cell*, 142: 375-86.

Ingram, Kyle G., Carol D. Curtis, Robert Silasi-Mansat, Florea Lupu, and Courtney T. Griffin. 2013. 'The NuRD Chromatin-Remodeling Enzyme CHD4 Promotes Embryonic Vascular Integrity by Transcriptionally Regulating Extracellular Matrix Proteolysis', *PLoS Genet*, 9: e1004031.

Johansson, B M, and M V Wiles. 1995. 'Evidence for involvement of activin A and bone morphogenetic protein 4 in mammalian mesoderm and hematopoietic development', *Mol Cell Biol*, 15: 141-51.

Kamisago , Mitsuhiro, Sapna D. Sharma , Steven R. DePalma , Scott Solomon , Pankaj Sharma , Barbara McDonough , Leslie Smoot , Mary P. Mullen , Paul K. Woolf , E. Douglas Wigle , J.G. Seidman , John Jarcho , Lawrence R. Shapiro , and Christine E. Seidman 2000. 'Mutations in Sarcomere Protein Genes as a Cause of Dilated Cardiomyopathy', *New England Journal of Medicine*, 343: 1688-96.

Kaski, J P, P Syrris, M Burch, M-T Tomé-Esteban, M Fenton, M Christiansen, P S Andersen, N Sebire, M Ashworth, J E Deanfield, W J McKenna, and P M Elliott. 2008. 'Idiopathic restrictive cardiomyopathy in children is caused by mutations in cardiac sarcomere protein genes', *Heart*, 94: 1478-84.

Klaassen, S., S. Probst, E. Oechslin, B. Gerull, G. Krings, P. Schuler, M. Greutmann, D. Hurlimann, M. Yegitbasi, L. Pons, M. Gramlich, J. D. Drenckhahn, A. Heuser, F. Berger, R. Jenni, and L. Thierfelder. 2008. 'Mutations in sarcomere protein genes in left ventricular noncompaction', *Circulation*, 117: 2893-901.

Lien, C.L., C. Wu, B. Mercer, R. Webb, J.A. Richardson, and E.N. Olson. 1999. 'Control of early cardiac-specific transcription of Nkx2-5 by a GATA-dependent enhancer', *Development*, 126: 75-84.

Lin, Q., J. Schwarz, C. Bucana, and E. N. Olson. 1997. 'Control of mouse cardiac morphogenesis and myogenesis by transcription factor MEF2C', *Science*, 276: 1404-7.

Linder, B., E. Mentele, K. Mansperger, T. Straub, E. Kremmer, and R. A. Rupp. 2007. 'CHD4/Mi-2beta activity is required for the positioning of the mesoderm/neuroectoderm boundary in Xenopus', *Genes Dev*, 21: 973-83.

Luna-Zurita, L., C. U. Stirnimann, S. Glatt, B. L. Kaynak, S. Thomas, F. Baudin, M. A. Samee, D. He, E. M. Small, M. Mileikovsky, A. Nagy, A. K. Holloway, K. S. Pollard, C. W. Muller, and B. G. Bruneau. 2016. 'Complex Interdependence Regulates Heterotypic Transcription Factor Distribution and Coordinates Cardiogenesis', *Cell*.

Luo, Z. L., H. Sun, Z. Q. Yang, Y. H. Ma, Y. Gu, Y. Q. He, D. Wei, L. B. Xia, B. H. Yang, and T. Guo. 2014. 'Genetic variations of ISL1 associated with human congenital heart disease in Chinese Han people', *Genet Mol Res*, 13: 1329-38.

Mahmoud, A. I., F. Kocabas, S. A. Muralidhar, W. Kimura, A. S. Koura, S. Thet, E. R. Porrello, and H. A. Sadek. 2013. 'Meis1 regulates postnatal cardiomyocyte cell cycle arrest', *Nature*, 497: 249-53.

Montgomery, Rusty L., Christopher A. Davis, Matthew J. Potthoff, Michael Haberland, Jens Fielitz, Xiaoxia Qi, Joseph A. Hill, James A. Richardson, and Eric N. Olson. 2007. 'Histone deacetylases 1 and 2 redundantly regulate cardiac morphogenesis, growth, and contractility', *Genes Dev*, 21: 1790-802.

Morita , Hiroyuki, Heidi L. Rehm , Andres Menesses , Barbara McDonough , Amy E. Roberts , Raju Kucherlapati , Jeffrey A. Towbin , J.G. Seidman , and Christine E. Seidman 2008. 'Shared Genetic Causes of Cardiac Hypertrophy in Children and Adults', *New England Journal of Medicine*, 358: 1899-908.

Moses, K. A., F. DeMayo, R. M. Braun, J. L. Reecy, and R. J. Schwartz. 2001. 'Embryonic expression of an Nkx2-5/Cre gene using ROSA26 reporter mice', *Genesis*, 31: 176-80.

Musselman, C. A., J. Ramirez, J. K. Sims, R. E. Mansfield, S. S. Oliver, J. M. Denu, J. P. Mackay, P. A. Wade, J. Hagman, and T. G. Kutateladze. 2012. 'Bivalent recognition of nucleosomes by the tandem PHD fingers of the CHD4 ATPase is required for CHD4-mediated repression', *Proc Natl Acad Sci U S A*, 109: 787-92.

Nishii, K., S. Morimoto, R. Minakami, Y. Miyano, K. Hashizume, M. Ohta, D. Y. Zhan, Q. W. Lu, and Y. Shibata. 2008. 'Targeted disruption of the cardiac troponin T gene causes sarcomere disassembly and defects in heartbeat within the early mouse embryo', *Dev Biol*, 322: 65-73.

O'Shaughnessy-Kirwan, A., J. Signolet, I. Costello, S. Gharbi, and B. Hendrich. 2015. 'Constraint of gene expression by the chromatin remodelling protein CHD4 facilitates lineage specification', *Development*, 142: 2586-97.

Paige, Sharon L, Sean Thomas, Cristi L Stoick-Cooper, Hao Wang, Lisa Maves, Richard Sandstrom, Lil Pabon, Hans Reinecke, Gabriel Pratt, Gordon Keller, Randall T Moon, John Stamatoyannopoulos, and Charles E Murry. 2012. 'A Temporal Chromatin Signature in Human Embryonic Stem Cells Identifies Regulators of Cardiac Development', *Cell*, 151: 221-32.

Petchey, L. K., C. A. Risebro, J. M. Vieira, T. Roberts, J. B. Bryson, L. Greensmith, M. F. Lythgoe, and P. R. Riley. 2014. 'Loss of Prox1 in striated muscle causes slow to fast skeletal muscle fiber conversion and dilated cardiomyopathy', *Proc Natl Acad Sci U S A*, 111: 9515-20.

Pfefferli, Catherine, Fritz Muller, Anna Jazwinska, and Chantal Wicky. 2014. 'Specific NuRD components are required for fin regeneration in zebrafish', *BMC Biol*, 12: 30.

Qian, Li, Yu Huang, C. Ian Spencer, Amy Foley, Vasanth Vedantham, Lei Liu, Simon J. Conway, Ji-dong Fu, and Deepak Srivastava. 2012. 'In vivo reprogramming of murine cardiac fibroblasts into induced cardiomyocytes', *Nature*, 485: 593-98.

Risebro, C. A., R. G. Searles, A. A. Melville, E. Ehler, N. Jina, S. Shah, J. Pallas, M. Hubank, M. Dillard, N. L. Harvey, R. J. Schwartz, K. R. Chien, G. Oliver, and P. R. Riley. 2009. 'Prox1 maintains muscle structure and growth in the developing heart', *Development*, 136: 495-505.

Saga, Y., S. Miyagawa-Tomita, A. Takagi, S. Kitajima, Ji Miyazaki, and T. Inoue. 1999. 'MesP1 is expressed in the heart precursor cells and required for the formation of a single heart tube', *Development*, 126: 3437-47.

Sehnert, A. J., A. Huq, B. M. Weinstein, C. Walker, M. Fishman, and D. Y. Stainier. 2002. 'Cardiac troponin T is essential in sarcomere assembly and cardiac contractility', *Nat Genet*, 31: 106-10.

Shimbo, T., Y. Du, S. A. Grimm, A. Dhasarathy, D. Mav, R. R. Shah, H. Shi, and P. A. Wade. 2013. 'MBD3 localizes at promoters, gene bodies and enhancers of active genes', *PLoS Genet*, 9: e1004028.

Shimono, Y., H. Murakami, K. Kawai, P. A. Wade, K. Shimokata, and M. Takahashi. 2003. 'Mi-2 beta associates with BRG1 and RET finger protein at the distinct regions with transcriptional activating and repressing abilities', *J Biol Chem*, 278: 51638-45.

Sifrim, A., M. P. Hitz, A. Wilsdon, J. Breckpot, S. H. Turki, B. Thienpont, J. McRae, T. W. Fitzgerald, T. Singh, G. J. Swaminathan, E. Prigmore, D. Rajan, H. Abdul-Khaliq, S. Banka, U. M. Bauer, J. Bentham, F. Berger, S. Bhattacharya, F. Bu'Lock, N. Canham, I. G. Colgiu, C. Cosgrove, H. Cox, I. Daehnert, A. Daly, J. Danesh, A. Fryer, M. Gewillig, E. Hobson, K. Hoff, T. Homfray, A. K. Kahlert, A. Ketley, H. H. Kramer, K. Lachlan, A. K. Lampe, J. J. Louw, A. K. Manickara, D. Manase, K. P. McCarthy, K. Metcalfe, C. Moore, R. Newbury-Ecob, S. O. Omer, W. H. Ouwehand, S. M. Park, M. J. Parker, T. Pickardt, M. O. Pollard, L. Robert, D. J. Roberts, J. Sambrook, K. Setchfield, B. Stiller, C. Thornborough, O. Toka, H. Watkins, D. Williams, M. Wright, S. Mital, P. E. Daubeney, B. Keavney, J. Goodship, R. M. Abu-Sulaiman, S. Klaassen, C. F. Wright, H. V. Firth, J. C. Barrett, K. Devriendt, D. R. FitzPatrick, J. D. Brook, and M. E. Hurles. 2016. 'Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing', *Nat Genet*, 48: 1060-5.

Suurmeijer, A. J., S. Clement, A. Francesconi, L. Bocchi, A. Angelini, D. J. Van Veldhuisen, L. G. Spagnoli, G. Gabbiani, and A. Orlandi. 2003. 'Alpha-actin isoform distribution in normal and failing human heart: a morphological, morphometric, and biochemical study', *J Pathol*, 199: 387-97.

Svensson, Eric C., Rachel L. Tufts, Christine E. Polk, and Jeffrey M. Leiden. 1999. 'Molecular cloning of FOG-2: A modulator of transcription factor GATA-4 in cardiomyocytes', *Proceedings of the National Academy of Sciences*, 96: 956-61.

Takimoto, Eiki, and David A. Kass. 2012. 'Chapter 21 - Regulation of Cardiac Systolic Function and Contractility A2 - Hill, Joseph A.' in Eric N. Olson (ed.), *Muscle* (Academic Press: Boston/Waltham).

van Eldik, W., and R. Passier. 2013. 'Signalling in sarcomeres in development and disease', *Netherlands Heart Journal*, 21: 367-71.

Waldron, L., J. D. Steimle, T. M. Greco, N. C. Gomez, K. M. Dorr, J. Kweon, B. Temple, X. H. Yang, C. M. Wilczewski, I. J. Davis, I. M. Cristea, I. P. Moskowitz, and F. L. Conlon. 2016. 'The Cardiac TBX5 Interactome Reveals a Chromatin Remodeling Network Essential for Cardiac Septation', *Dev Cell*, 36: 262-75.

Wamstad, Joseph A, Jeffrey M Alexander, Rebecca M Truty, Avanti Shrikumar, Fugen Li, Kirsten E Eilertson, Huiming Ding, John N Wylie, Alexander R Pico, John A Capra, Genevieve Erwin, Steven J Kattman, Gordon M Keller, Deepak Srivastava, Stuart S Levine, Katherine S Pollard, Alisha K Holloway, Laurie A Boyer, and Benoit G Bruneau. 2012. 'Dynamic and Coordinated Epigenetic Regulation of Developmental Transitions in the Cardiac Lineage', *Cell*, 151: 206-20.

Weiss, K., P. A. Terhal, L. Cohen, M. Bruccoleri, M. Irving, A. F. Martinez, J. A. Rosenfeld, K. Machol, Y. Yang, P. Liu, M. Walkiewicz, J. Beuten, N. Gomez-Ospina, K. Haude, C. T. Fong, G. M. Enns, J. A. Bernstein, J. Fan, G. Gotway, M. Ghorbani, K. van Gassen, G. R. Monroe, G. van Haaften, L. Basel-Vanagaite, X. J. Yang, P. M. Campeau, and M. Muenke. 2016. 'De Novo Mutations in CHD4, an ATP-Dependent

Chromatin Remodeler Gene, Cause an Intellectual Disability Syndrome with Distinctive Dysmorphisms', *Am J Hum Genet*.

Wu, S. M., Y. Fujiwara, S. M. Cibulsky, D. E. Clapham, C. L. Lien, T. M. Schultheiss, and S. H. Orkin. 2006. 'Developmental origin of a bipotential myocardial and smooth muscle cell precursor in the mammalian heart', *Cell*, 127: 1137-50.

Xin, M., Y. Kim, L. B. Sutherland, X. Qi, J. McAnally, R. J. Schwartz, J. A. Richardson, R. Bassel-Duby, and E. N. Olson. 2011. 'Regulation of insulin-like growth factor signaling by Yap governs cardiomyocyte proliferation and embryonic heart size', *Sci Signal*, 4: ra70.

Yamada, Tomoko, Yue Yang, Martin Hemberg, Toshimi Yoshida, Ha Young Cho, J. Patrick Murphy, Diasynou Fioravante, Wade G Regehr, Steven P Gygi, Katia Georgopoulos, and Azad Bonni. 2014. 'Promoter Decommissioning by the NuRD Chromatin Remodeling Complex Triggers Synaptic Connectivity in the Mammalian Brain', *Neuron*, 83: 122-34.

APPENDIX 1: The cardiac TBX5 interactome reveals a chromatin remodeling network essential for cardiac septation[2]

Introduction

Congenital malformations, or structural birth defects, are the leading cause of infant mortality in the US and Europe (Dolk et al., 2010; Heron and Tejada-Vera, 2009). Congenital heart disease (CHD) remains the most common congenital malformation, with atrial septal defects (ASD) present in 116 per 100,000 live births and ventricular septal defects (VSD) present in 307 per 100,000 live births (Gittenberger-de Groot et al., 2014). Holt Oram Syndrome (HOS) is an autosomal disorder associated with cardiac septal defects and is caused by dominant mutations in the T-box transcription factor *Tbx5* (Basson et al., 1997; Basson et al., 1994; Basson et al., 1999; Cross et al., 2000; Holt and Oram, 1960; Li et al., 1997). Mice heterozygous for a null mutation in *Tbx5* display many of the phenotypic abnormalities of CHD, while mice homozygous for a null mutation in *Tbx5* die by E10.5 due to cardiac defects (Bruneau et al., 1999; Bruneau et al., 2001; Moskowitz et al., 2007; Moskowitz et al., 2004). In addition to its role in cardiogenesis, *Tbx5* is one of the key factors essential for cellular reprogramming of fibroblasts into induced cardiomyocytes (Ieda et al., 2010; Qian et al., 2012).

While *Tbx5* is an essential transcription factor for heart development and reprogramming, and its disease relevance is well established, there are many critical

[2]This appendix previously appeared as an article in the journal *Developmental Cell*. The original citation is as follows: Waldron L, Steimle JD, Greco TM, Gomez NC, Dorr KM, Kweon J, Temple B, Yang XH, Wilczewski CM, Davis IJ, Cristea IM, Moskowitz IP, Conlon FL. “The cardiac TBX5 interactome reveals a chromatin remodeling network essential for cardiac septation,” *Developmental Cell* Volume 36 (February 2016): 3.

unanswered questions about the mechanism of TBX5 function. In particular, the TBX5 interactome has not been defined in-vivo, leaving unanswered the question of how TBX5 interacting proteins regulate TBX5's choice of distinct transcriptional targets, or how these interactions function to activate and/or repress target gene transcription. It is also not known how these events, when disrupted, contribute to CHD. To these ends, we initiated a directed proteomic-based approach to identify endogenous cardiac proteins that function in physical association with TBX5.

We report the discovery that TBX5 interacts biochemically and genetically with the transcriptional repression machinery of the Nucleosome Remodeling and Deacetylase (NuRD) complex to repress inappropriate expression of gene programs in the developing heart. We have further uncovered a novel domain of TBX5, show that it forms an α-Helix, and that TBX5 mis-sense mutations associated with septal defects in human patients map to this domain. Furthermore, HOS patient mis-sense mutations in this domain disrupt the TBX5-NuRD interaction and elicit misexpression of inappropriate gene programs that are normally repressed by TBX5. Through phylogenetic analysis across numerous vertebrate species, we find that the NuRD-interaction domain of TBX5, and hence the interaction of TBX5 with NuRD, evolved during the early diversification of vertebrates, simultaneous with the evolution of cardiac septation. Thus, the TBX5-NuRD interaction identified here plays a central role in both cardiac development and human congenital heart defects, as well as the evolution of the mammalian heart.

Results

Generation of the Tbx5^{Avi} allele

Past attempts to isolate endogenous T-box protein complexes have been hampered by their relative low abundance and insolubility. Therefore, we introduced an Avitag epitope into the murine *Tbx5* locus via homologous recombination to generate a *Tbx5^{Avi}* allele (Figure A1.1A-D). The Avitag is a small peptide tag that is biotinylated in the presence of the bacterial enzyme BirA. Thus, the Avitag/BirA system combines minimal structural invasiveness with the specificity and strength of the biotin-streptavidin interaction, the strongest non-covalent peptide-ligand interaction identified to date (Bayer and Wilchek, 1980; Maine et al., 2010; Roesli et al., 2006; van Werven and Timmers, 2006; Wang et al., 2006); *Tbx5^{Avi}* mice were mated to mice ubiquitously expressing BirA (*Rosa26^{BirA}*) (Driegen et al., 2005). The *Tbx5^{Avi}*; *Rosa26^{BirA}* compound homozygous mice show no overt phenotype, have no cardiac phenotype, and are fertile. Thus, the Avitag appears to have no effect on TBX5 activity, and expression of biotinylated TBX5^{Avi} is not deleterious to embryonic development.

Isolation and characterization of the endogenous TBX5 interactome

To define the TBX5 cardiac interactome, we performed mass spectrometry analysis of affinity-purified TBX5 complexes from *Tbx5$^{Avi/Avi}$;Rosa26$^{BirA/BirA}$* cardiac nuclei (Figure A1.1E). We used an unbiased gene ontology-based bioinformatics classification to screen the functions of proteins associated with TBX5 at a statistical relevance versus *Tbx5$^{Avi/Avi}$* alone (no BirA) and *Rosa26$^{BirA/BirA}$* alone. This analysis defined a subset of 58 candidate interactions (Table SA1.1). We analyzed these candidates by

functional network analysis; drawing upon the STRING database of protein-protein interactions, 40 out of the 58 candidates were assigned to a single interconnected network (Figure A1.1F, SA1.1A).

TBX5 interacts with components of the Nucleosome Remodeling and Deacetylase (NuRD) complex in vivo

Analysis of the TBX5 interaction network identified the Heterodimeric Facilitates Chromatin Transcription (FACT) complex and several members of the transcription initiation factor TFIID complex, both of which play roles in RNA Pol II-dependent gene transcription (Berk, 1999; Saunders et al., 2003) (Figure A1.1F). TBX5 has been long considered to be a cardiac transcriptional activator. Surprisingly, in our analysis we identified TBX5 in association with multiple components of the NuRD chromatin-modifying complex (Figure A1.1F, G), which in most instances functions to repress gene transcription (Wade et al., 1998; Xue et al., 1998). We find TBX5 in association with six components of the NuRD transcriptional repression complex: CHD4, MTA1, RBBP4, GATAD2a, GATAD2b, and HDAC2. All the components required for a functional NuRD complex were identified in association with TBX5, as CHD3 commonly acts in a mutually exclusive manner with CHD4, RBBP7 and RBBP4 act redundantly, and an additional component of the NuRD complex, MBD3, has been suggested to be dispensable for transcription factor recruitment (e.g., by hunchback (Kehle et al., 1998)). We confirmed the interaction of TBX5 with multiple members of the NuRD complex in adult cardiac nuclear extracts (Figure A1.1G).

As TBX5 has been shown to be required for cardiogenesis at E9.5 (Bruneau et al., 2001), we established expression of components of the NuRD complex in cardiac tissue at E9.5. We found that members of the NuRD complex are strongly expressed in the heart at this stage by RT-PCR (Figure SA1.1B). We further demonstrate that endogenous TBX5 (untagged TBX5) interacts with CHD4 at E9.5 (Figure A1.1H).

TBX5 and MTA1 and cardiac septation

To assess whether the TBX5-NuRD complex interaction was physiologically relevant, we analyzed mutant alleles of *Tbx5* and a central component of the NuRD complex, *Mta1*, for a genetic interaction (Figure A1.2, SA1.2). At E13.5, *Mta1*$^{+/-}$ mice have no apparent cardiac defects (Figure A1.2C). *Tbx5*$^{+/-}$ mice exhibit partially penetrant cardiac defects as previously described, including ASD, VSD, and complete common atrioventricular canal (CCAVC), which includes both ASD and VSD components (Figure A1.2D) (Baban et al., 2014; Basson et al., 1994; Basson et al., 1999; Benson et al., 1996; Bruneau et al., 1999; Bruneau et al., 2001; McDermott et al., 2005). *Tbx5*$^{+/-}$; *Mta1*$^{+/-}$ heterozygous null embryos all demonstrated septal defects (19/19), a higher frequency of septal defects than either single heterozygote alone (0/14 for *Mta1*$^{+/-}$ and 12/16 for *Tbx5*$^{+/-}$) ($p = 7.1E-9$ vs. WT and $p = 0.035$ vs *Tbx5*$^{+/-}$) (Figure A1.2E-J). Furthermore, *Mta1*$^{-/-}$ hearts also do not have septal defects at this stage (Figure SA1.3). We observed no change in the left ventricular chamber width or left ventricular wall thickness (Figure SA1.2D-E) or inferior or superior cushions size (Figure A1.2F-I). To rule out a developmental delay in septation in the *Tbx5*$^{+/-}$; *Mta1*$^{+/-}$ heterozygous null embryos, we examined *Tbx5*$^{+/-}$; *Mta1*$^{+/-}$ heterozygous null embryos at E14.5 and

consistently observe CCAVCs at this time point (Figure SA1.2F-G). These results demonstrate a genetic interaction between *Tbx5* and *Mta1* in intro-cardiac septation providing genetic evidence for a role for the TBX5-NuRD interaction.

.

TBX5 binds to the same consensus DNA site in activated and repressed target genes

The finding that TBX5 and NuRD complex components interact biochemically and genetically led us to hypothesize that TBX5 can function to repress aberrant gene expression in the developing heart. To test this hypothesis, we performed RNA-seq analysis and identified genes significantly downregulated and significantly upregulated in *Tbx5* null verses wildtype heart tissue (Figure A1.3, Table SA1.2). To determine which genes were direct targets of TBX5, we overlaid our data set to that of a TBX5 ChIP-seq data set (He et al., 2011). From our ChIP-seq data, 71,624 TBX5 peak regions were identified by enrichment of signal using MACS2 (Zhang et al., 2008). These peak regions demonstrated a significant association with genes that were differentially regulated between *Tbx5* wild type and *Tbx5*$^{-/-}$ hearts.

Given a large number of peaks can skew results in gene-association analyses, we varied the number of analyzed peaks from 10,000 to 70,000 in increments of 10,000. In every instance, TBX5 was significantly associated with differentially expressed genes (data not shown). We selected the top 40,000 peaks for subsequent analyses since the inclusion of additional peaks minimally increased the number of associated genes (data not shown). We identified an association between TBX5 and 82% of genes differentially expressed between *Tbx5* wild type and *Tbx5*$^{-/-}$ hearts (2988 of 3634, Figure A1.3A). Of these genes, about half were activated and half were repressed by TBX5. Both groups

demonstrated a statistically significant enrichment in TBX5 binding sites in the TBX5 ChIP-peak regions (activated 4.88×10^{-141}, repressed $p = 1.29 \times 10^{-142}$, KS test, Figure A1.3B). Genes repressed by TBX5 were enriched in pathways involved in cancer as well as cardiac diseases, while genes activated by TBX5 were enriched in pathways involved in cell cycle and DNA replication, a known function of TBX5 (Figure A1.3C) (Hatcher et al., 2001, Goetz et al., 2006).

TBX5 acts to directly repress inappropriate expression of genes in cardiac tissue

To validate the genes putatively repressed by TBX5, we rank ordered our list of repressed genes based first on relative change in expression between wild-type and *Tbx5*$^{-/-}$ hearts and the number of reads from the ChIP-seq data set (Table SA1.3). From this rank ordered list, we cloned the top 15 of the ChIP peak sequence elements from 14 genes and performed transcriptional assays (Figure A1.4, SA1.4). Of the candidate genes tested, 11 were repressed by TBX5. This set of genes could be divided into two categories: *Casz1*, *Fgf11*, *Gad1, Kcne3, Kctd16, Klf4, Nxph4, Plekha2,* and *Sncb* are expressed in neural lineages (Diez-Roux et al., 2011; Dorr et al., 2015; Liu et al., 2011; Metz et al., 2011; Smallwood et al., 1996; Sopher et al., 2001; Trifonov et al., 2014; Wang et al., 2014), while *Casz1*, *Col20a1, Fgf11*, *Gad1*, *Klf4, Mef2b,* and *Plekha2* are expressed in a divergent subset of cancers (Costantini et al., 2009; Evans and Liu, 2008; Hu et al., 2015; Huang et al., 2013; Kimura et al., 2013; Ying et al., 2013). Altogether, these data suggest that TBX5 function is essential for *in vivo* transcriptional repression and prevention of ectopic expression of inappropriate gene programs in the heart.

TBX5 interaction with CHD4 does not require either the CHD4 chromo domains or the Phd domains.

CHD4 is a highly conserved protein that serves as the central catalytic core component of the NuRD complex (Denslow and Wade, 2007). Studies across plants and animals have shown that CHD4 rather functions through interaction with DNA binding proteins by its pair of chromo domains. In addition CHD4 also contains a tandem set of Phd (plant homology domains) and an ATPase cassette (Denslow and Wade, 2007). To gain insight into the mechanisms of TBX5-NuRD mediated repression, we mapped the protein interaction domains of TBX5 of CHD4 (Figure A1.5A-F). Streptavidin pulldown of TBX5 complexes demonstrated that an N-terminal human CHD4 region containing the pair of Phd domains and the pair of chromo domains was required for interaction with mouse TBX5 (Figure A1.5A, C). Surprisingly, we found CHD4 deletions of either the Phd domains or the chromo domains still interact with TBX5(Figure A1.5D) implying that TBX5 interacts with a previously undefined set of residues in the amino terminal portion of CHD4.

TBX5 interacts with the NuRD complex through a coil-α-helix domain

In vitro based approaches have led to the identification of a number of proteins that interact with TBX5 through two regions of TBX5; one in the amino-terminal portion of the protein including the most amino-terminal portion of the T-box DNA binding domain (Brown et al., 2005; Garg et al., 2003; Ghosh et al., 2009; Hiroi et al., 2001; Wang et al., 2011) and a second in the C-terminus of TBX5 containing a putative

nuclear localization signal (Krause et al., 2004; Murakami et al., 2005). To determine if TBX5 associates with the NuRD complex through these regions we mapped the residues of TBX5 that was essential for interaction with CHD4. From these studies we found a region C-terminal to the T-box was necessary and sufficient for interaction with CHD4 (amino acids 238-340) (Figure A1.5B, E-F). A conserved function for this region is supported by our observation that the minimal region of TBX5 that interacts with CHD4 is highly conserved across 48 TBX5 orthologues (Figure A1.5H).

Since the region of TBX5 that interacts with CHD4 has not been included in TBX5 crystal structural analysis (Stirnimann et al., 2010) we sought to determine if the region of TBX5 that interacted with CHD4 represented a true TBX5 structural domain. We therefore conducted molecular 3D structural modeling of the entire TBX5 protein as well as the minimal region of TBX5 that we showed interacted with CHD4. Results from these analyses revealed that TBX5 residues 255-264 is highly predicted to form a small coil-to α-Helix region (Figure A1.5I,J). We have termed this structural-functional region of TBX5 the NuRD Interaction Domain ($TBX5^{NID}$) (Figure A1.5G).

The TBX5-NuRD interaction domain is essential for cardiac development and function in humans

Having demonstrated TBX5 and the NuRD complex genetically and biochemically interact, and having established that $TBX5^{NID}$ represents a true structural functional domain, we next determined if the interaction between TBX5 and the NuRD complex is essential for TBX5 function. For these studies we undertook a CHD patient driven approach based on the observations that the vast majority of molecularly

characterized TBX5 mutations associated with CHD are mis-sense mutations within the coding region of TBX5. Most of the CHD associated mutations are located within the TBX5 T-box causing either a loss of DNA binding or degradation by non-sense mediated decay (Fan et al., 2003a; Fan et al., 2003b; Mori and Bruneau, 2004). However in addition, there exists a subset of CHD-associated mutations that lie in the C-terminus of TBX5, outside the T-box; the resulting mutant forms of TBX5 are stable, nuclear localized and bind DNA. The molecular basis for these CHD-associated mutations however, has been unknown.

Multiple mis-sense mutations that correspond to TBX5-associated CHD in the C-terminal region of TBX5 that fall within the $TBX5^{NID}$ including: $TBX5^{S252I}$, $TBX5^{S261C}$, $TBX5^{V263M}$, $TBX5^{K266R}$, and $TBX5^{Q292R}$ (Heinritz et al., 2005) (Figure A1.5G). We investigated the physical location of these TBX5 amino acids on the predicted NID structural domain and observed the amino acids align along a single interaction surface of the $TBX5^{NID}$ α-Helix (Figure A1.5I-J). We hypothesized the CHD mis-sense mutations in the $TBX5^{NID}$ disrupt the TBX5-CHD4 interaction.

We tested the hypothesis that TBX5 mis-sense mutations in the $TBX5^{NID}$ alter the TBX5-NuRD association. Three single mis-sense mutations within the $TBX5^{NID}$ ($TBX5^{S261C}$, $TBX5^{V263M}$, and $TBX5^{K266R}$) each abolished the interaction between TBX5 and both CHD4 and MTA1, and an additional two other CHD single mis-sense mutations ($TBX5^{S252I}$ and $TBX5^{Q292R}$) reduced the interactions (Figure A1.5K). These results suggest that disruption of the TBX5-NuRD complex *in vivo* causes cardiac abnormalities associated with HOS.

TBX5 can repress target genes in a NuRD dependent manner

Having demonstrated that the TBX5-NuRD interaction is disrupted by CHD-associated mutations in the NID, we queried if repression of TBX5 target genes was NuRD dependent. We performed transcriptional assays and observed, of the 11 validated targets repressed by TBX5, *Kctd16* and *Mef2b* fail to be repressed by TBX5 containing the CHD associated mutation $TBX5^{S261C}$ (Figure A1.6A-B, S5). Furthermore, $TBX5^{S261C}$ failed to repress transcription of multiple elements of the *Mef2b* gene (Figure A1.6B-D). Furthermore, we were able to ChIP CHD4 in the presence of TBX5 to these elements (Figure A1.6E-F). Taken together, these results imply the TBX5-NuRD complex is required to repress inappropriate gene expression in the heart and further suggest a subset of CHD mutants leads to the mis-expression of genes including *Mef2b* in the developing heart.

The TBX5-NuRD interaction arose concomitantly with the evolution of cardiac septation

Phylogenetic analysis of forty-eight TBX5 orthologues demonstrated there is complete conservation of the $TBX5^{NID}$ amino acids mutated in TBX5-associated CHD in animals with a septated heart ($TBX5^{S261}$, $TBX5^{V263}$, $TBX5^{K266}$) (e.g. *Xenopus* to human) (Figure A1.7). In contrast, orthologues of TBX5 in vertebrates that lack a septated heart (e.g. zebrafish) have an amino acid substitution present in human CHD patients, and furthermore, this substitution ablates the TBX5-NuRD interaction (Figure A1.5K). We hypothesized that the *Xenopus* ortholog of Tbx5 would interact with the NuRD complex, but a *Xenopus* Tbx5 ortholog with zebrafish NID amino acid substitutions would fail to interact with the NuRD complex. To test this hypothesis we generated the K266R

mutation in the *Xenopus* ortholog of Tbx5 (xTbx5^{K266R}). We observed that while mTBX5 and xTbx5 interact with the NuRD complex, xTbx5^{K266R} has a disrupted interaction with CHD4 (Figure A1.7B). Thus, our data suggested that the TBX5-NuRD interaction arose coincidentally with the evolution of cardiac septation and it is an essential component of an evolutionarily conserved mechanism for cardiac septum formation.

Discussion

Here we report the isolation and characterization of the first *in vivo* TBX5 interactome. For over 20 years, studies have defined TBX5 as an activator of transcription. However, our analysis demonstrates that TBX5 interacts biochemically and genetically with the transcriptional repression machinery of the NuRD complex. We further report TBX5 acts to repress inappropriate gene programs in cardiac tissue. We observed that only a subset of TBX5-repressed target genes failed to be repressed by the mutation of the TBX5-NuRD interacting domain (e.g. HOS mutation TBX5^{S261C}; Figure 6A-D, S5) implying that TBX5-mediated repression occurs in both a NuRD-dependent and NuRD-independent manner. We favor a model by which TBX5 regulates growth and differentiation of cardiomyocytes in part through the repression of alternate lineage-specific gene regulatory networks.

To date, all other predicted TBX5 interactions have been identified through in vitro studies, such as GST pulldowns in bacterial or cell lysates or yeast 2-hybrid approaches (Brown et al., 2005; Garg et al., 2003; Ghosh et al., 2009; Hiroi et al., 2001; Wang et al., 2011)(Krause et al., 2004; Murakami et al., 2005). In our study we isolate and characterize the endogenous TBX5 interactome from cardiac tissue under

physiological conditions. Therefore, this is the first in vivo isolation of any T-box protein in any species, which is highly significant given mutations in T-box genes are causative for over 30 human diseases. Though there is no in vivo evidence for interaction of proteins previously identified as being associated with TBX5 (BRG1, NKX2.5, GATA4, TBX20, Myocardin, TAZ and BAF60C), it remains formally possible that the proteins also interact with TBX5 in vivo at different time points of cardiac development then the period we analyzed or may not have identified in our study for technical reasons (e.g. distinct lysis buffer conditions for different studies). To the later point we note if we slightly relax the statistical cutoffs, we identified BRG1 (SMARCA4) (Takeuchi et al., 2011) in complex with TBX5.

TBX5 and cardiac septation

Our data suggest the TBX5-NuRD interaction arose coincidentally with cardiac septation. This hypothesis is further supported by our observations that *Tbx5*$^{+/-}$*; Mta1*$^{+/-}$ compound heterozygous mice have a higher frequency of cardiac septal defects than either single heterozygote alone (Figure 2J). AVSDs have been reported in 18 HOS patients, and 10 have been characterized to date (Baban et al., 2014). Of these, 6 of the mutations lead to the introduction of a stop codon before the TBX5^{NID}, and one mis-sense mutation maps within the TBX5^{NID} and ablates the TBX5-NuRD interaction (Figure 5K). Thus, we find a correlation between the phenotype of patients displaying intracardiac septal defects and the specific genotype of TBX5 patient mutations that causes absent or altered TBX5^{NID}.

Conclusions

Our studies demonstrate that a proteomic-based approach coupled with protein modeling provides a powerful predictive strategy in prioritizing patient mutations. This type of approach should be broadly applicable to the analysis of protein complexes in any cell or tissue type. The power of such an approach is underscored by recent high-throughput sequencing studies that have led to the identification of a vast number of somatic mutations in coding regions of potential disease causing genes. Exploiting such large datasets has proved problematic because it has not been possible to determine which mutations are naturally occurring, non-disease associated variants and which are disease-causing. Results presented here demonstrate a proteomic-modeling based strategy that can serve to prioritize mis-sense mutations and, therefore, correlate disease associated phenotypes with specific mutations.

Experimental Proceedures

Generation of Tbx5^{Avi} mice

The targeting construct to create *Tbx5^{Avi}* was created in collaboration with the UNC Animal Models Core and the UNC BAC Core (Chapel Hill). Briefly, the biotin acceptor peptide (Avi) targeting cassette was inserted in frame to the terminal exon of the *Tbx5* gene. Targeted ES cells were selected and introduced into B6 mouse blastocysts. High-chimera males were mated with B6 females to produce F1 heterozygotes (*Tbx5$^{Avi\text{-}Neo/+}$*). Incorporation of the *Tbx5^{Avi}* construct was confirmed by PCR and Southern analysis (Figure 1B), presence of *Tbx5^{Avi}* mRNA confirmed by RT-PCR (Figure 1D). F1 heterozygotes were crossed with *Sox2-Cre* mice (Hayashi et al.,

2002) to remove the neomycin cassette in the F2 generation (Figure 1C). F2 compound heterozygous mice were intercrossed to remove the *Sox2-Cre* allele and establish *Tbx5*$^{Avi/Avi}$ homozygous lines. *Tbx5*$^{Avi/Avi}$ mice were crossed to *Rosa26*^{BirA} mice (JAX), (Driegen et al., 2005). Resultant progeny were intercrossed to obtain compound homozygotes. Genotyping was performed by PCR. Genotyping primers in Supplemental Experimental Procedures. Southern probe was designed to *Tbx5* intron 8. Genomic DNA was digested with EcoRV or BamHI. Animal care and animal experiments were conducted in accordance with the Animal Care Committee at the University of North Carolina, Chapel Hill.

Preparation of cardiac nuclei

Frozen hearts from 4 week old *Tbx5*$^{Avi/Avi}$*; Rosa26*$^{BirA/BirA}$ (n=3, 2 replicates) and *Tbx5*$^{Avi/Avi}$ or *Rosa26*$^{BirA/BirA}$ control mice (n=3, 1 replicate per genotype) were homogenized using a mortar and pestle in liquid nitrogen. Nuclei were prepared as previously described (Franklin et al., 2011) and snap frozen in liquid N_2.

Solubilization of protein complexes

Plasmids were transfected into HEK-293 cells, harvested and lysed as previously described (Kaltenbrun et al., 2013). E9.5 hearts were harvested in PBS and lysed as described (Kaltenbrun et al., 2013). Cells and nuclear pellets were resuspended in lysis buffer (200mM K-HEPES pH 7.4, 1.1M KOAc, 20mM $MgCl_2$, 1% Tween-20, 10μM $ZnCl_2$, 10mM $CaCl_2$, 0.5% Triton X-100, 500 mM NaCl, 1X protease inhibitors (Roche), 1X phosphatase inhibitors (Sigma), optimized for TBX5 complexes. Nuclei were

homogenized using a Polytron (Kinematica), and processed for affinity purification or immunoprecipitation as described (Kaltenbrun et al., 2013).

Conjugation of magnetic beads and affinity purification

Conjugation of V5 (Invitrogen R960-CUS) or TBX5 (Aviva ARP33403-P050) antibody to magnetic beads (Invitrogen), and immunoisolation of protein complexes was performed as described (Cristea et al., 2005; Kaltenbrun et al., 2013), with following modifications for streptavidin-coated beads. Affinity purifications performed using streptavidin conjugated Dynabeads (Invitrogen) by slow rocking at RT, 30 minutes. The isolated protein complex was eluted for 20 min at 95°C in 40μL (80mM NaOAc pH 9.0, 95% formamide). Western blots were probed with mouse anti-V5 (1:5000, Invitrogen R960-CUS), rabbit anti-CHD4 (1:500, Active Motif 39289), mouse anti-HDAC2 (1:1000, Abcam Ab51832), rabbit anti-MTA1 (1:5000, Bethyl A300-280A), rabbit anti-TBX5 (1:1000, Aviva ARP33403-P050, using light chain-specific rabbit secondary), goat anti-TBX5 (1:500, Santa Cruz SC-17866), and mouse anti-GFP (1:10000, Clontech 63281).

Mass spectrometry

Isolated protein complexes were analyzed by mass spectrometry as previously described (Kaltenbrun et al., 2013). Briefly, tandem mass spectra were extracted by Proteome Discoverer (Thermo Fisher Scientific), and all MS/MS samples were analyzed with SEQUEST (Thermo Fisher Scientific; version 1.2.0.208), set up to search the human and mouse UniProt-SwissProt protein sequence database, assuming digestion pattern with trypsin. Scaffold (version Scaffold_3_00_06, Proteome Software Inc.,

Portland, OR) was used to validate MS/MS-based peptide and protein identifications. Peptide sequences were deemed a match if they could be established at >95.0% probability as specified by the Peptide Prophet algorithm (Keller et al., 2002). In turn, protein identifications were deemed a match if they could be established at >99.0% probability by the ProteinProphet algorithm and have at least two sequenced peptides. Protein identifications and associated spectral counts were exported to Excel for further data processing.

Interaction bioinformatics analysis

Protein identifications were filtered to exclude non-specific associations. Proteins were retained as specific candidates if they (1) had at least two spectral counts in both *Tbx5*$^{Avi/Avi}$*; Rosa26*$^{BirA/BirA}$ replicates with at least 4 spectral counts in one replicate, and (2) were uniquely identified or had at least a 2.5-fold spectral count enrichment in the *Tbx5*$^{Avi/Avi}$*; Rosa26*$^{BirA/BirA}$ condition versus the controls. Candidates were assigned gene ontology classification using the UniProt GOA annotations from within the Cytoscape (v3.2) platform (Cline et al., 2007). Gene symbols were submitted to the web-based STRING database (Franceschini et al., 2013) for interaction network analysis. Interactions with a combined STRING score of > 0.4 (medium confidence) were retained, exported, and further visualized in Cytoscape.

RNA extraction and RT-PCR

Adult (n=2) hearts were cryolysed (2.5 min/cycle, 30 Hz, 20 cycles) in Trizol (Invitrogen) and RNA extracted using Trizol and purified on RNeasy columns (Qiagen).

cDNA synthesis was performed from 0.5-1 μg of RNA using random primers and SuperScript II reverse transcriptase (Invitrogen). Expression levels assessed using GoTaq Green Master Mix (Promega) and Taq polymerase by GeneAmp PCR System (Applied Biosystems). PCR products were analyzed by gel electrophoresis. RT-PCR primer sequences are provided in supplemental Experimental Procedures.

Transcriptional assays

ChIP peak regions from candidate genes were amplified from mouse genomic DNA and subcloned into the PGL3-Promoter Luciferase vector (Promega). Primer sequences in Supplemental Experimental Procedures. HEK-293 cells were transfected with 600ng total DNA in 12 well plates using FuGene6 (Promega) and harvested 48 hours later for transcription assay which were performed using Dual Reporter Assay System (Promega) according to the manufacturer's protocol in three biological and three technical replicates per assay.

DNA constructs

Generation of the *Tbx5*^{Avi} targeting construct for embryonic stem cells and generation of chimeric mice was performed by JrGang Chen (UNC BAC Core-Chapel Hill). *Tbx5-Avi-V5* was generated by synthesizing an oligo encoding the Avi tag (5' ggc ctg aac gac atc ttc gag gct cag aaa atc gaa tgg cac gaa 3') and subcloning it into the pcDNA3.1/V5-His TOPO vector (Invitrogen). Full length and truncated versions of mouse *Tbx5* were then subcloned into pcDNA3.1-Avi-V5. pEF1α-BirA-V5-Neo was generously provided by Stuart Orkin (Harvard University) (Kim et al., 2009). GFP tagged

full length, C-terminal, and N-terminal human CHD4 constructs were generously provided by Stephen Jackson (Cambridge) (Polo et al., 2010). Site-directed mutagenesis primers are available in Supplemental Experimental Procedures.

Genetic interaction and histology

Tbx5$^{fl/fl}$ mice were generously provided by Jon and Christine Seidman (Harvard Medical School). The *Tbx5$^{+/-}$* mouse has been reported (Bruneau et al., 2001). The *Mta1$^{+/-}$* mouse line was generated by Harold Olivey (Indiana University Northwest) and generously provided by Eric Svensson (Novartis Institutes for Biomedial Research). All mice are in a mixed B6/129/SvEv/CD-1 background and experiments were performed according to a protocol reviewed and approved by the Institutional Animal Care and Use Committee of the University of Chicago, in compliance with the USA Public Health Service Policy on Humane Care and Use of Laboratory Animals. E13.5 and E14.5 wild type, *Tbx5$^{+/-}$*, *Mta1$^{+/,-}$* and *Tbx5$^{+/-}$;Mta1$^{+/-}$* embryos were dissected from timed matings. Embryos fixed in 4% PFA overnight and sent to the Human Tissue Resource Center at the University of Chicago for paraffin embedding, sectioning, and H&E staining. Slides were imaged on a Leica DM2500 microscope and QImaging Retiga 2000R 1394 Color Cooled camera. *Mta1$^{-/-}$* null hearts were sectioned and prepared as previously described (Dorr et al., 2015).

RNA-seq

Heart tube dissections were performed as described (Hoffmann et al., 2014). Briefly, E9.5 *Tbx5$^{+/+}$* and *Tbx5$^{-/-}$* heart tubes were dissected away from the body nearest

the myocardial reflections and four heart tubes were pooled to isolate sufficient RNA. Genotypes assigned via PCR. Pools of tissue were mechanically homogenized in TRIzol Reagent (Invitrogen, Carlsbad CA) and RNA was isolated in accordance with the manufacturer's instructions. RNA resuspended in water was further purified with 500 µl of 1-butanol (4 times) and 500 µl diethyl ether (twice) (Krebs et al., 2009). Isolated RNA underwent library preparation and 51-bp single-ended RNA sequencing at the Genomics Core Facility at The University of Chicago. Briefly, mRNA was selected using an oligo-dT pulldown, and barcoded libraries were prepared according to Illumina's instructions accompanying the TruSeq RNA Sample prep kit v2 (Part# RS-122-2001). Library fragments of ~275 bps (insert plus adaptor and PCR primer sequences) were quantitated using the Agilent Bio-analyzer 2100 and pooled in equimolar amounts. The pooled libraries were sequenced on the HiSeq2500 in Rapid Run Mode following the manufacturer's protocols (Invitrogen, 2013).

The raw data output from the Illumina Genome Analyzer was in phred-scaled base-64 fastq format, representing the sequence and quality scores for each read. The single-ended 51-bp reads were filtered based on their quality scores using FastQC toolkit (Andrews S, 2010) , selecting for the 37-bp reads on the right with a minimum quality score of 35 in 75% of the bases of the read. The reads passing the filter were aligned to the mouse genome mm10 assembly by Tophat (version 2.0.6) using default parameters (Kim et al., 2013). The output SAM files were filtered for alignment sorted, and converted to BAM files using Samtools (Li et al., 2009). Hit counts were estimated and normalized by both each transcript's length and the total yield of the sequencer using Cufflinks (version 2.1.1, default parameters and "-p 2") (Trapnell et al., 2012). Among all Cufflinks

reported transcript abundances, the Fragments Per Kilobase of exon per Million fragments mapped (FPKM), values larger than 0.125 in all samples were kept for the following statistical analysis. This filter guaranteed the median correlation coefficient >0.96 between any two biological replicates while kept ~65% of all transcripts, including 11789 coding genes (Ensembl annotation, mm10, 2014 April version). Differential expression between mutant samples (n=2) and wild types (n=4) was tested using Cuffdiff (v2.1.1) (Trapnell et al., 2012) Although Cuffdiff lacks power for designs with fewer than three replicates per group, the low variability of the two biological replicates in all cases (r>0.96) guarantees the statistic power of a fold change to reveal biologically meaningful results. All genes with an adjusted p-value less than 0.05 and the absolute fold change larger than 1.57 were called significant.

Large Scale bioinformatics analysis of TBX5 binding sites

Differentially expressed genes were determined by q-value (<0.05) and status ("OK"). Those genes that did not meet the status threshold were given a non-significant q-value (qval=1). TBX5 ChIP-Seq FASTQ data was downloaded from GEO (GSE21529). Tags were filtered using Tag-dust (Lassmann et al., 2009) and aligned to the mouse genome (mm9) using bowtie (Langmead et al., 2009). Peaks were called using MACS2 (Zhang et al., 2008) with default parameters. Peaks were associated with differentially expressed genes and differential motif enrichment was determined using BETA (Wang et al., 2013) with the following options –-pn 40000 and –-df 0.05 –-da 1. Because BETA uses a curated database that lacks the TBX5 motif, TBX5 peaks associated with differentially regulated genes were also analyzed for motif enrichment using HOMER

(Heinz et al., 2010). Genes associated with TBX5 regulation were then analyzed for pathway enrichment using DAVID (Huang da et al., 2009). The top 10 (by p-value) KEGG pathways are shown for both activated and repressed genes.

Chromatin IP

HEK-293 cells were transfected with the *Kctd16-Luciferase* construct used in transcriptional assays, as well as wild type TBX5 or TBX5^{S261C} (20 μg total DNA) using FuGene 6 (Promega). Cells were fixed in 1% PFA for 10 minutes at RT. 10 million cells were used per ChIP. Cells were lysed in ChIP Buffer (50mM Tris pH 7.5, 140 mM NaCl, 1mM EDTA, 1 mM EGTA, 0.1% sodium deoxycholate, 0.1% SDS), sonicated using a Branson 450d, 12 cycles, 30s/cycle (1s on, 0.5s off), 30% amplitude), and spun at top speed to isolate chromatin. Triton X-100 was added to 1%, and the chromatin was added to CHD4-conjugated Protein G beads (Life Technologies) at 4°C O/N. ChIP was resumed using ChIP Protocol (Agilent). ChIP primer sequence is available in Supplemental Experimental Procedures.

TBX5 NID Modelling

The mouse TBX5 sequence was submitted to the DisProt PONDR-FIT disorder predictor in order to identify disordered regions in the protein. Both the entire protein and just the NID region were submitted to the fold recognition server HHpred (http://toolkit.tuebingen.mpg.de/hhpred) to determine if any suitable templates were available for structural modeling. A small coil-to-helical region involving residues 255-264 was predicted by HHpred. This region of TBX5 was modeled based on the

correspondingly small region of the structure of the CRISPR-associated protein [PDB ID 3VZI].

Author Contributions:

K.D. performed Southern analysis to determine *Tbx5*^{Avi} positive clones. J.K. and J.S. performed all experiments and analysis related to the *Tbx5*$^{+/-}$*; Mta1*$^{+/-}$ mouse, and C.W. performed the *Mta1*$^{-/-}$ experiment and analysis. B.T. modeled the TBX5 protein structure, and generated the TBX5 evolutionary tree, T.G. analyzed proteomics data. N.C.G. performed large scale bioinformatics analysis of TBX5 binding motifs and pathway analysis. I.J.D. provided additional bioinformatics expertise. I.L.C. provided additional proteomics expertise. I.P.M provided RNA-seq data. L.W. performed all other experiments. L.W. and F.L.C. designed the experiments and wrote the paper. All authors discussed the results and commented on the manuscript.

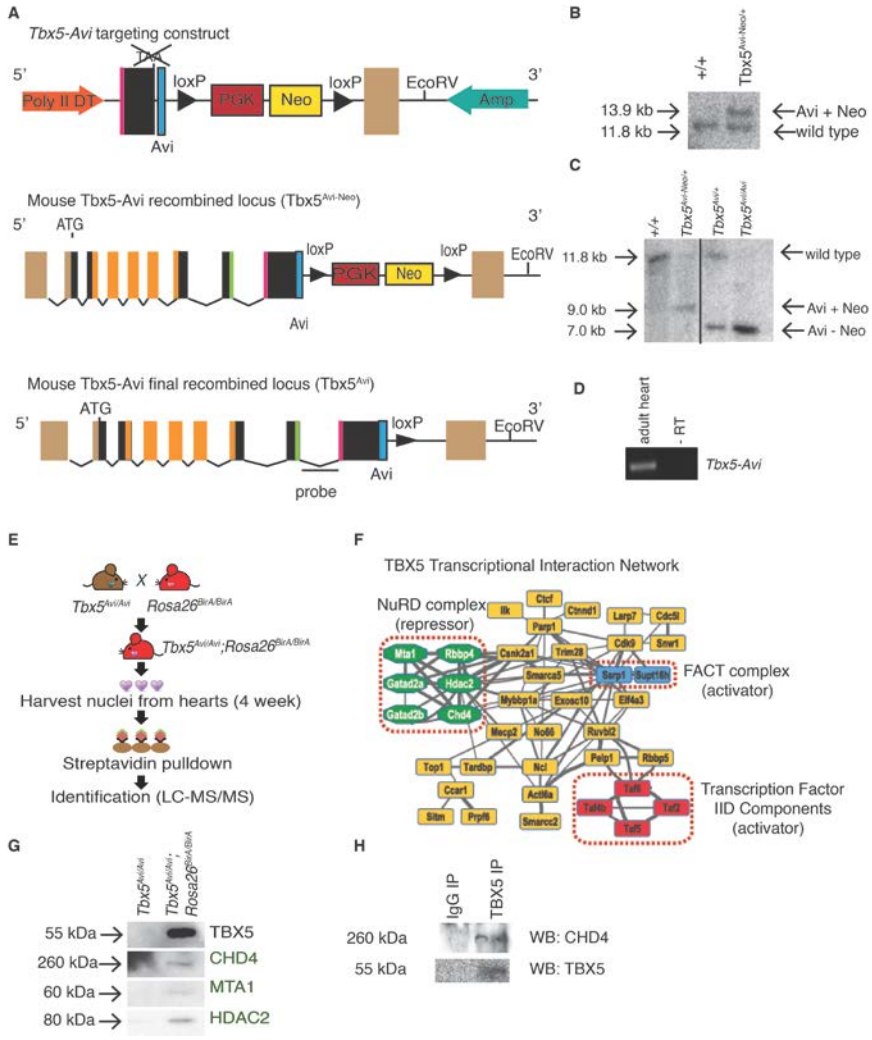
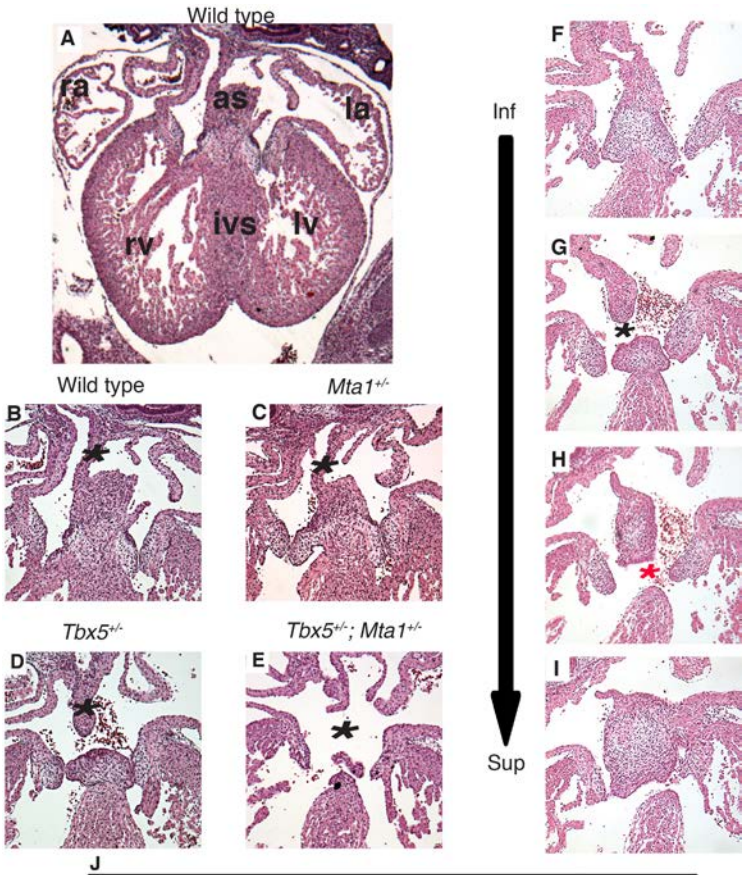


Figure A1.1. TBX5 interacts with the NuRD complex. (**A**) (Top) The *Tbx5-Avi* targeting cassette contains the biotin acceptor peptide (Avi) fused in frame to the terminal coding exon of *Tbx5*. Neo-neomycin resistance gene. PGK- phosphoglycerate kinase-1 promoter. Amp-Ampicillan resistance gene. Exons colored brown represent untranslated exons (5' and 3' UTR). Exons colored orange represent the T-box DNA binding domain. Exons colored green represent the nuclear localization signal. Exons colored pink represent the activation domain. (Middle) Schematic of the *Tbx5* recombined locus (*Tbx5*$^{Avi\text{-}Neo}$) upon homologous recombination. (Bottom) Schematic of the *Tbx5* locus upon introduction of *Cre* (*Tbx5*^{Avi}). (**B**) Southern blot of positive

embryonic stem cell clone confirms homologous recombination. (**C**) Southern blot of F2 mice confirms removal of Neo cassette upon introduction of *Cre*. Samples were run on the same gel but were noncontiguous. (**D**) RT-PCR analysis of *Tbx5*$^{Avi/+}$ adult hearts demonstrates presence of *Tbx5*^{Avi} mRNA. **(E)** Overview of the isolation and characterization of the endogenous cardiac TBX5 interactome. (**F**) TBX5 transcriptional interaction network. Network nodes, labeled with mouse gene symbols, are candidate direct or indirect TBX5 protein interactions identified from affinity enrichment of biotinylated TBX5. Network edges represent known and/or predicted functional interactions in the STRING database. Edge thickness reflects the combined STRING evidence score for each binary relationship. Thicker edges represent increased interaction evidence. Selected transcriptional complexes within the network are highlighted in red dashed boxes. (**G**) Affinity isolation of endogenous TBX5^{Avi} from 4 week cardiac nuclei confirms interaction of TBX5 with endogenous components of the NuRD complex. (**H**) Affinity isolation of TBX5 from E9.5 hearts further confirms interaction of TBX5 with the NuRD component CHD4 in the embryonic heart.



J

	Wild type	*Mta1*$^{+/-}$	*Tbx5*$^{+/-}$	*Tbx5*$^{+/-}$*; Mta1*$^{+/-}$
Normal	12	14	4	0
ASD	0	0	1	0
VSD	0	0	3	8
CAVC	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 A1.2. TBX5 and the NuRD complex genetically interact. Histology (Hematoxylin and Eosin) 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*$^{+/-}$*; Mta1*$^{+/-}$ heart from inferior to superior revealing ASD (**G**) (black asterisk) and VSD (red asterisk) (**H, 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 = intervectricular septum, la = left atrium, lv = left ventricle, ra = right atrium, rv = right ventricle.

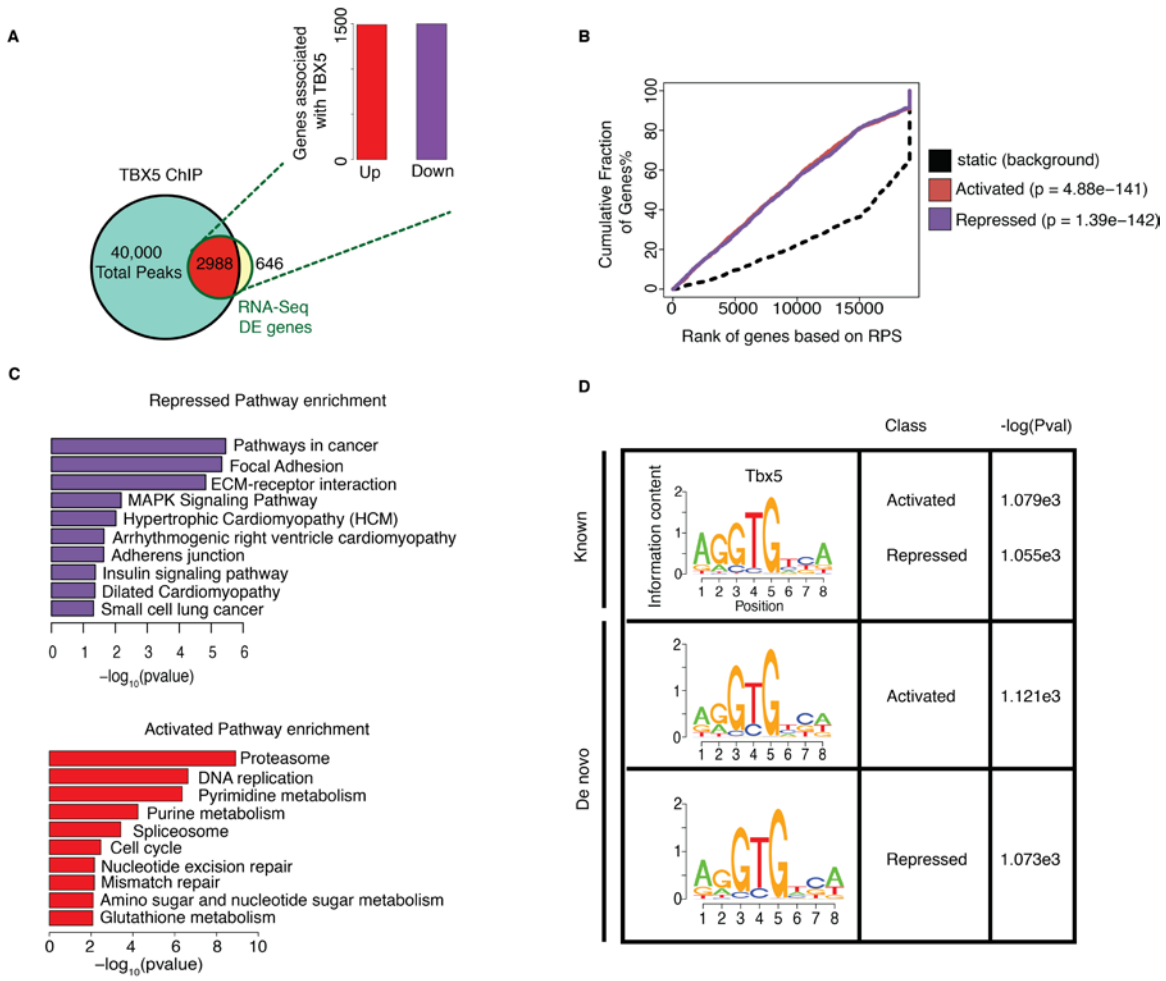


Figure A1.3. Analysis of TBX5 binding motifs in activated vs repressed genes. (**A**) Schematic overlay between TBX5 peaks (He et al., 2011) and differentially expressed genes between wild-type and *Tbx5* null heart tissue. 2988 genes are differentially expressed and associated with a TBX5 peak. Inset is number of up and downregulated genes. (**B**) Activating/Repressive function prediction of TBX5 peaks. Red and purple lines represent up and down-regulated 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 represent significance of up or down-regulated group relative to background (static genes) by Kolmogorov-Smirnov test. (**C**)

Kegg Pathway enrichment of down-regulated genes associated with TBX5 binding ranked by $-\log_{10}$(pvalue). (**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.

A

Gene	Log2 fold change	ChIP peak location	ChIP peak value(s)	Size of element
Sncb	5.01	intron 4	18	321 bp
Nxph4	4.93	5' UTR (5.7 kb from TSS)	41	490 bp
Mef2b	4.61	5' UTR (7.4 kb from TSS)	27	347 bp
Kctd16	4.53	5' UTR (1.4 kb from TSS)	17	426 bp
Gad1	4.19	intron 1	17	276 bp
Kcne3	3.63	5' UTR (100bp from TSS)	23	350 bp
Fgf11	3.49	intron 1	45	1,200 bp
Col20a1	1.95	5' UTR (2.7 kb from TSS)	17, 35	1,100 bp
Klf4	1.74	exon 3	57, 56	3,291 bp
Fgf8	1.71	5' UTR (2.8 kb from TSS)	80	384 bp
Fbxl20	1.37	5' UTR/ exon 1/ intron 1	25, 56, 30	1,600 bp
Casz1-1	1.26	exon 1/ intron 1	60, 29, 43	3,200 bp
Casz1-2	1.26	intron 1	32, 22, 18	2,400 bp
Kcnn2	1.23	exon 1	13	140 bp
Plekha2	1.21	5' UTR (1.3 kb from TSS)	49, 33	1,700 bp

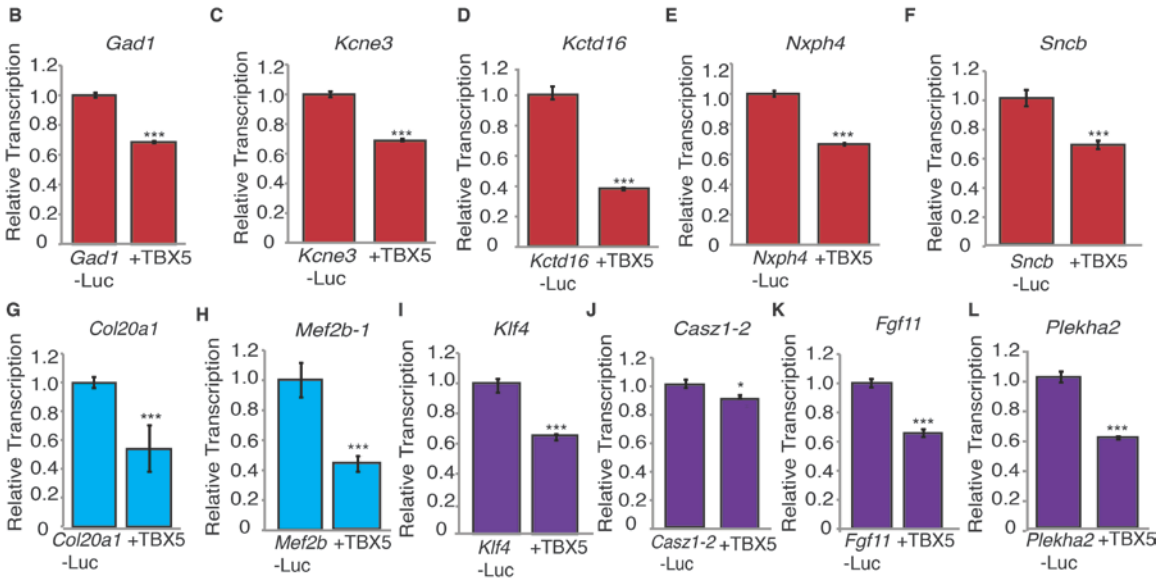


Figure A1.4. TBX5 functions to represses misexpression of genes in cardiac tissue. (**A**) Summary of fold change (RNA-seq) and ChIP-seq peak values of the 15 top rank ordered genes. (**B-L**) Gene reporter elements cloned from potential TBX5 targets in the presence or absence of TBX5. Transcriptional assays of target genes show TBX5 acts as a transcriptional repressor in 11 of 15 elements tested (additional data in Figure S4). Genes defined as neural are shown in red, expressed in cancer as blue, or both as purple. Graphs are plotted as the mean luciferase value ± SEM. Student's two tailed t-test was used to perform statistical analysis. * $p< 0.05$ *** $p < 0.001$.

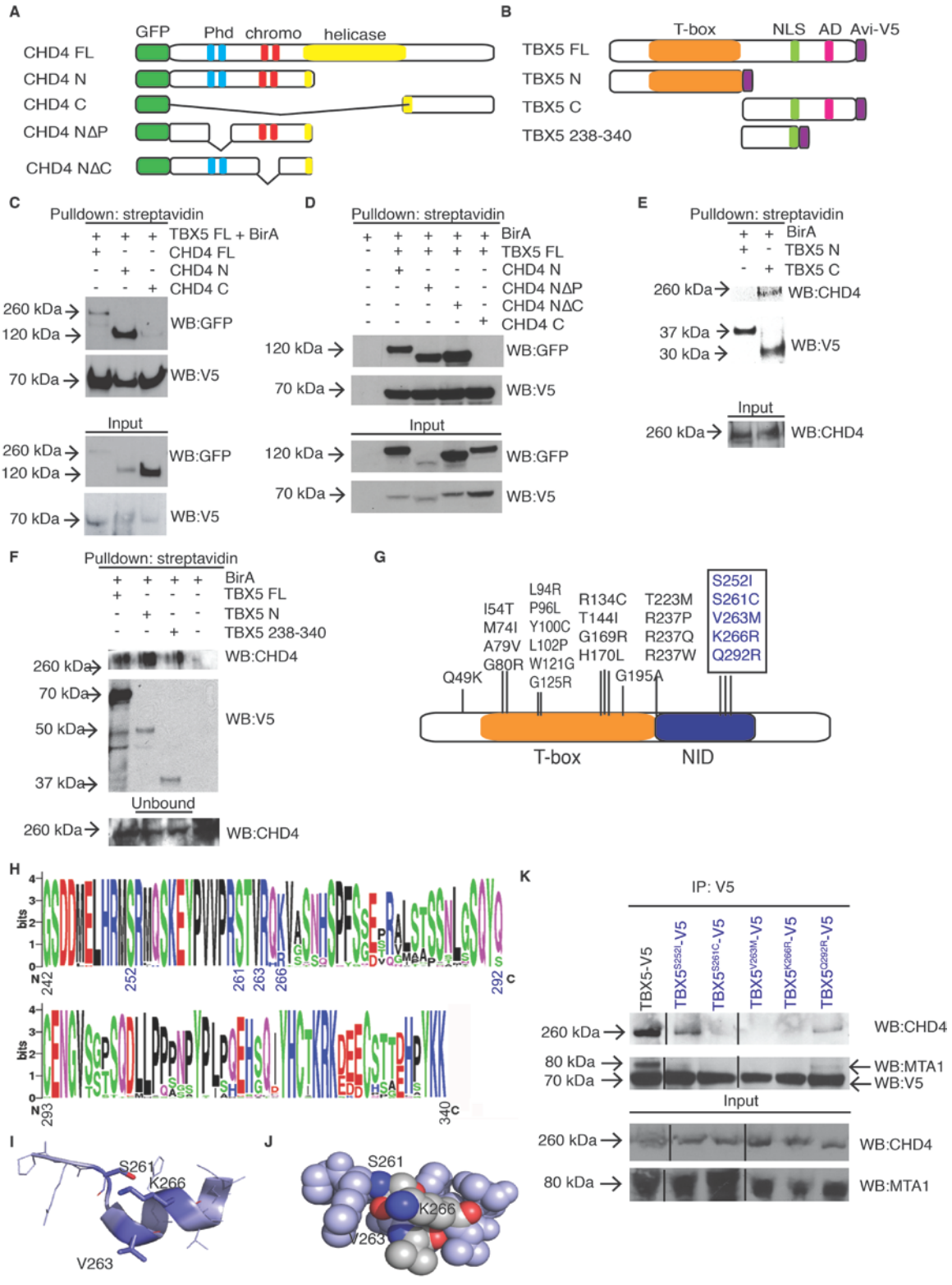

A
GFP
Phd
chromo
helicase
CHD4 FL
CHD4 N
CHD4 C
CHD4 NΔP
CHD4 NΔC
B
T-box
NLS
AD
Avi-V5
TBX5 FL
TBX5 N
TBX5 C
TBX5 238-340
C
Pulldown: streptavidin
TBX5 FL + BirA
CHD4 FL
CHD4 N
CHD4 C
260 kDa
120 kDa
WB:GFP
70 kDa
WB:V5
Input
D
Pulldown: streptavidin
BirA
TBX5 FL
CHD4 N
CHD4 NΔP
CHD4 NΔC
CHD4 C
E
Pulldown: streptavidin
BirA
TBX5 N
TBX5 C
260 kDa
WB:CHD4
37 kDa
30 kDa
WB:V5
Input
F
Pulldown: streptavidin
BirA
TBX5 FL
TBX5 N
TBX5 238-340
260 kDa
WB:CHD4
70 kDa
50 kDa
37 kDa
WB:V5
Unbound
G
I54T
M74I
A79V
G80R
L94R
P96L
Y100C
L102P
W121G
G125R
R134C
T144I
G169R
H170L
T223M
R237P
R237Q
R237W
S252I
S261C
V263M
K266R
Q292R
Q49K
G195A
T-box
NID
H
bits
N
C
I
S261
K266
V263
J
S261
K266
V263
K
IP: V5
TBX5-V5
TBX5S252I-V5
TBX5S261C-V5
TBX5V263M-V5
TBX5K266R-V5
TBX5Q292R-V5
260 kDa
WB:CHD4
80 kDa
70 kDa
WB:MTA1
WB:V5
Input
260 kDa
WB:CHD4
80 kDa
WB:MTA1

Figure A1.5. Congenital heart disease associated mutations of TBX5 disrupt TBX5-NuRD complex interaction and activity. (**A**) Schematic of CHD4 deletion constructs. Phd= Plant homeo domain. CHD4 FL = aa 1-1937, CHD4 N = aa 1-758, CHD4 C = aa 1183-1937, TBX5 FL = aa 1-518, TBX5 N = aa 1-237, TBX5 C = aa 238-518. (**B**) Schematic of TBX5 deletion constructs. NLS = Nuclear Localization Signal. AD = Activation domain. (**C-D**) Affinity isolation of full-length TBX5-Avi-V5 with GFP-tagged CHD4 deletions reveals that CHD4 interacts with TBX5 via sequences in the N-terminus of CHD4, but that the Phd and chromo domains are not individually required for the TBX5-CHD4 interaction. **(E-F**) Affinity isolation of CHD4 with a TBX5-Avi-V5 deletion series demonstrates TBX5 sequences aa 238-340 are both necessary and sufficient for interaction with CHD4. This region has been termed the NuRD interaction domain (NID). (**G**) Schematic of human TBX5 protein showing location of a subset of CHD associated mis-sense mutations. NID= NuRD interacting domain (aa 238-340). (**H**) Sequence alignment of the $TBX5^{NID}$ from 48 TBX5 ortholgues**.** Height of letters is relative conservation at that residue. (**I**) Structural prediction of the $TBX5^{NID}$ shows the $TBX5^{NID}$ is comprised of a coil region followed an α-helix. (**J**) CHD associated mis-sense mutations align along a single surface of the α-helix and are predicted to disrupt protein-protein interactions (representative residues $TBX5^{S261}$, $TBX5^{V263}$, and $TBX5^{K266}$ shown in dark blue). (**K**) Immunoisolation of V5-tagged CHD associated mis-sense mutations of TBX5 probed for CHD4 and MTA1.

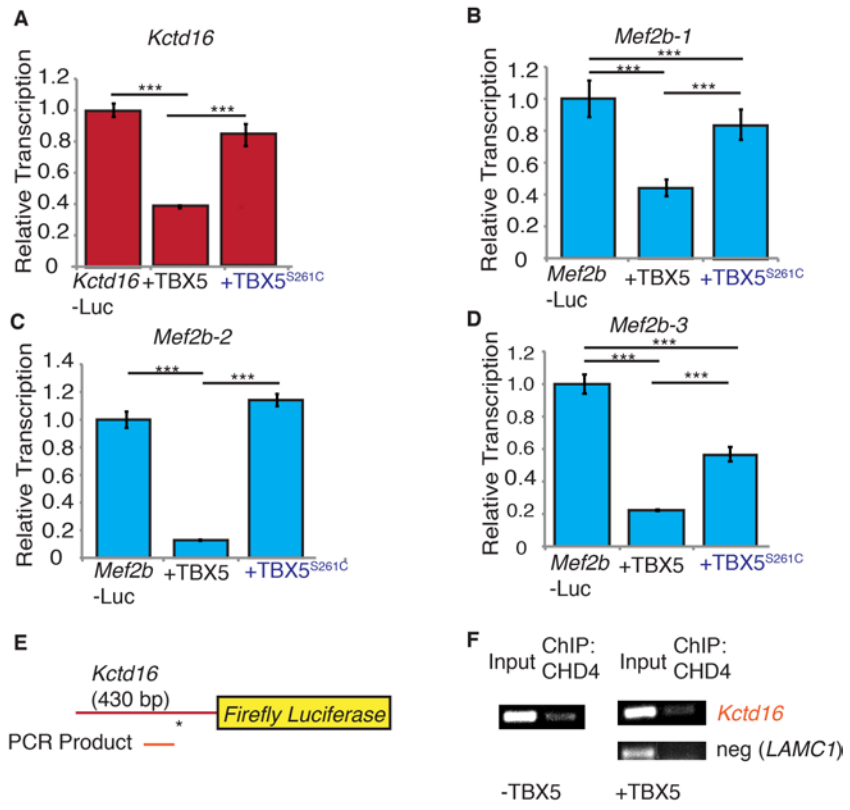


Figure A1.6. Congenital heart disease associated mutations of TBX5 disrupt TBX5-NuRD complex activity. (**A-D**) Transcription of the regulatory regions of TBX5 target genes in response to wild type TBX5 (second column) or TBX5^{S261C} (third column). Additional gene targets shown in Figure S7. n= 9 per transcriptional assay. Graphs are plotted as the mean luciferase value ± SEM. Student's two tailed t-test was used to perform statistical analysis. *** $p < 0.001$. (**E**) Schematic of ChIP primer location on *Kctd16* element. The T-box binding element (asterisk) is noted. (**F**) ChIP of CHD4 from cells co-transfected with *Kctd16-Luciferase* in the presence of absence of TBX5.

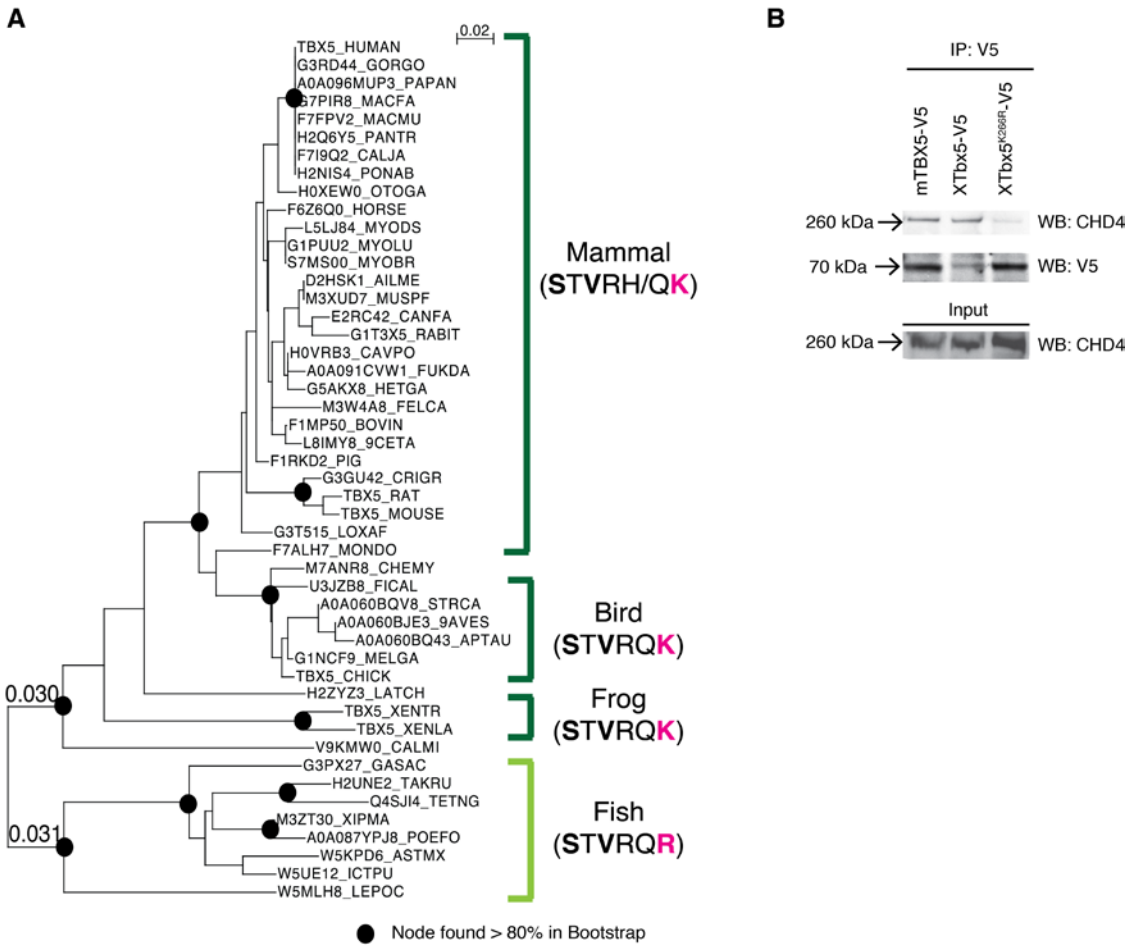


Figure A1.7. The TBX5-NuRD interaction evolved concurrently with cardiac septation. (A) Phylogenetic comparison of TBX5 orthologues with boot-strap values. Dark green brackets represent animals with a septated heart. Light green brackets represent animals without a septated heart. Sequence of core amino acid residues required for the TBX5-NuRD interaction are shown to the right. Amino acid residues mutated in TBX5-associated CHD are bolded, and a K->R substitution from fish to frog is highlighted in pink. (**B**) Immunoisolation of V5-tagged TBX5 orthologs probed for CHD4.

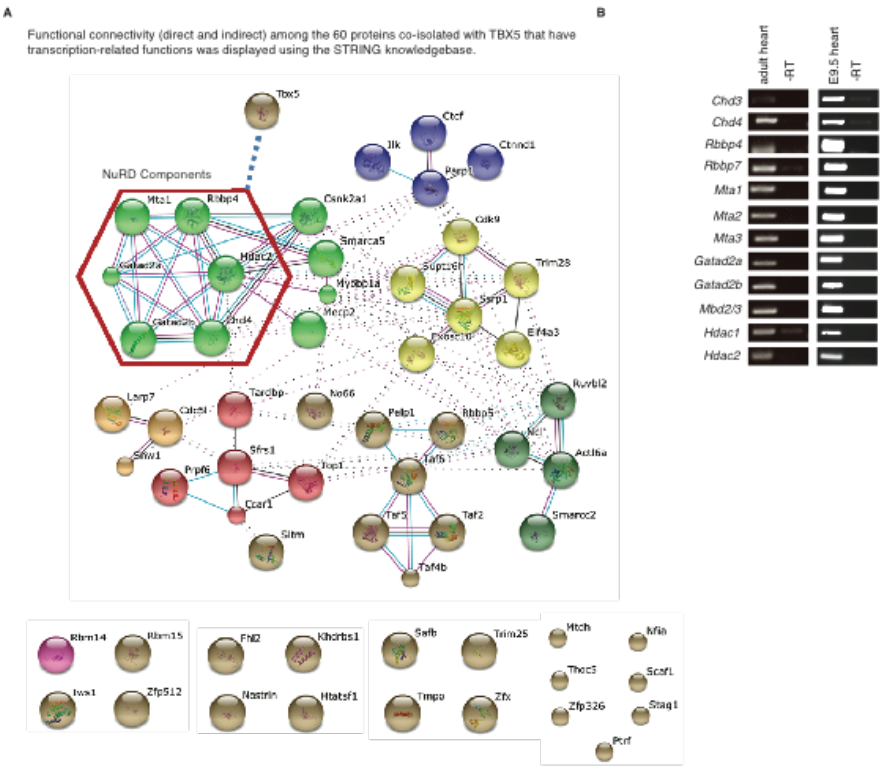


Figure SA1.1. The TBX5 transcription interaction network. (**A**)Sixty mouse genes corresponding to TBX5 candidate interactions as identified by the gene ontology function of transcription were uploaded to the STRING database (www.string-db.org). STRING functional relationships were evaluated with default settings with text mining disabled. Forty proteins were clustered in a single protein interaction network, reflecting a minimum combined STRING score of > 0.4 (medium confidence) between each binary relationship (left). This network was imported to Cytoscape for further processing (see Fig 1). The remaining 20 nodes with scores < 0.4 are illustrated as disconnected (bottom). (**B**) RT-PCR analysis of all possible members of the NuRD complex demonstrates the NuRD complex is present in both the adult and embryonic heart.

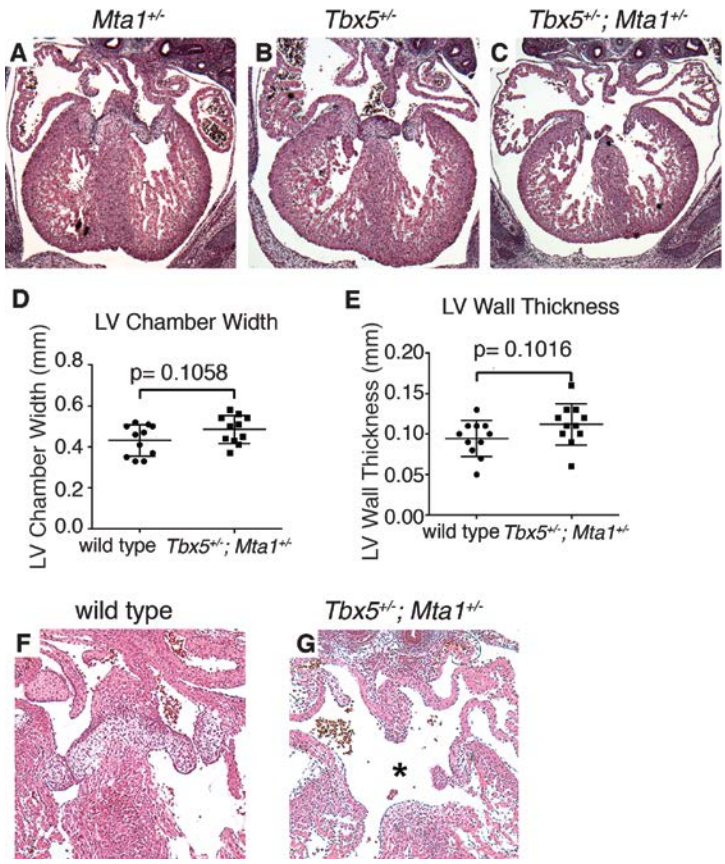


Figure SA1.2. TBX5 and MTA1 and cardiac septation. Histology staining (Hematoxylin and Eosin) of (**A**) *Mta1*$^{+/-}$, (**B**) *Tbx5*$^{+/-}$ and, (**C**) *Tbx5*$^{+/-}$*; Mta1*$^{+/-}$ mice at E13.5. (**D-E**) Measurements of left ventricular (LV) chamber width (**D**) and wall thickness (**E**) of wild type and *Tbx5*$^{+/-}$; *Mta1*$^{+/-}$ embryonic hearts (n=11 per genotype) at E13.5 from the 4-chamber view at the widest part. Data is reported as either chamber width or wall thickness in millimeters ± SEM. P value was determined by Student's T-test. (**F**-**G**) Histology staining of (**F**) wild type and (**G**) *Tbx5*$^{+/-}$*; Mta1*$^{+/-}$ mice at E14.5 demonstrates a CCAVC at this stage (asterisk).

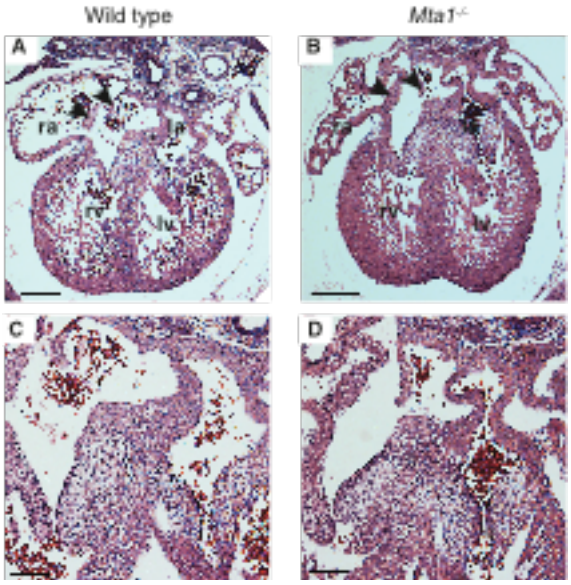


Figure SA1.3. *Mta1*$^{-/-}$ embryos undergo normal heart chamber septation. Histology staining (Hematoxylin and Eosin) of heart tissue at E13.5. Low magnification transverse section through (**A**) wild-type heart and (**B**) *Mta1*$^{-/-}$ heart. Asterisks denote region of higher magnification, arrows denote leaflets of venous valve of inferior vena cava, scale bar=310μm. High magnification of asterisk region in (**C**) wild-type heart and (**D**) *Mta1*$^{-/-}$ heart at point of chamber septation. Scale bar=130μm. la = left atrium, lv = left ventricle, ra = right atrium, rv = right ventricle.

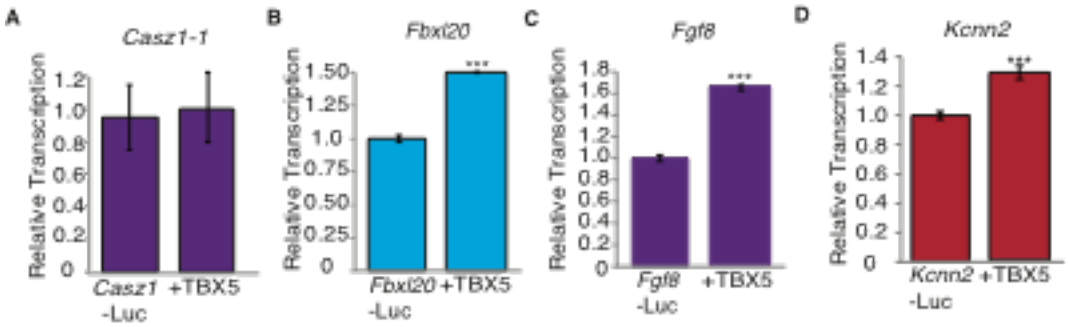


Figure SA1.4. Transcriptional targets not repressed by TBX5. (**A-D)** Gene reporter elements cloned from potential TBX5 targets in the presence or absence of TBX5. Data is reported as the mean ± SEM. Student's T-test was used to determine significance. *** $p < 0.001$.

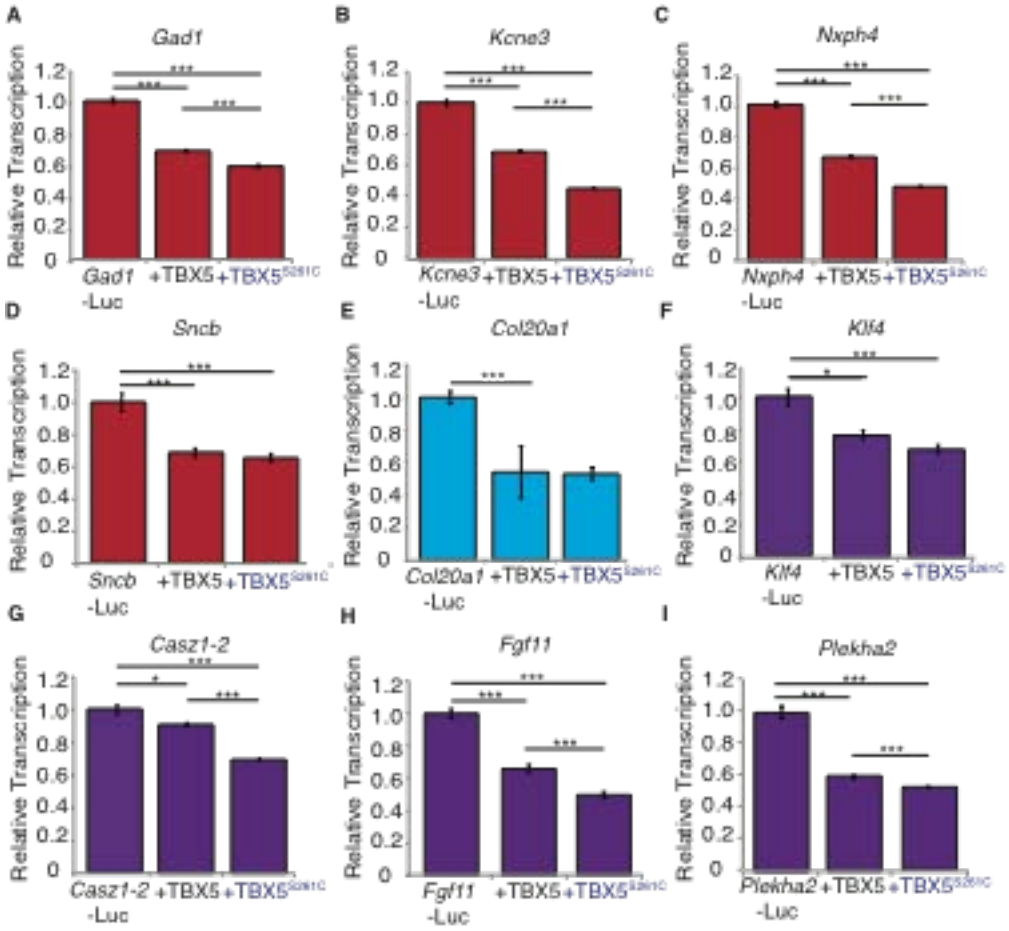


Figure SA1.5. Targets repressed by TBX5 in a NuRD independent manner. (**A-I)** TBX5-repressed gene reporter elements and in the co-transfected with wild type TBX5 or $TBX5^{S261C}$. Genes sorted into neural (red), cancer (blue), or both (purple). Data is reported as the mean ± SEM. Student's T-test was used to determine significance * p< 0.05 *** p < 0.001.

Table S1. Spectral counts for TBX5 interacting proteins, related to Figure 1 (Excel File upon request)

Table S2. Complete list of genes misregulated in *Tbx5*$^{-/-}$ embryonic heart, related to Figures 3-4 (Excel File upon request)

Table S3. TBX5 ChIP data for genes upregulated in *Tbx5*$^{-/-}$ embryonic hearts, related to Figures 3-4. (Excel File upon request)

Supplemental Experimental Proceedures

Generation of Tbx5^{Avi} mice

Genotyping primers for *Tbx5$^{Avi/Avi}$; Rosa26$^{BirA/BirA}$* mice.

Genotyping Primers	Sequence
Wild type *Tbx5*	F: CGGCGTGCCCAGGACCCTGT R: GGCACAGGTCAGCCTTTAGC
Tbx5^{Avi}	F: CGGCGTGCCCAGGACCCTGT R: GCGCGCCTTCGTGCCATTCGA
Sox2-Cre	F: GTCCCCTTCTCCATCTCCAG R: GCAAACGGACAGAAGCATTT
Neo	F: AGGATCTCCTGTCATCTCACCTT R: CTAGAGAATTGATCCCCTCAGAAG
Presence of BirA	F: TTACGCAAGCTGGGTGCAGA R: TTACGCAAGCTGGGTGCAGA
Absence of BirA	F: GTGTAACTGTGGACAGAGGAG R: GAACTTGATGTGTAGACCAGG

Transcriptional assays

TBX5 target gene primers for transcriptional assays.

ChIP cloning PCR Primers	Sequence
Casz1-1	F: ATCGGGTACCTGTCCCTTTCTTCTGTCCCA R: ATCGGGTACCAGGCATAAGATCGCTGGAAA
Casz1-2	F: ATCGGGTACCACCATCCATCCTCGTGTAGC R: ATCGGCTAGCGCCAAAAACAGTGGCCTAAA
Col20a1	F: ATCGGCTAGCATCGAGTGTGTCAGCTGTGG R: ATCGAAGCTTCCGTGGTGCTGTGTTCTTAC
Fbxl20	F: GATCACGCGTGGCCCAAGGAATGGGTACT R: GATCACGCGTCCACAGTCGCCTCTATCGTC
Fgf8	F: CGAGGCTAGCCCAGCTCCCTCCTTCCTTTA R: CGAGGGTACCGGCAGCTTTCTGTCTTGGTT
Fgf11	F: GCACAGATCTGGGACTCCCTAACTGTCGTACC R: GCACACGCGTCCGCTGGAGCTAGTCAGG
Gad1	F: CGAGGGTACCCACTGGAATTCCTACCGTGA R: CGAGGCTAGCCAAGGCGTTCTGGTCTAAGG
Kcne3	F: CGCTACGCGTGTGCCACACAGCTACAATAA R: CGCTCTCGAGTCCGGTGAACACTGCAATAA
Kcnn2	F: GATCGGTACCCTGGTCTCCCTGCAGCTTTA R: GATCGCTACGGTTGTCCTGGCTCTGTTGCT
Kctd16	F: AGGACGCGTTGCTTGTGCAAACTCCAGAC R: GACCAGATCTAGGCAGGGCGACAGATAGAT
Klf4	F: ATCGACGCGTCTTGGCGTGAGGAACTCTCT R: ATCGAGATCTCGCTCAAATGGGCCTCTA
Mef2b-1	F: CGAGGTACCTCTTCTCTCGGACGGACCTA R: CGAGGCTAGCCACACACACTCACGGCTCTC
Mef2b-2	F: CGAGGCTAGCCTGCACAGGGACCCACAG R: CGAGGCTAGCCTGCACAGGGACCCACAG

Mef2b-3	F: CGAGGGTACCGCGTGCGATCACCATAACT R: CGAGGCTAGCAGGGGCTGCGATGTAGGT
Nxph-4	F: CGAGGGTACCGCTTGCAGAAGGGTATCTGG R: CGAGGCTAGCTAAGCCTCCAGGCATTCAAC
Plekha2	F: GATCGGTACCTAACTTGGAACCGACCTACG R: GATCGCTAGCTTTCACACCTGAGCGAACTG
Sncb	F: CACGGGTACCTCCCTCCACTGCCTCCAC R: CAGCGCTAGCGCTCCAGGGTCCTCCTAGTC

DNA constructs

Site directed mutagenesis primer pairs. Amino acid changes are represented in bold.

mTBX5^{S252I}	F: GATGTCTCGGATGCAA**ATT**AAAGAGTATCCTGTGGTTC R: GAACCACAGGATACTCTTT**AAT**TTGCATCCGAGACATC
mTBX5^{S261C}	F: GTATCCTGTGGTTCCCAGG**TCG**ACAGTGAGGCACAAAG R: CTTTGTGCCTCACTGT**CGA**CCTGGGAACCACAGGATAC
mTBX5^{V263M}	F: GGTTCCCAGGAGCACA**ATG**AGGCACAAAGTCACCTCC R: GGAGGTGACTTTGTGCCT**CAT**TGTGCTCCTGGGAACC
mTBX5^{K266R}	F: GAGCACAGTGAGGCAC**AGA**GTCACCTCCAACCACAGC R: GCTGTGGTTGGAGGTGAC**TCT**GTGCCTCACTGTGCTC
mTBX5^{Q292R}	F: CAATTTAGGGTCCCAGTAC**CGG**TGTGAGAATGGTGTC R: GACACCATTCTCACA**CCG**GTACTGGGACCCTAAATTG
xTbx5^{K266R}	F: GTCCGTCAG**AGG**GTGTCCTCT R: TGTGCATCTTGGGACTAC

Chromatin IP

Kctd16 intron 1 ChIP-PCR primer sequences

Kctd16 intron 1	F: GTTTCTTGCCTGCACTGCT R: GAAGGAAAAGCAGCCGGGAA

RT-PCR

Primer sequences for NuRD complex members

Chd3	F: AGAAGGAAAACAAGCCAGG R: TTGTGTCCCCCATCCCCTT
Chd4	F: GGGTGAAATGAATCGTGGC R: TTGTGTAGGACTGCATTGGC
Rbbp4	F: ACTGTTGCCTTGTGGGATCT R: AGATCCCAGACATTCAGCCT
Rbbp7	F: GGTGGTTGCTCGAGTTCATA R: GTGGCAATGATGTGAGGATTC
Mta1	F: AGAGCCCACTGGTGCTGAAG R: CATCTACCCCATTGTGCTGC
Mta2	F: GCTCGATCCGACTTCCTAAG R: GTCTCATAGGAACGTTTCACC
Mta3	F: TGAGAATTCCTCCAGCAACCC R: TTAGCATGCTTGTCGGCGAG
Gatad2a	F: GCCCAGTGCTGCCAATAATG R: CAGCGGCATGTGAAGTCTGT
Gatad2b	F: TTACAGCAGCAGGCAGCA R: GCCACCTGGCACAGACAATT
Mbd2/3	F: GGAGAAGAACCCTGGTGTGT R: ACAGGCCTTGTCCAGTGGT
Hdac2	F: ATGAGGCTTCATGGGATGAC R: CATGGCGTACAGTCAAGGAG

REFERENCES

Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Baban, A., Pitto, L., Pulignani, S., Cresci, M., Mariani, L., Gambacciani, C., Digilio, M.C., Pongiglione, G., and Albanese, S. (2014). Holt-Oram syndrome with intermediate atrioventricular canal defect, and aortic coarctation: Functional characterization of a de novo TBX5 mutation. Am J Med Genet A.

Basson, C.T., Bachinsky, D.R., Lin, R.C., Levi, T., Elkins, J.A., Soults, J., Grayzel, D., Kroumpouzou, E., Traill, T.A., Leblanc-Straceski, J.*, et al.* (1997). Mutations in human TBX5 [corrected] cause limb and cardiac malformation in Holt-Oram syndrome. Nat Genet *15*, 30-35.

Basson, C.T., Cowley, G.S., Solomon, S.D., Weissman, B., Poznanski, A.K., Traill, T.A., Seidman, J.G., and Seidman, C.E. (1994). The clinical and genetic spectrum of the Holt-Oram syndrome (heart-hand syndrome). N Engl J Med *330*, 885-891.

Basson, C.T., Huang, T., Lin, R.C., Bachinsky, D.R., Weremowicz, S., Vaglio, A., Bruzzone, R., Quadrelli, R., Lerone, M., Romeo, G.*, et al.* (1999). Different TBX5 interactions in heart and limb defined by Holt-Oram syndrome mutations. Proc Natl Acad Sci U S A *96*, 2919-2924.

Bayer, E.A., and Wilchek, M. (1980). The use of the avidin-biotin complex as a tool in molecular biology. Methods Biochem Anal *26*, 1-45.

Benson, D.W., Basson, C.T., and MacRae, C.A. (1996). New understandings in the genetics of congenital heart disease. Curr Opin Pediatr *8*, 505-511.

Berk, A.J. (1999). Activation of RNA polymerase II transcription. Curr Opin Cell Biol *11*, 330-335.

Brown, D.D., Martz, S.N., Binder, O., Goetz, S.C., Price, B.M., Smith, J.C., and Conlon, F.L. (2005). Tbx5 and Tbx20 act synergistically to control vertebrate heart morphogenesis. Development *132*, 553-563.

Bruneau, B.G., Logan, M., Davis, N., Levi, T., Tabin, C.J., Seidman, J.G., and Seidman, C.E. (1999). Chamber-specific cardiac expression of Tbx5 and heart defects in Holt-Oram syndrome. Dev Biol *211*, 100-108.

Bruneau, B.G., Nemer, G., Schmitt, J.P., Charron, F., Robitaille, L., Caron, S., Conner, D.A., Gessler, M., Nemer, M., Seidman, C.E.*, et al.* (2001). A murine model of Holt-Oram syndrome defines roles of the T-box transcription factor Tbx5 in cardiogenesis and disease. Cell *106*, 709-721.

Cline, M.S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., Christmas, R., Avila-Campilo, I., Creech, M., Gross, B.*, et al.* (2007). Integration of biological networks and gene expression data using Cytoscape. Nat Protoc *2*, 2366-2382.

Costantini, J.L., Cheung, S.M., Hou, S., Li, H., Kung, S.K., Johnston, J.B., Wilkins, J.A., Gibson, S.B., and Marshall, A.J. (2009). TAPP2 links phosphoinositide 3-kinase signaling to B-cell adhesion through interaction with the cytoskeletal protein utrophin: expression of a novel cell adhesion-promoting complex in B-cell leukemia. Blood *114*, 4703-4712.

Cristea, I.M., Williams, R., Chait, B.T., and Rout, M.P. (2005). Fluorescent proteins as proteomic probes. Mol Cell Proteomics *4*, 1933-1941.

Cross, S.J., Ching, Y.H., Li, Q.Y., Armstrong-Buisseret, L., Spranger, S., Lyonnet, S., Bonnet, D., Penttinen, M., Jonveaux, P., Leheup, B.*, et al.* (2000). The mutation spectrum in Holt-Oram syndrome. J Med Genet *37*, 785-787.

Denslow, S.A., and Wade, P.A. (2007). The human Mi-2/NuRD complex and gene regulation. Oncogene *26*, 5433-5438.

Diez-Roux, G., Banfi, S., Sultan, M., Geffers, L., Anand, S., Rozado, D., Magen, A., Canidio, E., Pagani, M., Peluso, I.*, et al.* (2011). A high-resolution anatomical atlas of the transcriptome in the mouse embryo. PLoS Biol *9*, e1000582.

Dolk, H., Loane, M., and Garne, E. (2010). The prevalence of congenital anomalies in Europe. Adv Exp Med Biol *686*, 349-364.

Dorr, K.M., Amin, N.M., Kuchenbrod, L.M., Labiner, H., Charpentier, M.S., Pevny, L.H., Wessels, A., and Conlon, F.L. (2015). Casz1 is required for cardiomyocyte G1-to-S phase progression during mammalian cardiac development. Development *142*, 2037-2047.

Driegen, S., Ferreira, R., van Zon, A., Strouboulis, J., Jaegle, M., Grosveld, F., Philipsen, S., and Meijer, D. (2005). A generic tool for biotinylation of tagged proteins in transgenic mice. Transgenic Res *14*, 477-482.

Evans, P.M., and Liu, C. (2008). Roles of Krupel-like factor 4 in normal homeostasis, cancer and stem cells. Acta Biochim Biophys Sin (Shanghai) *40*, 554-564.

Fan, C., Duhagon, M.A., Oberti, C., Chen, S., Hiroi, Y., Komuro, I., Duhagon, P.I., Canessa, R., and Wang, Q. (2003a). Novel TBX5 mutations and molecular mechanism for Holt-Oram syndrome. J Med Genet *40*, e29.

Fan, C., Liu, M., and Wang, Q. (2003b). Functional analysis of TBX5 missense mutations associated with Holt-Oram syndrome. J Biol Chem *278*, 8780-8785.

Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., von Mering, C.*, et al.* (2013). STRING v9.1: protein-protein

interaction networks, with increased coverage and integration. Nucleic Acids Res *41*, D808-815.

Franklin, S., Zhang, M.J., Chen, H., Paulsson, A.K., Mitchell-Jordan, S.A., Li, Y., Ping, P., and Vondriska, T.M. (2011). Specialized compartments of cardiac nuclei exhibit distinct proteomic anatomy. Mol Cell Proteomics *10*, M110 000703.

Garg, V., Kathiriya, I.S., Barnes, R., Schluterman, M.K., King, I.N., Butler, C.A., Rothrock, C.R., Eapen, R.S., Hirayama-Yamada, K., Joo, K.*, et al.* (2003). GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. Nature *424*, 443-447.

Ghosh, T.K., Song, F.F., Packham, E.A., Buxton, S., Robinson, T.E., Ronksley, J., Self, T., Bonser, A.J., and Brook, J.D. (2009). Physical interaction between TBX5 and MEF2C is required for early heart development. Mol Cell Biol *29*, 2205-2218.

Gittenberger-de Groot, A.C., Calkoen, E.E., Poelmann, R.E., Bartelings, M.M., and Jongbloed, M.R. (2014). Morphogenesis and molecular considerations on congenital cardiac septal defects. Ann Med *46*, 640-652.

Goetz, S.C., Brown, D.D., and Conlon, F.L. (2006). TBX5 is required for embryonic cardiac cell cycle progression. Development *133*, 2575-2584.

Hatcher, C.J., Kim, M.S., Mah, C.S., Goldstein, M.M., Wong, B., Mikawa, T., and Basson, C.T. (2001). TBX5 transcription factor regulates cell proliferation during cardiogenesis. Dev Biol *230*, 177-188.

Hayashi, S., Lewis, P., Pevny, L., and McMahon, A.P. (2002). Efficient gene modulation in mouse epiblast using a Sox2Cre transgenic mouse strain. Mech Dev *119 Suppl 1*, S97-S101.

He, A., Kong, S.W., Ma, Q., and Pu, W.T. (2011). Co-occupancy by multiple cardiac transcription factors identifies transcriptional enhancers active in heart. Proc Natl Acad Sci U S A *108*, 5632-5637.

Heinritz, W., Shou, L., Moschik, A., and Froster, U.G. (2005). The human TBX5 gene mutation database. Hum Mutat *26*, 397.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell *38*, 576-589.

Heron, M., and Tejada-Vera, B. (2009). Deaths: leading causes for 2005. Natl Vital Stat Rep *58*, 1-97.

Hiroi, Y., Kudoh, S., Monzen, K., Ikeda, Y., Yazaki, Y., Nagai, R., and Komuro, I. (2001). Tbx5 associates with Nkx2-5 and synergistically promotes cardiomyocyte differentiation. Nat Genet *28*, 276-280.

Hoffmann, A.D., Yang, X.H., Burnicka-Turek, O., Bosman, J.D., Ren, X., Steimle, J.D., Vokes, S.A., McMahon, A.P., Kalinichenko, V.V., and Moskowitz, I.P. (2014). Foxf genes integrate tbx5 and hedgehog pathways in the second heart field for cardiac septation. PLoS Genet *10*, e1004604.

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

Hu, S., Li, L., Yeh, S., Cui, Y., Li, X., Chang, H.C., Jin, J., and Chang, C. (2015). Infiltrating T cells promote prostate cancer metastasis via modulation of FGF11-->miRNA-541-->androgen receptor (AR)-->MMP9 signaling. Mol Oncol *9*, 44-57.

Huang, C.C., Tu, S.H., Lien, H.H., Jeng, J.Y., Huang, C.S., Huang, C.J., Lai, L.C., and Chuang, E.Y. (2013). Concurrent gene signatures for han chinese breast cancers. PLoS One *8*, e76421.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc *4*, 44-57.

Ieda, M., Fu, J.D., Delgado-Olguin, P., Vedantham, V., Hayashi, Y., Bruneau, B.G., and Srivastava, D. (2010). Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. Cell *142*, 375-386.

Kaltenbrun, E., Greco, T.M., Slagle, C.E., Kennedy, L.M., Li, T., Cristea, I.M., and Conlon, F.L. (2013). A Gro/TLE-NuRD corepressor complex facilitates Tbx20-dependent transcriptional repression. J Proteome Res *12*, 5395-5409.

Kehle, J., Beuchle, D., Treuheit, S., Christen, B., Kennison, J.A., Bienz, M., and Muller, J. (1998). dMi-2, a hunchback-interacting protein that functions in polycomb repression. Science *282*, 1897-1900.

Keller, A., Nesvizhskii, A.I., Kolker, E., and Aebersold, R. (2002). Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal Chem *74*, 5383-5392.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol *14*, R36.

Kim, J., Cantor, A.B., Orkin, S.H., and Wang, J. (2009). Use of in vivo biotinylation to study protein-protein and protein-DNA interactions in mouse embryonic stem cells. Nat Protoc *4*, 506-517.

Kimura, R., Kasamatsu, A., Koyama, T., Fukumoto, C., Kouzu, Y., Higo, M., Endo-Sakamoto, Y., Ogawara, K., Shiiba, M., Tanzawa, H.*, et al.* (2013). Glutamate acid decarboxylase 1 promotes metastasis of human oral cancer by beta-catenin translocation and MMP7 activation. BMC Cancer *13*, 555.

Krebs, S., Fischaleck, M., and Blum, H. (2009). A simple and loss-free method to remove TRIzol contaminations from minute RNA samples. Anal Biochem *387*, 136-138.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.

Lassmann, T., Hayashizaki, Y., and Daub, C.O. (2009). TagDust--a program to eliminate artifacts from next generation sequencing data. Bioinformatics *25*, 2839-2840.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Li, Q.Y., Newbury-Ecob, R.A., Terrett, J.A., Wilson, D.I., Curtis, A.R., Yi, C.H., Gebuhr, T., Bullen, P.J., Robson, S.C., Strachan, T.*, et al.* (1997). Holt-Oram syndrome is caused by mutations in TBX5, a member of the Brachyury (T) gene family. Nat Genet *15*, 21-29.

Liu, Z., Yang, X., Li, Z., McMahon, C., Sizer, C., Barenboim-Stapleton, L., Bliskovsky, V., Mock, B., Ried, T., London, W.B.*, et al.* (2011). CASZ1, a candidate tumor-suppressor gene, suppresses neuroblastoma tumor growth through reprogramming gene expression. Cell Death Differ *18*, 1174-1183.

Maine, G.N., Li, H., Zaidi, I.W., Basrur, V., Elenitoba-Johnson, K.S., and Burstein, E. (2010). A bimolecular affinity purification method under denaturing conditions for rapid isolation of a ubiquitinated protein for mass spectrometry analysis. Nat Protoc *5*, 1447-1459.

McDermott, D.A., Bressan, M.C., He, J., Lee, J.S., Aftimos, S., Brueckner, M., Gilbert, F., Graham, G.E., Hannibal, M.C., Innis, J.W.*, et al.* (2005). TBX5 genetic testing validates strict clinical criteria for Holt-Oram syndrome. Pediatr Res *58*, 981-986.

Metz, M., Gassmann, M., Fakler, B., Schaeren-Wiemers, N., and Bettler, B. (2011). Distribution of the auxiliary GABAB receptor subunits KCTD8, 12, 12b, and 16 in the mouse brain. J Comp Neurol *519*, 1435-1454.

Mori, A.D., and Bruneau, B.G. (2004). TBX5 mutations and congenital heart disease: Holt-Oram syndrome revealed. Curr Opin Cardiol *19*, 211-215.

Morris, C. (2010). Epidemiology of congenital heart disease. In Cardiology, D.J. Crawford MH, Paulus WJ, ed. (Philadelphia, Mosby), pp. 1381-1389.

Moskowitz, I.P., Kim, J.B., Moore, M.L., Wolf, C.M., Peterson, M.A., Shendure, J., Nobrega, M.A., Yokota, Y., Berul, C., Izumo, S.*, et al.* (2007). A molecular pathway including Id2, Tbx5, and Nkx2-5 required for cardiac conduction system development. Cell *129*, 1365-1376.

Moskowitz, I.P., Pizard, A., Patel, V.V., Bruneau, B.G., Kim, J.B., Kupershmidt, S., Roden, D., Berul, C.I., Seidman, C.E., and Seidman, J.G. (2004). The T-Box transcription factor Tbx5 is required for the patterning and maturation of the murine cardiac conduction system. Development *131*, 4107-4116.

Polo, S.E., Kaidi, A., Baskcomb, L., Galanty, Y., and Jackson, S.P. (2010). Regulation of DNA-damage responses and cell-cycle progression by the chromatin remodelling factor CHD4. EMBO J *29*, 3130-3139.

Qian, L., Huang, Y., Spencer, C.I., Foley, A., Vedantham, V., Liu, L., Conway, S.J., Fu, J.D., and Srivastava, D. (2012). In vivo reprogramming of murine cardiac fibroblasts into induced cardiomyocytes. Nature *485*, 593-598.

Roesli, C., Neri, D., and Rybak, J.N. (2006). In vivo protein biotinylation and sample preparation for the proteomic identification of organ- and disease-specific antigens accessible from the vasculature. Nat Protoc *1*, 192-199.

Saunders, A., Werner, J., Andrulis, E.D., Nakayama, T., Hirose, S., Reinberg, D., and Lis, J.T. (2003). Tracking FACT and the RNA polymerase II elongation complex through chromatin in vivo. Science *301*, 1094-1096.

Sedmera, D., and Thompson, R.P. (2011). Myocyte proliferation in the developing heart. Dev Dyn *240*, 1322-1334.

Smallwood, P.M., Munoz-Sanjuan, I., Tong, P., Macke, J.P., Hendry, S.H., Gilbert, D.J., Copeland, N.G., Jenkins, N.A., and Nathans, J. (1996). Fibroblast growth factor (FGF) homologous factors: new members of the FGF family implicated in nervous system development. Proc Natl Acad Sci U S A *93*, 9850-9857.

Sopher, B.L., Koszdin, K.L., McClain, M.E., Myrick, S.B., Martinez, R.A., Smith, A.C., and La Spada, A.R. (2001). Genomic organization, chromosome location, and expression analysis of mouse beta-synuclein, a candidate for involvement in neurodegeneration. Cytogenet Cell Genet *93*, 117-123.

Stirnimann, C.U., Ptchelkine, D., Grimm, C., and Muller, C.W. (2010). Structural basis of TBX5-DNA recognition: the T-box domain in its DNA-bound and -unbound form. J Mol Biol *400*, 71-81.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc *7*, 562-578.

Trifonov, S., Yamashita, Y., Kase, M., Maruyama, M., and Sugimoto, T. (2014). Glutamic acid decarboxylase 1 alternative splicing isoforms: characterization, expression and quantification in the mouse brain. BMC Neurosci *15*, 114.

van Werven, F.J., and Timmers, H.T. (2006). The use of biotin tagging in Saccharomyces cerevisiae improves the sensitivity of chromatin immunoprecipitation. Nucleic Acids Res *34*, e33.

Wade, P.A., Jones, P.L., Vermaak, D., and Wolffe, A.P. (1998). A multiple subunit Mi-2 histone deacetylase from Xenopus laevis cofractionates with an associated Snf2 superfamily ATPase. Curr Biol *8*, 843-846.

Wang, J., Rao, S., Chu, J., Shen, X., Levasseur, D.N., Theunissen, T.W., and Orkin, S.H. (2006). A protein interaction network for pluripotency of embryonic stem cells. Nature *444*, 364-368.

Wang, C., Cao, D., Wang, Q., and Wang, D.Z. (2011). Synergistic activation of cardiac genes by myocardin and tbx5. PLoS One *6*, e24242.

Wang, S., Sun, H., Ma, J., Zang, C., Wang, C., Wang, J., Tang, Q., Meyer, C.A., Zhang, Y., and Liu, X.S. (2013). Target analysis by integration of transcriptome and ChIP-seq data with BETA. Nat Protoc *8*, 2502-2515.

Wang, W., Kim, H.J., Lee, J.H., Wong, V., Sihn, C.R., Lv, P., Perez Flores, M.C., Mousavi-Nik, A., Doyle, K.J., Xu, Y*., et al.* (2014). Functional significance of K+ channel beta-subunit KCNE3 in auditory neurons. J Biol Chem *289*, 16802-16813.

Xue, Y., Wong, J., Moreno, G.T., Young, M.K., Cote, J., and Wang, W. (1998). NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. Mol Cell *2*, 851-861.

Ying, C.Y., Dominguez-Sola, D., Fabi, M., Lorenz, I.C., Hussein, S., Bansal, M., Califano, A., Pasqualucci, L., Basso, K., and Dalla-Favera, R. (2013). MEF2B mutations lead to deregulated expression of the oncogene BCL6 in diffuse large B cell lymphoma. Nat Immunol *14*, 1084-1092.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W*., et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.

APPENDIX 2: The Lhx9-Integrin pathway is essesntial for positioning of the proepicardial organ[3]

Introduction

Cardiovascular disease is one of the highest causes of mortality and reduced quality of life in the US (Roger et al., 2012). To better optimize therapeutics to preserve and maintain heart physiology it is pertinent that we understand the mechanisms involved in all aspects of heart development, growth and repair. This will further enable us to target resident cardiac cell populations and gene programs for clinical applications.

The heart initially forms in vertebrates as a bilaminar tube comprised of an inner endocardium and outer myocardial layer. At later stages of development a third layer is added to the heart termed the epicardium. In many vertebrates this layer forms from a dynamic precursor cell population, the proepicardial (PE) organ, which arises as a cluster of cells on the right side of the septum transversum adjacent to the heart (Schulte et al., 2007, Jahr et al., 2008, Manner, 1992, Pombal et al., 2008, Serluca, 2008). Whilst FGF and BMP signaling, as well as the transcription factors Snai1 and Twist, have been implicated in PE induction (Schlueter and Brand, 2009, Schulte et al., 2007, Schlueter and Brand, 2013), there is little known about the mechanisms that direct the formation of the PE cell cluster. Once the cluster has formed and is positioned correctly it is induced to bridge towards the developing heart, directed by BMP signaling

[3]This appendix previously appeared as an article in the journal *Development*. The original citation is as follows: Tandon P, Wilczewski CM, Williams CE, Conlon FL. “The Lhx9-integrin pathway is essential for positioning of the proepicardial organ,” *Development* no. 143 (March 2016): 5.

from the atrioventricular sulcus (AVS) (Ishii et al., 2010). Upon attachment, epicardial cells then traverse over the heart surface as an epithelial-like sheet termed the epicardium.

The epicardial layer is an essential cardiac component as it provides mitogenic stimulation to the developing myocardium (Kang et al., 2008, Lavine et al., 2005, Li et al., 2011, Pennisi et al., 2003, Stuckmann and Lassar, 2002). In addition, the epicardium is the source of epicardial-derived cells (EPDCs) whereby subpopulations of this layer undergo epithelial-mesenchymal transition and invade the underlying myocardium. Here they differentiate into vital cardiac cells such as smooth muscle cells and pericytes of the vasculature and fibroblasts that contribute to the scaffold and support of the heart (Gittenberger-de Groot et al., 2010, Lie-Venema et al., 2007, Manner et al., 2001, Ratajska et al., 2008, von Gise and Pu, 2012). Current understanding is that EPDC lineage specificity is determined early within the precursor PE cell cluster prior to epithelial-mesenchymal transition (EMT) (Acharya et al., 2012, Guadix et al., 2006, Manner, 1999, Mikawa and Gourdie, 1996, Katz et al., 2012, Wei et al., 2015). In the event of a cardiac infarction or damage in the adult, research has also demonstrated the utility of the epicardium in providing regenerative support and cardiac cells to the damaged tissue (Lepilina et al., 2006, Limana et al., 2010, Winter and Gittenberger-de Groot, 2007, Winter et al., 2009). Therefore, better comprehension of how this source of pluripotent cells is specified and develops is vitally important to understanding mechanisms of cardiac development, disease and repair.

Here we report that LIM Homeobox transcription factor 9 (Lhx9) is essential for the formation and function of the epicardium. We show that loss of Lhx9 in *Xenopus*

leads to defects in the assembly of the PE cluster and subsequently failed migration and spreading of epicardial cells onto the surface of the developing heart. Using a small pharmacological molecule screen in *Xenopus* we ascertained that Integrin-Paxillin signaling is required for proper PE clustering, which is confirmed by a significant decrease in epicardial *integrin alpha 4* (*itga4*) expression upon loss of Lhx9 function. Given that interactions between cells and the surrounding ECM are vital for epicardial formation and its role in cardiac repair (Kalman et al., 1995, Nahirney et al., 2003, Pae et al., 2008, Wang et al., 2013, Benesh et al., 2013, Fransen and Lemanski, 1991, Mercer et al., 2013, Rongish et al., 1996, Yang et al., 1995), we further show a novel function for Lhx9 in mediating Integrin-adhesion mechanisms for correct PE cell clustering. Therefore Lhx9 is vital for the formation of the epicardium and development of the heart.

Results

Lhx9 is expressed in a temporally dynamic pattern during epicardial formation.

Animals lacking a functional epicardium die prematurely and display hemorrhaging hearts with thin ventricular walls (Moore et al., 1998, Lie-Venema et al., 2005, Gittenberger-de Groot et al., 2000, Acharya et al., 2012, Combs et al., 2011). Many vertebrates show a right side clustering of PE cells during epicardial formation (Viragh et al., 1993, Kalman et al., 1995, Nahirney et al., 2003, Serluca, 2008, Pombal et al., 2008, Jahr et al., 2008). Recently we have demonstrated that the LIM homeodomain transcription factor family *lhx9* is transiently expressed in the *Xenopus* PEO (Tandon et al., 2013). We therefore conducted a detailed analysis to establish the

relationship between Lhx9 expression and formation of the PEO on the septum transversum.

The sequence and structure of Lhx9 has been shown to be highly conserved across vertebrates (Bertuzzi et al., 1999, Retaux et al., 1999, Failli et al., 2000, Bachy et al., 2001, Ottolenghi et al., 2001, Alunni et al., 2007, Oshima et al., 2007, Peukert et al., 2011) being comprised of two LIM protein binding domains and a DNA-binding homeodomain (HD). Consistent with other vertebrates we identified two isoform variants of *lhx9* in *Xenopus*, one with a full length HD (*lhx9HD*) and one with a truncated HD (*lhx9α*) generated by alternative splicing of exon 5 (Failli et al., 2000, Molle et al., 2004) (Fig. SA2.1).

To ascertain the endogenous role of Lhx9 during epicardial development we performed a detailed spatio-temporal analysis of both *lhx9* isoforms. Results from these studies show *lhx9α* is initially expressed bilaterally on the septum transversum prior to PEO formation (Fig. A2.1A, E, Fig. SA2.2-SA2.4). In all vertebrates, the lateral confinement of the PEO cells to the right side of the embryo can be observed by the expression of the conserved PEO marker *tbx18* (Fig. A2.1K-N, Fig. SA2.3) (Tandon et al., 2013). Consistently we observed a dramatic restriction of *lhx9α* expression to the right side of the septum transversum (stages 39-46) (Fig. A2.1B-D, F) and subsequently *lhx9α* is down regulated, relative to *tbx18*, as the PEO cells mature and migrate onto the heart (Fig. A2.1D, N, Fig. SA2.3) (Tandon et al., 2013). An identical pattern of expression was obtained with an independent probe recognizing the LIM domains of *lhx9α* and *lhx9HD* (Fig.A2.1E, F, Fig. SA2.1, Fig. SA2.4, Fig. SA2.11). We further observed a highly similar pattern of expression for *lhx9HD* in the septum transversum

(Fig. A2.1G-J, Fig. SA2.5) however we note *lhx9HD* is more weakly expressed and dramatically down regulated upon PE clustering compared to *lhx9α* (Fig. A2.A1I, Fig. SA2.5). We also note we did not observe detectable expression of *lhx2*, a closely related LIM-HD transcription factor to Lhx9, in the developing epicardium at any stage (Fig. SA2.6) (Viczian et al., 2006). Taken together these studies demonstrate *lhx9α* and *lhx9HD* isoforms have overlapping dynamic expression during epicardial development and *lhx9α* is the predominant isoform expressed during PEO formation.

Lhx9 is required for epicardial formation.

To determine the requirement for Lhx9 in the vertebrate PEO we depleted Lhx9 in *Xenopus laevis* (*X. laevis*) using Morpholinos (see Supplementary Methods for details).

Depletion of Lhx9 function in *Xenopus* does not impede early embryonic development. However, analysis of the developing heart at tadpole stages revealed a severe epicardial defect. Whilst control embryos displayed PE cells clustering to the AVS on the right side of the heart and a layer of epicardial cells on the ventricle surface (Fig. A2.2A, C, red arrowheads marking PE cell cluster, *tbx18* in situ hybridization), embryos lacking Lhx9 failed to accumulate PE cells to the right-side AVS and displayed a reduced and discontinuous epicardial layer on the ventricular surface (Fig. A2.2B, D, E, F, red arrowheads).

Lhx9 is required for clustering of proepicardial cells.

Having demonstrated proper epicardial formation requires Lhx9 function and because *lhx9* expression is detected on the septum transversum and PE cluster (Fig. A2.1A-F, Fig. SA2.2), we assessed whether Lhx9 functions during the earlier stages of epicardial specification. At stage 38 the conserved pan-marker of the epicardial lineage, *tbx18* was observed on the septum transversum and deemed indistinguishable between control and Lhx9-depleted embryos (Fig. A2.2G, H). Therefore Lhx9 is not required for specification of the epicardial lineage. However, by stage 41 when PE cells begin to cluster to the right side of embryos (Fig. A2.2I, red arrowhead), Lhx9-depleted stage-matched sibling embryos maintained diffuse *tbx18* expression bilaterally along the septum transversum (Fig. A2.2J, green arrowheads, Fig. SA2.9A). By stage 42, the PE cluster had attached to the AVS on the right side of the heart in controls whereas in Lhx9-depleted embryos there was either no detectable cluster and *tbx18* expression persisted on the anterior endoderm/septum transversum or the cluster was mispositioned caudal to the heart (Fig. A2.2L, M, N).

Consistently, we observed the spatial distribution of two independent epicardial markers, *tcf21* and *wt1*, were similarly disrupted in Lhx9-depleted embryos. Both *tcf21* and *wt1* expression marks the punctate cluster of PE cells at stage 41 on the right side of the embryo (Fig. A2.2O, Q, S, red arrowhead), whilst expression of these epicardial markers is reduced and mislocalized in Lhx9-depleted embryos (Fig. A2.2P, R, T, U, red arrowheads, Fig. SA2.9B). This data collectively implies Lhx9 function is critical for normal clustering of PE cells during early stages of development and in its absence epicardial formation is disrupted.

Proepicardial cell cluster positioning is driven by Integrin-mediated mechanisms.

The cellular and molecular mechanisms which lead to the clustering of the PEO on the right side of the septum transversum remain entirely unknown. To assess whether active cell migration is required for asymmetric PE clustering we conducted a screen in wild type embryos with a defined bank of small molecule inhibitors known to disrupt cell migration and adhesion in *Xenopus* (Harata et al., 2013, Broders-Bondon et al., 2007). Whilst the inhibition of myosin II (Blebbistatin) resulted in tadpoles with severe pericardial edema (possibly due to reduced heart beat) and paralysis, we did observe clustered PE cells (marked by *tbx18*) to the right of the embryo, albeit due to the pronounced edema we were unable to determine attachment to the heart (data not shown). Inhibition of Rac1 activity (NSC23766 trihydrochloride) which has been extensively shown to disrupt cell migration (Diaz et al., 2014, Huang et al., 2013) did not affect PE cell clustering compared to controls as depicted by *tbx18* expression on the AVS (Fig. A2.3A, B, G-J). Little *tbx18* expression was preserved on the septum transversum in these embryos implying PE clustering is not dependent on classical cell migration mechanisms involving actin-myosin interactions and GTPase activity.

Interestingly, exposure from stages 38-41 when PE cell clustering is occurring, to a small molecule that disrupts Integrin-Paxillin interactions (6-B345TTQ) led to a PE cluster positioning defect phenocopying loss of Lhx9 function. Integrin binds to ECM components and induces the recruitment of proteins, such as Paxillin, to focal adhesions at the site of cell attachment. These protein complexes then function to reorganize the actin cytoskeleton and are essential for the coordinated adhesion and

motility of cells through an ECM environment (Bellis et al., 1995). Paxillin is a multi-LIM domain protein has been shown to directly and specifically bind to the cytoplasmic tail of Itgα4 to coordinate cell spreading, adhesion and migration (Liu et al., 1999, Alon et al., 2005). Whilst some clustering of PE cells was evident in cultured tadpoles, we observed mispositioning of the cluster to the caudal side of the heart as well as persistent *tbx18* expression on the septum transversum (Fig. A2.3C-F, K) as seen in Lhx9-depleted embryos (Fig. A2.2H, J, L, M). This data therefore suggests Integrin-mediated mechanisms plays an important role in the correct positioning of the PEO cell cluster in developing embryos.

Lhx9 is required for proepicardial clustering via an Integrin-Paxillin interaction.

Our data demonstrate Lhx9 as well as Integrin-mediated cellular functions are essential for the correct clustering and positioning of PE cells. We previously reported *lhx9* and *itgα4* are co-regulated in response to loss of the epicardial transcription factor Tcf21 (Tandon et al., 2013). Based on these observations we hypothesized Lhx9 acts upstream of *itgα4* to correctly position the PEO. Consistent with this hypothesis we find *itgα4* expression is detected subsequent to *lhx9* in the clustered PE cells from stage 39-46 (Fig.A2.4A-D, red arrowheads, Fig. SA2.3). We thus assessed the expression of *itgα4* in Lhx9-depleted hearts. Consistent with *itgα4* acting downstream of Lhx9 we find *itgα4* is reduced and misplaced in Lhx9-depleted embryos (Fig. A2.4E, F, O, Fig. SA2.8, Fig. SA2.9C). By contrast, the expression of Integrin β1 (Itgβ1) is strongly localized to the epicardium (as marked by *tcf21*) and appears to be independent of Lhx9 function (Fig. A2.4G-J', Fig. SA2.11) therefore enabling its use as a marker of epicardial cells.

Due to the clustering defects observed upon reducing the Itgα4-Paxillin interaction biochemically (Fig. A2.3) and reduced *itgα4* expression in Lhx9-depleted embryos, we sought to determine if Lhx9 is responsible for Integrin-mediated cell adhesion and migration processes in PE cells. To address this possibility we assayed for phosphorylated Paxillin (Y118) which has been used to demonstrate Integrin activation at focal adhesions and shown to modulate cell spreading, migration and invasion (Bellis et al., 1995, Burridge et al., 1992, Turner, 1991, Nakamura et al., 2000, Lewis and Schwartz, 1998, Zaidel-Bar et al., 2007, Sachdev et al., 2009). Our data indicated a significant decrease in phosphorylated Y118 Paxillin expression specifically in the attached PE cluster in Lhx9-depleted embryos compared to controls (Fig. A2.4K-N', P). Taken together these data strongly imply Lhx9 acts through Itgα4 to correctly position the PEO.

Lhx9-regulated Integrin signaling is essential for correct formation of the epicardial layer.

Collectively our data demonstrates a role for Lhx9-Itgα4 in the correct positioning and clustering of the PEO. Our data also suggests Integrin activity at focal adhesions is required for the clustering of PE cells. Previous data has implicated Itgα4 as being essential for epicardial development particularly in regulating PE cell clustering, epicardial adhesion and epicardial migration (Sengbusch et al., 2002, Pinco et al., 2001, Yang et al., 1995, Dettman et al., 2003, Dokic and Dettman, 2006, Hirose et al., 2006, Pae et al., 2008, Bax et al., 2010). Integrin heterodimers have been shown to coordinate cytoskeletal interactions and adhesion with the surrounding ECM

environment (Gardiner, 2011, Gehler et al., 2013, Huttenlocher and Horwitz, 2011, Manninen, 2015, Wolfenson et al., 2013), pertinent processes for epicardial formation and function (Kalman et al., 1995, Nahirney et al., 2003).

To determine if the depleted adhesive properties of the PEO and its mis-positioning have persistent biological consequence in the development of the vertebrate epicardium and heart, we examined the epicardial ECM environment of control and Lhx9-depleted hearts. Using Itgβ1 as a marker of epicardial cells we analyzed the expression of known Itgα4 ligands and ECM components. Whilst we observed no discernible alterations in the expression of the Itgα4-Itgβ1 (Very Late Antigen 4, VLA-4) ligand *vascular cell adhesion molecule 1* (*vcam1*, Fig.S10) in Lhx9-depleted embryos, we did detect a decrease in Fibronectin within the PE cluster as well as in the heart relative to controls (Fig. A2.5A-E) (Humphries et al., 1995, Wu et al., 1995). This suggests alterations to Integrin function in response to loss of Lhx9 is specific to Itgα4 pathway components. In addition these data further implies the adhesive qualities and positioning of the PEO is critical for the proper deposition of epicardial ECM.

To address this further we examined the localization of epicardial-ECM markers Laminin and β-dystroglycan. In control embryos we observed weak Laminin deposition surrounding attached PE cells (Fig. A2.5F, F', white arrowheads) but a continuous smooth layer of Laminin was present under the mature epicardial layer (Fig. A2.5F", white arrowheads). However in Lhx9-depleted embryos, when epicardial cells were present on the myocardial surface, we observed an increased accumulation of Laminin in the basal portion of attached PE cells (Fig. A2.6G, G', white arrowheads) as well as discontinuous deposits within epicardial cells on the surface of the heart (Fig. A2.5G",

white arrowheads). To further confirm these findings, we analyzed β-dystroglycan, a transmembrane glycoprotein that attaches cells to the basement membrane through Laminin binding (Ervasti and Campbell, 1993, Klietsch et al., 1993, Smalheiser and Schwartz, 1987). In Lhx9-depleted embryos β-dystroglycan is reduced or disrupted at the point of contact between the PE cluster and heart surface (as indicated by *tcf21*) (Fig. A2.5J-K') compared to controls (Fig. A2.5H-I'). Note β-dystroglycan is prominent in the endocardium of both control and Lhx9-depleted hearts showing this defect is specific to the epicardium (Fig. A2.5I', K', yellow arrowheads). Taken together these results demonstrate PE cell adhesive integrity is required for the proper formation of the epicardial layer.

Collectively, our data are consistent with a role for Lhx9 in the clustering and positioning of the PEO (Fig. A2.6A-B). Our data further show the ability of Lhx9 to regulate Itgα4 signaling is vital for the correct establishment of clustering, attachment and migration of PE cells to the heart (Fig, A2.6A, C, D). These signals are in turn necessary for the correct deposition of ECM components thus, facilitating epicardial cells to form a cohesive layer over the myocardial surface (Fig. A2.6E, F).

Discussion

Our data collectively shows Lhx9 is essential for proepicardial positioning and that loss of Lhx9 is associated with aberrant deposition of essential ECM components during crucial stages of epicardial clustering and migration. We further observe that loss of Lhx9 correlates with a decrease in Integrin-ECM signaling and the ability of PE cells to form cohesive adhesions downstream of Integrin activity. Because the PE cluster

defect in Lhx9-depleted embryos closely resembles the defect caused by disturbing the Itgα4-Paxillin interaction biochemically, we propose and demonstrate Lhx9 functions to modulate *itgα4* expression. Taken together this data establishes a role for Lhx9 in Integrin-mediated cell adhesion and motility crucial for the proper positioning and clustering of PE cells.

Though most vertebrates show a right side clustering of PE cells during epicardial formation (Viragh et al., 1993, Kalman et al., 1995, Nahirney et al., 2003, Serluca, 2008, Pombal et al., 2008, Jahr et al., 2008) it remains to be established if the sidedness of epicardial formation is essential for its function. In our studies we further find that alterations in the positioning of the epicardium on the septum transversum is essential for the correct attachment of the epicardium to the heart at the AVS. Thus, these imply that alterations in PE clustering predispose an embryo to cardiac defects.

Itga4 and epicardial development

The function of Itgα4 during epicardial development has been characterized in mice mutants who displayed absence of the epicardial layer and lack of coronary vessels at the AVS (Yang et al., 1995). This was shown to be due to defects in the migratory process (Sengbusch et al., 2002) and as demonstrated by this work. Therefore the role of Itgα4 in epicardial migration appears to be conserved regardless of the mode of epicardial formation; cluster bridge in frog or free-floating cysts in mouse. Our data provides new evidence that the expression of *itga4* is restricted to immature PE cells and downregulated as the epicardial layer matures, similar to the hematopoietic lineage (Arroyo et al., 1999, Lobb and Hemler, 1994). Using Itgβ1 as a

marker of epicardial cells we analyzed the expression of known Itgα4 ligands and ECM components. We report a down regulation of Fibronectin in embryos lacking Lhx9 but observe no alterations in expression of *vcam1*, the Itgα4-Itgβ1 ligand. Taken together these findings demonstrate that loss of Lhx9 functions upstream of *itgα4* and that the function of Lhx9 is specific to Itgα4 function.

Lhx9 and Tcf21 during epicardial formation

Our previous studies identified *lhx9* and *itgα4* as genes upregulated upon Tcf21 loss. The role of Lhx9 in cardiac development to date has been limited to studies on the effect of its persistent expression upon FOG2 ablation (Smagulova et al., 2008). Whilst these studies provide insight into the effects of potentially prolonged Lhx9 activity, the requirement for Lhx9 has not been addressed. Consistently with our finding on Tcf21, we observe here that depletion of Lhx9 results in decreased *itgα4* with a sparse and disorganized epicardial layer on the heart. Given their similar patterns of expression in the developing epicardium, this data suggests therefore that Tcf21 and Lhx9 function coordinately together to enable the correct clustering and migration of PE cells onto the heart surface through Integrin-directed mechanisms. It is of interesting note that Tcf21 and LIM homeodomain transcription factors are both found to be essential for the development of the gonad and craniofacial structures (Moncaut et al., 2012, Harel et al., 2012, Cui et al., 2004, Birk et al., 2000) suggesting that the transcriptional targets of these two genes and the pathways they regulate may unveil a conserved mechanism into regulating cell motility and thus, have implications in cancer as well as cardiac and organ development (Vladimirova et al., 2009, Yang et al., 2014, Miller et al., 2014,

Weiss et al., 2013, Miller et al., 2013, Zhang et al., 2012, Ye et al., 2012, Richards et al., 2011, Arab et al., 2011).

Materials and Methods

In situ hybridization

Whole-mount in situ hybridization (ISH) was carried out as previously described (Tandon et al., 2013, Harland, 1991, Charpentier et al., 2015), the pericardial cavity membrane in late tadpole stage embryos being removed post-fixation to improve resolution. Embryos were processed for gelatin (30 µm) sectioning as previously reported (Gessert and Kuhl, 2009, Tandon et al., 2013, Charpentier et al., 2015). All probes were previously used (Tandon et al., 2013) or generated by PCR cloning from *Xenopus* embryonic cDNA (Supplemental Table 1). Information on constructs is available upon request. All phenotypes were quantified and statistically assessed using Fishers Exact test or Chi-Square (Graphpad Prism 6). Figure images show representative phenotypes quantified. For spatiotemporal analysis of *lhx9, lhx2, tbx18* and *itga4* transcripts in whole wild type embryos (Supplemental Figs. 2-6), multiple focal plane images were compiled using auto-blend function (Photoshop CS4). Original images are available upon request.

Xenopus manipulations

Xenopus embryos were cultured, microinjected and staged according to Niewkoop and Faber (Nieuwkoop and Faber, 1967, Tandon et al., 2013). *Xla.Tg(Cardiac-actin:GFP)*^{Mohun} transgenic frogs were used as previously reported

(Latinkic et al., 2002, Tandon et al., 2013). The 5'UTR from both short and long *X. laevis lhx9* alloalleles was cloned and sequenced to verify design of the translation-blocking Morpholino (MOT). A 5-base mismatch MO was used as a control (Supplemental Table 2). Two transcription-blocking Morpholinos were designed against the splice donor-site of exon 1 from both *lhx9* alloallele genomic loci (MO1, Table 2) (Tandon et al., 2012) (Genetools). 30 ng of each MO was injected at the one-cell stage, with both MO1 being co-injected (Tandon et al., 2013, Tandon et al., 2012). To assess the specificity and efficiency of MOT, 1ng *lhx9*-5'UTR-GFP RNA (designed against each genomic loci) was co-injected together with the stated concentrations of MO. Stage 11 embryos were collected, lysed and analyzed using western-blot as previously performed (Tandon et al., 2012, Tandon et al., 2013) using GFP antibody (JL8, #632381, Clontech, 1:10,000) and Shp2 antibody (#610622, BD Transduction, 1:2500) as a loading control. To assess the efficiency of MO1, embryos were injected at the one-cell stage, and the hearts from 25 embryos collected at stage 42 from each condition. RT-PCR was performed (Superscript II, Invitrogen) for *lhx9* with *gapdh* as a loading control (Supplemental Table 1). See also Supplemental Methods.

Immunohistochemistry

Embryos were processed for agarose vibratome sectioning (100 µm) as reported (Wallingford, 2010, Tandon et al., 2013, Charpentier et al., 2015). Sections were then washed with PBS + 1% Triton X-100, blocked with wash buffer + 10% fetal calf serum and incubated with primary antibody as previously reported (Langdon et al., 2012, Tandon et al., 2013, Charpentier et al., 2015). Antibodies included mouse anti-

Tropomyosin (CH1, DSHB 1:25), rabbit anti-Laminin (L9393, Sigma, 1:100), mouse anti- β-dystroglycan (MANDAG2, 7D11, DSHB, 1:100), rabbit anti-Fibronectin (F3648, Sigma, 1:250), rabbit anti-phospho-Paxillin pTyr118 (44-722G, Invitrogen, 1:100) and mouse anti-Integrin β1 (8C8, DHSB, 1:100). CH1 was deposited to the DSHB by Lin J.J.-C, MANDAG2(7D11) by Morris G.E and 8C8 by Hausen P/ Gawantka V. Sections were then incubated in DAPI/PBS (200 ng/ml) and imaged using a Zeiss LSM700 Spectral Confocal Laser Scanning Microscope, with representative figures compiled using ImageJ and Photoshop. Image fluorescence, gauged by integrated pixel density, was measured using ImageJ and quantified with standard two-tailed student's t-test (Graphpad).

Live tadpole small molecule culture assay

Xenopus wild type embryos were incubated in the appropriate concentration of small molecule or DMSO control in 0.1 X MBS from stage 38 to stage 41 at room temperature in the dark. Blebbistatin (B0560, Sigma) was used at 10 µM (Straight et al., 2003), 6-B345TTQ (B7438, Sigma) was used at 5 and 10 µM (Kummer and Ginsberg, 2006, Kummer et al., 2010, Hung et al., 2013) and NSC23766 trihydrochloride (SML0952, Sigma) was used at 5 and 50 µM. Embryos were briefly rinsed in 0.1x MBS, fixed in 4% PFA and subjected to in situ hybridization as described.

Funding

We are thankful for funding from the National Institutes of Health (NICHD, R21HD073044 to FLC) and the American Heart Association (13POST16950044, postdoctoral fellowship to PT).

Author Contributions

PT developed the concepts, performed experiments and analyzed data for the manuscript as well as prepared and edited the paper prior to submission. CMW and CEW performed experiments and phenotypic analysis, FLC prepared and edited manuscript.

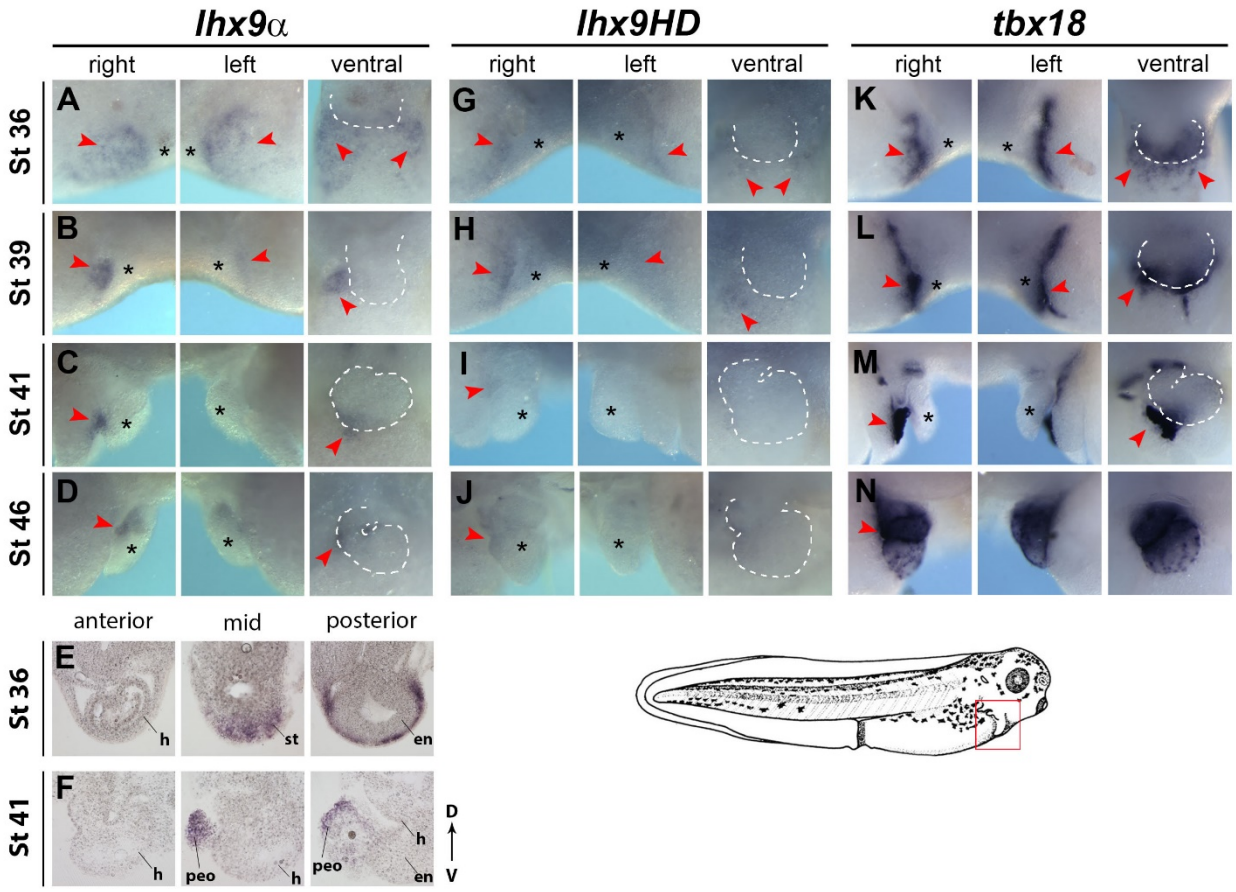


Figure A2.1. Spatio-temporal analysis of *lhx9* isoforms during *Xenopus* epicardial development. Whole embryo in situ hybridization was performed on *Xenopus* embryos at stages 36 to 46, using probes specific for *lhx9a* (A-D), *lhx9LIM* (E, F), *lhx9HD* (G-L) and *tbx18* (K-N) (See Fig. S2, S3, S4, S5). Views display embryos facing right, left and ventral with anterior to the top. Images show magnified cardiac regions as depicted by red box on tadpole schematic (bottom). Red arrowheads depict staining of septum transversum and PE clusters. Cardiac region is depicted with *. White dashed line outlines heart in ventral views. (E, F) Transverse gelatin section in situ hybridization images of cardiac region demonstrate *lhx9LIM* expression exclusively in the septum transversum (stage 36) and proepicardial tissue (stage 41). en, endocardium; h, heart; peo, proepicardial organ; st, septum transversum.

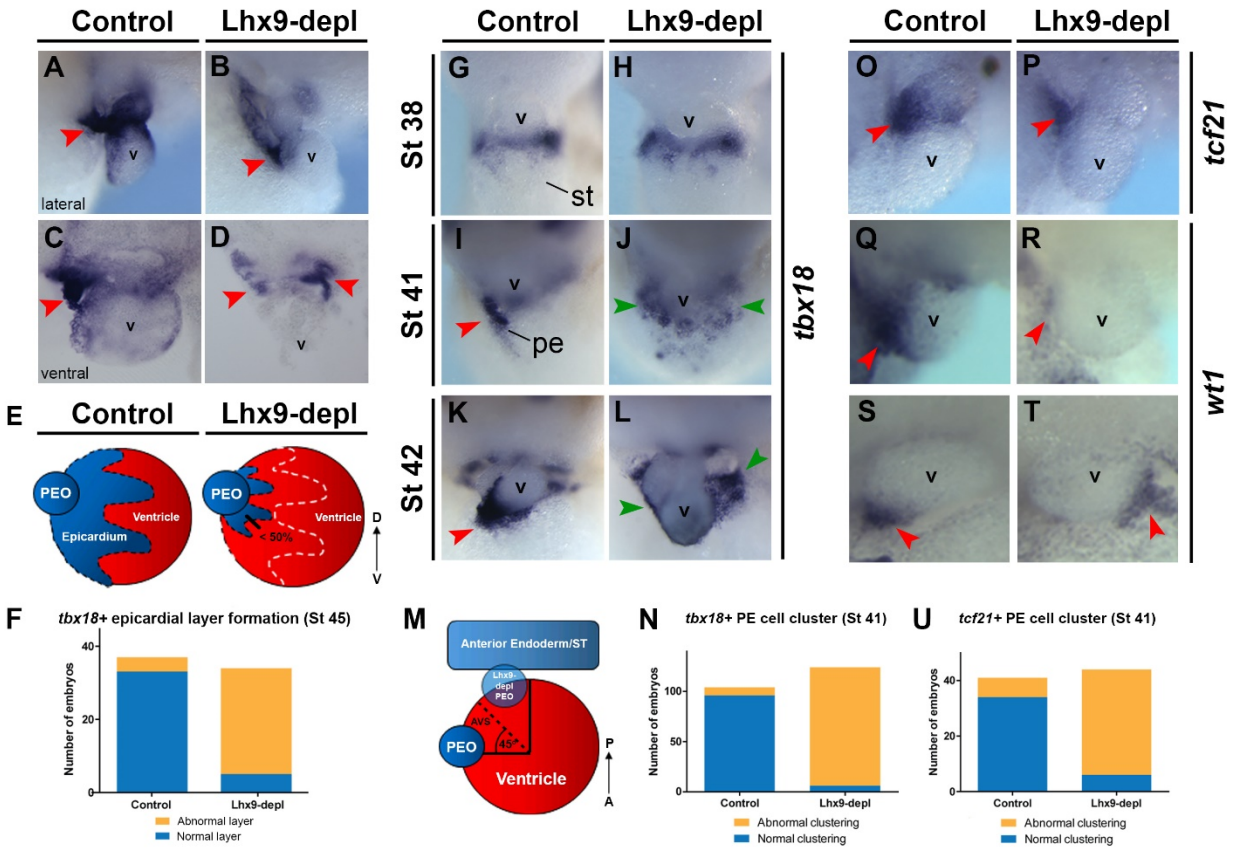


Figure A2.2. Lhx9 is required for proper epicardial layer and PE cluster formation.

Epicardial layer formation (as shown with *tbx18*) analysis in control (A, C) and Lhx9-depleted (B, D) embryos at stage 45. (A, B) Images show lateral view of cardiac region, dorsal to the top and anterior to right. (C, D) Transverse gelatin sections through representative embryo cardiac regions showing *tbx18* expression, dorsal to the top. PE cell clustering (red arrowheads) and epicardial cell layer on the ventricular surface is shown. (E) Epicardial layer formation defects were quantified as abnormal as ≤ 50% ventricular coverage as depicted in schematic. (F) Quantification of observed epicardial layer formation defects, as represented in A-D, from three independent experiments, p = <0.0001 by two-tailed Fisher's exact test. Proepicardial clustering morphology was analyzed using in situ hybridization for *tbx18* in control (G, I, K) and Lhx9-depleted (H, J,

L) embryos at stage 38 (G, H), stage 41 (I, J) and stage 42 (K, L). Ventral view showing cardiac region, anterior to the top. *Tbx18* expression is detected as a cluster of cells in control embryos on the right of the embryo near the atrioventricular sulcus (I, K, red arrowheads) whereas in Lhx9-depleted embryos, the cluster is either not detected and *tbx18* expression remains throughout the septum transversum region or is mis-positioned to the caudal side of the heart (abnormal clustering) (J, L, green arrowheads). (M) Quantification of PE clustering defects is depicted in the schematic as being abnormal by bilateral *tbx18* expression retention on septum transversum and/or cluster mis-positioning of ≥ 45^{o} caudal to the AVS compared to controls. (N) Quantification of observed clustering phenotype at stage 41 represented in C and D. Data taken from seven independent experiments, p = <0.0001 by two-tailed Fisher's exact test. (O-T) In situ hybridization analysis for *tcf21* (O, P, U) and *wt1* (Q-T) proepicardial expression at stage 41. Images depict lateral view of the cardiac region, dorsal to the top and anterior to the right (M-P) or ventral view, dorsal to top (Q, R). Red arrowheads mark clustered PE cells. (U) Quantification of observed phenotypes as depicted in M, data taken from six independent experiments, p = <0.0001 by two-tailed Fisher's exact test. avs, atrioventricular sulcus; pe/peo; proepicardial organ, st; septum transversum, v; ventricle.

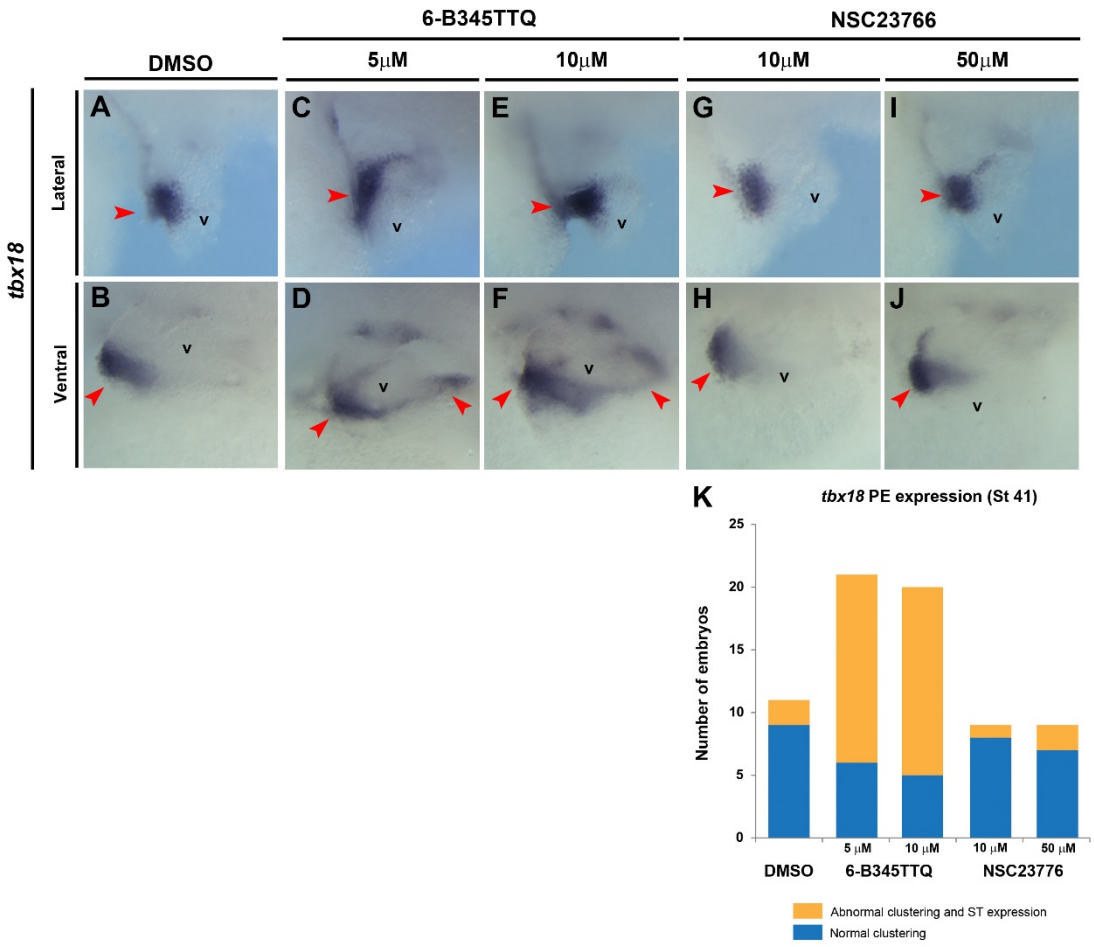


Figure A2.3. Integrin-Paxillin association is required for PE clustering. In situ hybridization for *tbx18* on whole embryos after incubation in stated concentrations of small molecules between stages 38 and 41. Lateral views (A, C, E, G, I) and ventral views (B, D, F, H, J) of the cardiac region from representative embryos are shown, red arrowheads mark PE cells. (C-F) Images depict maintained *tbx18* expression on septum transversum and clustering abnormalities as shown Fig.2M when embryos incubated in 6-B345TTQ. (K) Embryos taken from two independent experiments, p = 0.000248 by Chi Square test.v, ventricle.

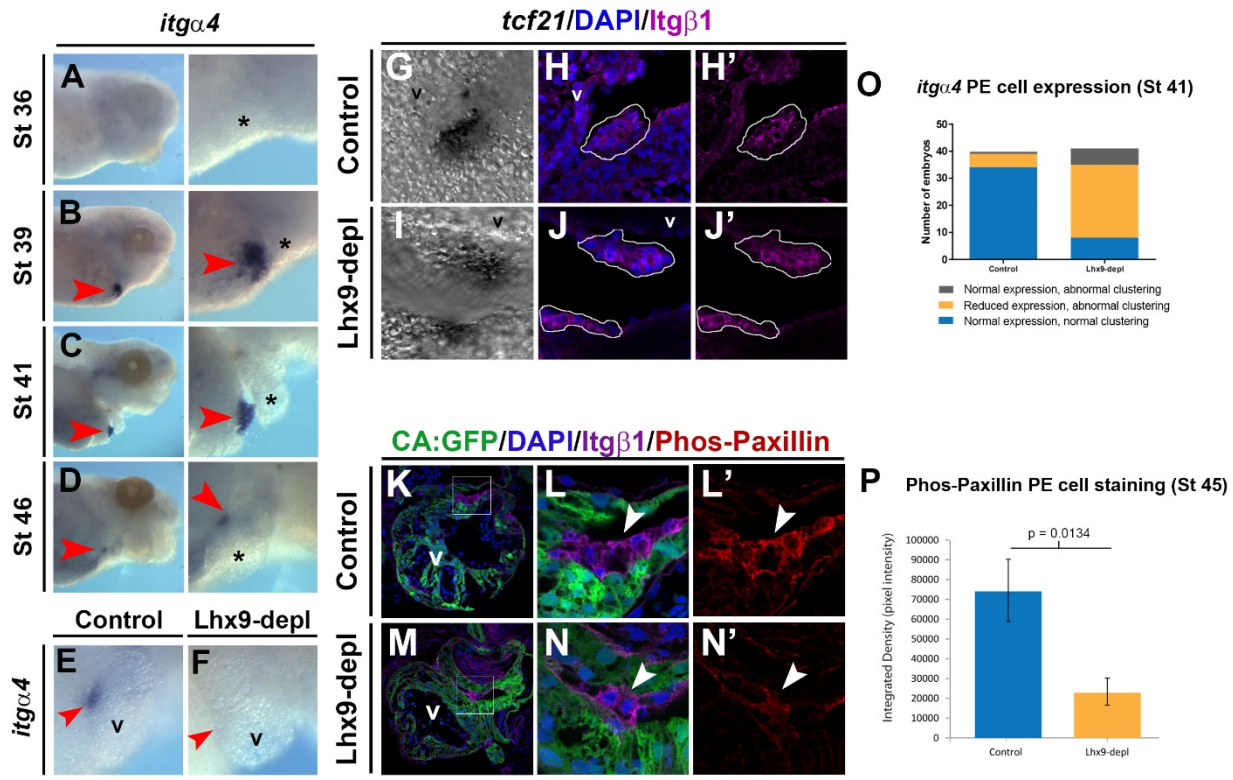


Figure A2.4. Lhx9 regulates Integrin α4-Paxillin signaling in PE cluster. (A-D) Whole embryo in situ hybridization for *itgα4*, stages 36 to 46. Lateral views of anterior portion (left panels) with magnified image of cardiac region (right panels). Red arrowhead marks PE cluster, * marks heart. (E, F) In situ hybridization for *itgα4* in control (E) and Lhx9-depleted embryos (F) at stage 41, representative cardiac region shown, anterior to right, dorsal to top, red arrowhead marks PE cluster. (O) Quantification of observed *itgα4* expression (reduced expression denoting ≤ 50% stain intensity to controls) and clustering defects shown in E, F (see Fig. 2M for phenotype assessment), , embryos taken from six independent experiments, p = <0.0001 by Chi-square. (G-J) Transverse cardiac region agarose sections from control (G, H) and Lhx9-depleted (I, J) stage 41 embryos post-in situ hybridization for *tcf21* (G, I), nuclei expression with DAPI (H, J, blue) and Itgβ1 immunohistochemistry (H, J, magenta).

White outlines (H, J) indicate PE cluster as depicted by *tcf21* and Itgβ1. (K-N) Transverse cardiac region agarose sections from control (K, L) and Lhx9-depleted (M, N) stage 45 *Xla.Tg(Cardiac-actin:GFP)*^{Mohun} embryos depicting representative immunohistochemical analysis for DAPI (blue), GFP (green) to mark cardiomyocytes, Itgβ1 (magenta) to mark epicardial cells and phosphorylated Y118-Paxillin (red). Magnified images (L, N) from white boxes (K, M). White arrowheads mark PE cluster. (P) Pixel intensity (integrated density) levels for phosphorylated Y118-Paxillin (M) from five control and ten Lhx9-depleted embryos, $p = 0.0134$ by two-tailed student t-test.v, ventricle.

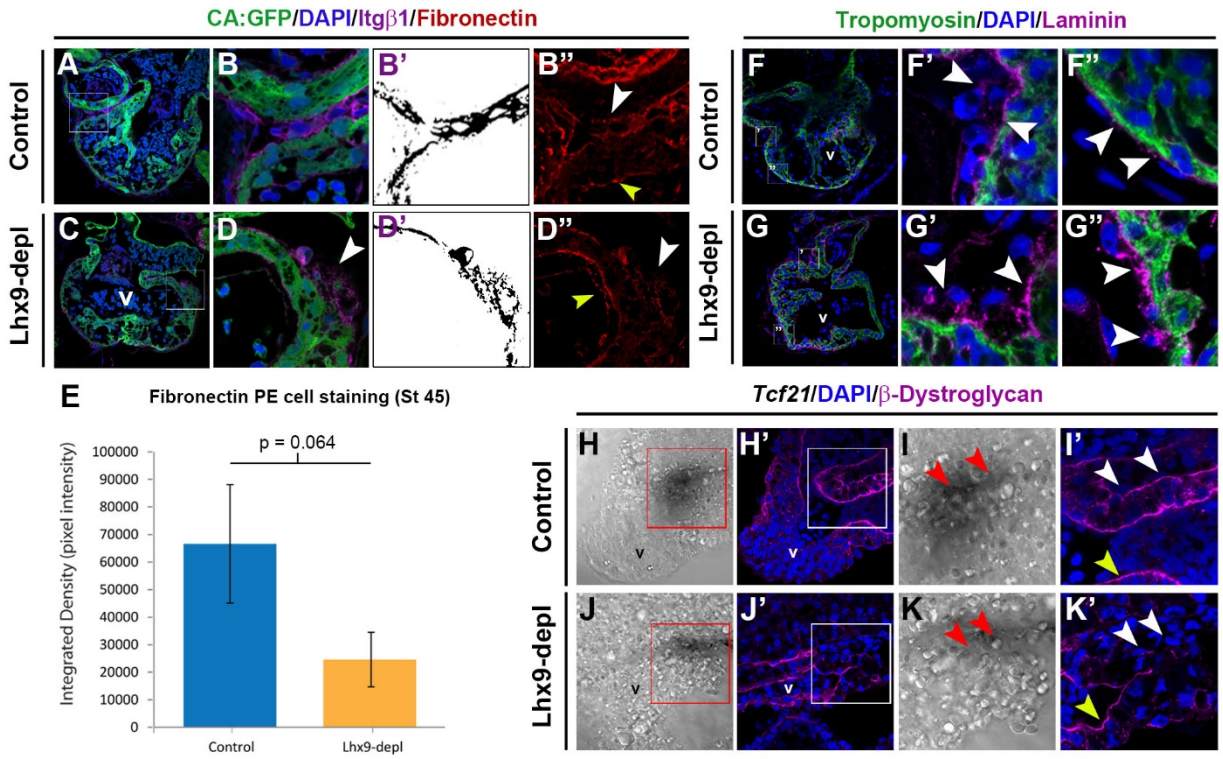


Figure A2.5. Disrupted Lhx9-Integrin signaling alters epicardial ECM environment. (A-E) Transverse cardiac agarose sections from control (A, B) and Lhx9-depleted (C, D) *Xla.Tg(Cardiac-actin:GFP)*^{Mohun} embryos at stage 45. Nuclei stained with DAPI (blue), GFP marking cardiomyocytes (green) and immunohistochemical expression for Itgβ1 (B, D, magenta) to mark epicardial cells, and Fibronectin (B", D", red). Threshold binary images (Image J) in B' and D' show Itgβ1-positive epicardial cells used to quantify pixel intensity in B" and D". (E) Pixel intensity (integrated density) levels for Fibronectin from five control and ten Lhx9-depleted embryos, p = 0.064 by two-tailed student t-test. (F-K) Transverse cardiac agarose sections from control (F, H, I) and Lhx9-depleted (G, J, K) embryos at stage 45. Nuclei stained with DAPI (blue) and immunohistochemical stain for tropomyosin (cardiomyocytes, F, G, green), laminin (F, G, magenta) and β-dystroglycan (H'-K', magenta). *Tcf21* (H-K) demonstrates PE cell attachment to the

heart. White and red boxes depict magnified images in ‘ and “ panels. White arrowheads mark PE (F’, G’, I’, K’) and migrating epicardial cells (F”, G”). Yellow arrowheads mark expression in endocardial tissue (B”, D”, I’, K’). v; ventricle. Representative images from seven (laminin) and six (β-dystroglycan) embryos per condition, from two independent experiments.

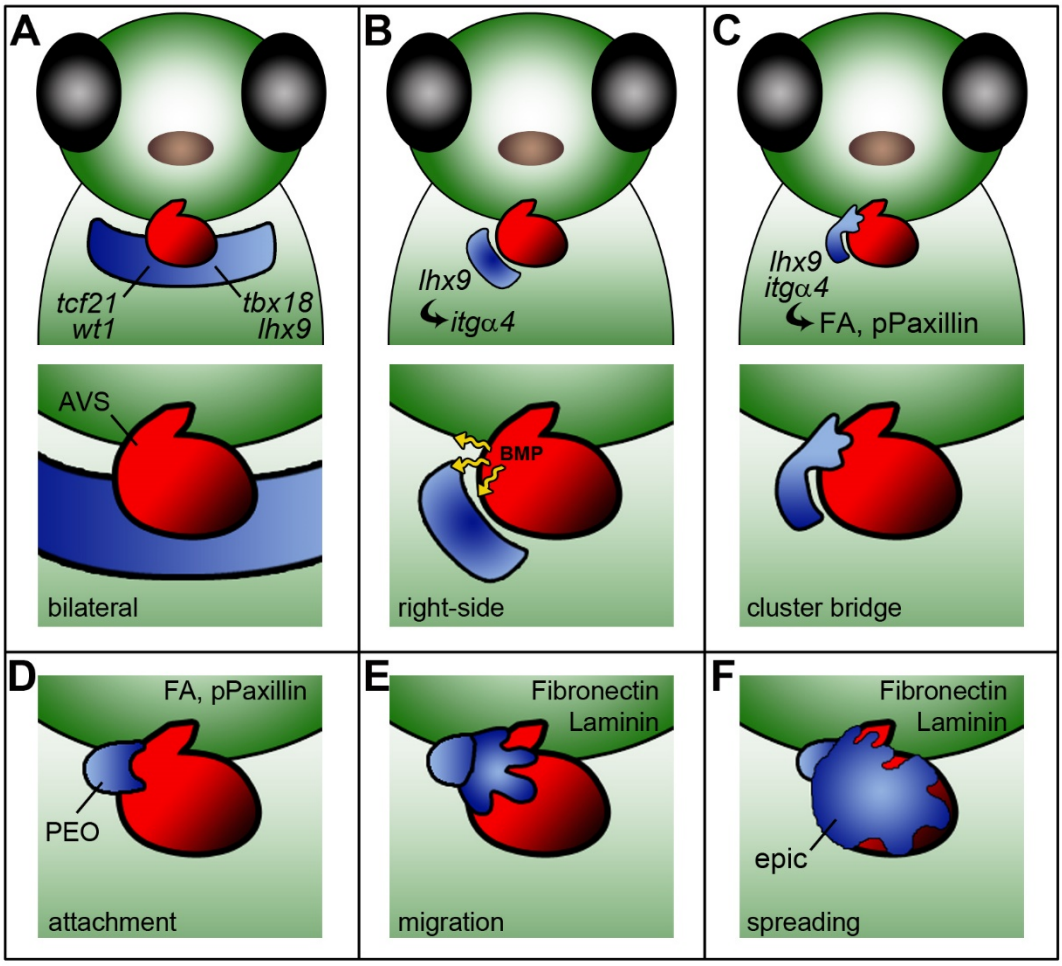


Figure A2.6. Model depicting role for Lhx9 in epicardial development in *Xenopus*.

(A) During early tadpole stages the epicardial lineage is determined and marked by transcription factors *tcf21*, *tbx18*, *wt1* and *lhx9* (blue) as a bilateral population of cells on the septum transversum caudal to the heart (red). (B) Lhx9 functions to drive clustering of cells to form the proepicardial cluster on the right side of the embryo (blue), whereby *itgα4* expression is activated. At this stage signaling factors, most likely BMP (yellow arrows), from the heart atrioventricular sulcus (AVS) direct epicardial migration. (C, D) Lhx9-Integrin-mediated signaling, including focal adhesion (FA) formation and

phosphorylation of Paxillin (pPaxillin), allow the PEO bridge (blue) to attach to the heart (red) at the AVS. (E, F) Once the PEO has attached to the heart, deposition of essential ECM components such as Fibronectin and Laminin are required for the epicardial layer to adhere and spread over the heart surface.

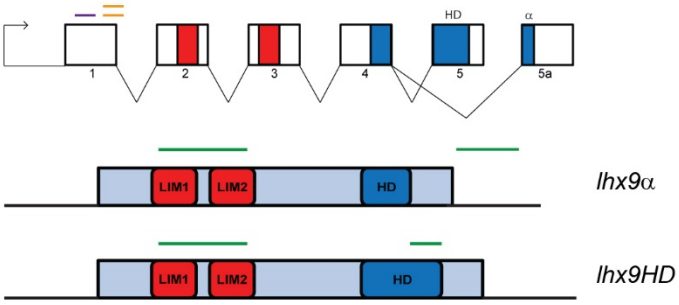


Figure SA2.1. Lhx9 genomic loci and isoform organization. Schematics of X. laevis Lhx9 genomic locus (top) and mRNA (bottom), showing localization of LIM protein binding domains (red) in exon 2 and 3, and the DNA-binding homeodomain in exons 4 and 5 (blue). Not to scale. Note that lhx9α isoform harbors a truncated HD due to alternative splicing to exon 5a. Translation-blocking (purple) and splice-blocking (orange) Morpholinos are depicted on exon 1 of genomic locus. Green bars on mRNA schematics depict localization of in situ hybridization probes.

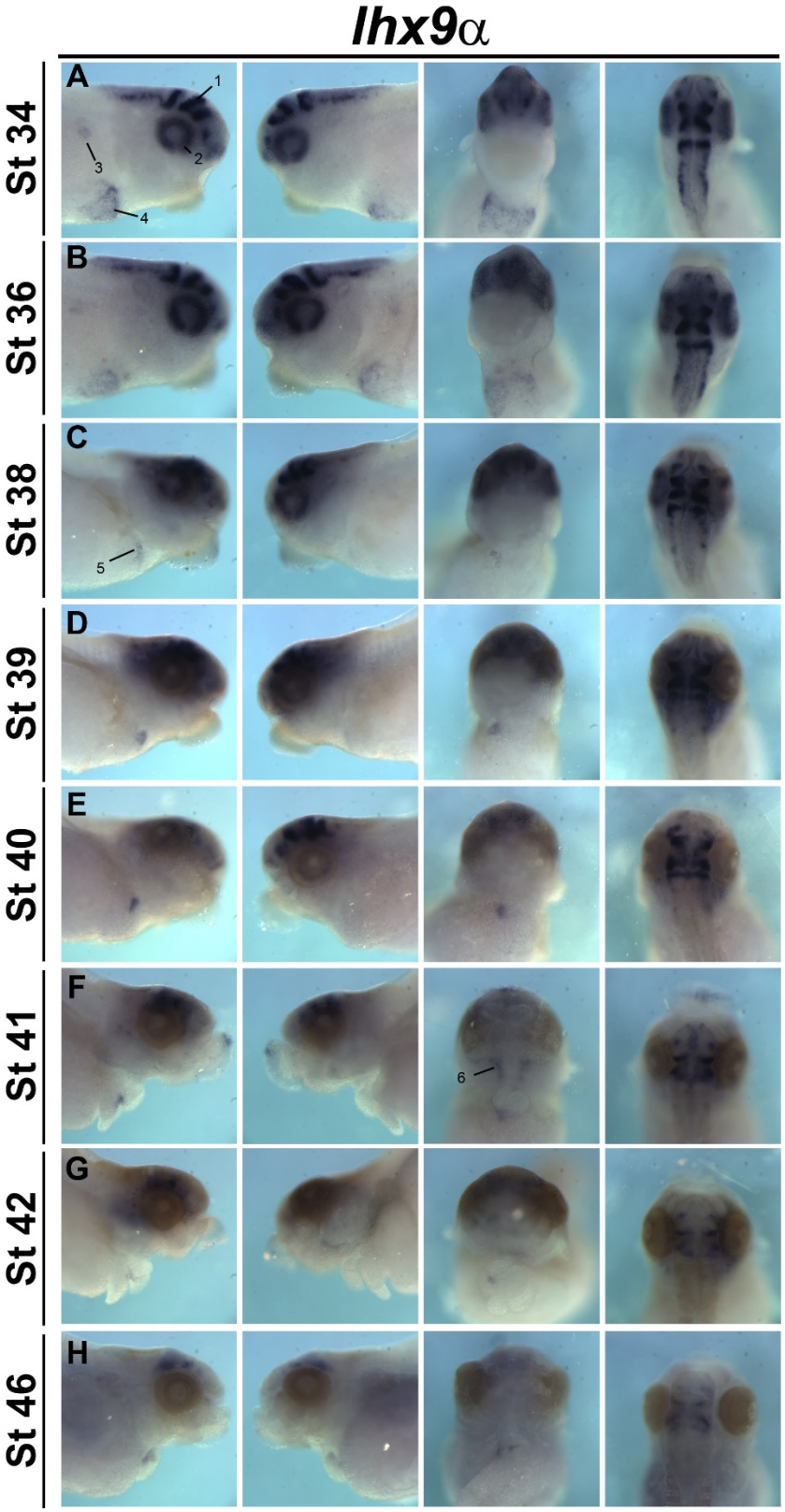

lhx9α
St 34
St 36
St 38
St 39
St 40
St 41
St 42
St 46
A
B
C
D
E
F
G
H
1
2
3
4
5
6

Figure SA2.2. Spatio-temporal analysis of lhx9α during Xenopus embryogenesis. AIn situ hybridization right, left, ventral and dorsal views of wild-type embryos showing lhx9a expression of the anterior region, from stage 34 to stage 46. 1; neural tube, 2; retina, 3; kidney, 4; septum transversum, 5; proepicardial cluster, 6; jaw cartilage.

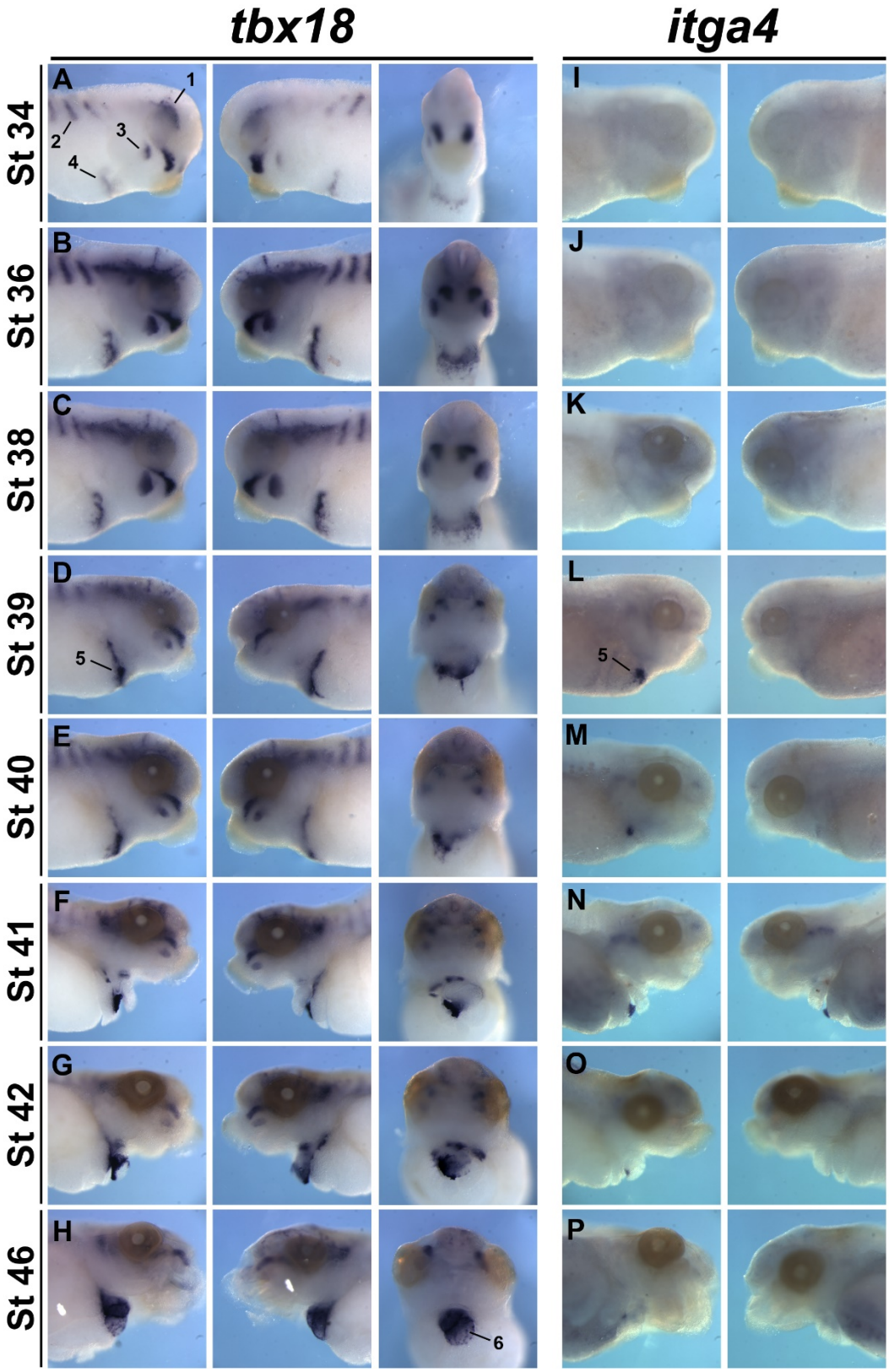

tbx18
itga4
St 34
St 36
St 38
St 39
St 40
St 41
St 42
St 46
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
1
2
3
4
5
6

Figure SA2.3. Spatio-temporal analysis of tbx18 and itga4 during Xenopus embryogenesis. (A-H) In situ hybridization of tbx18 showing right, left and ventral views of wild-type embryos from stage 34 to 46, anterior region of embryo. (I-P) In situ hybridization of itga4 showing right and left views of anterior region from stage 34-46. 1; cranial mesoderm, 2; somites, 3; branchial arches, 4; septum transversum, 5; proepicardial cluster, 6; epicardium.

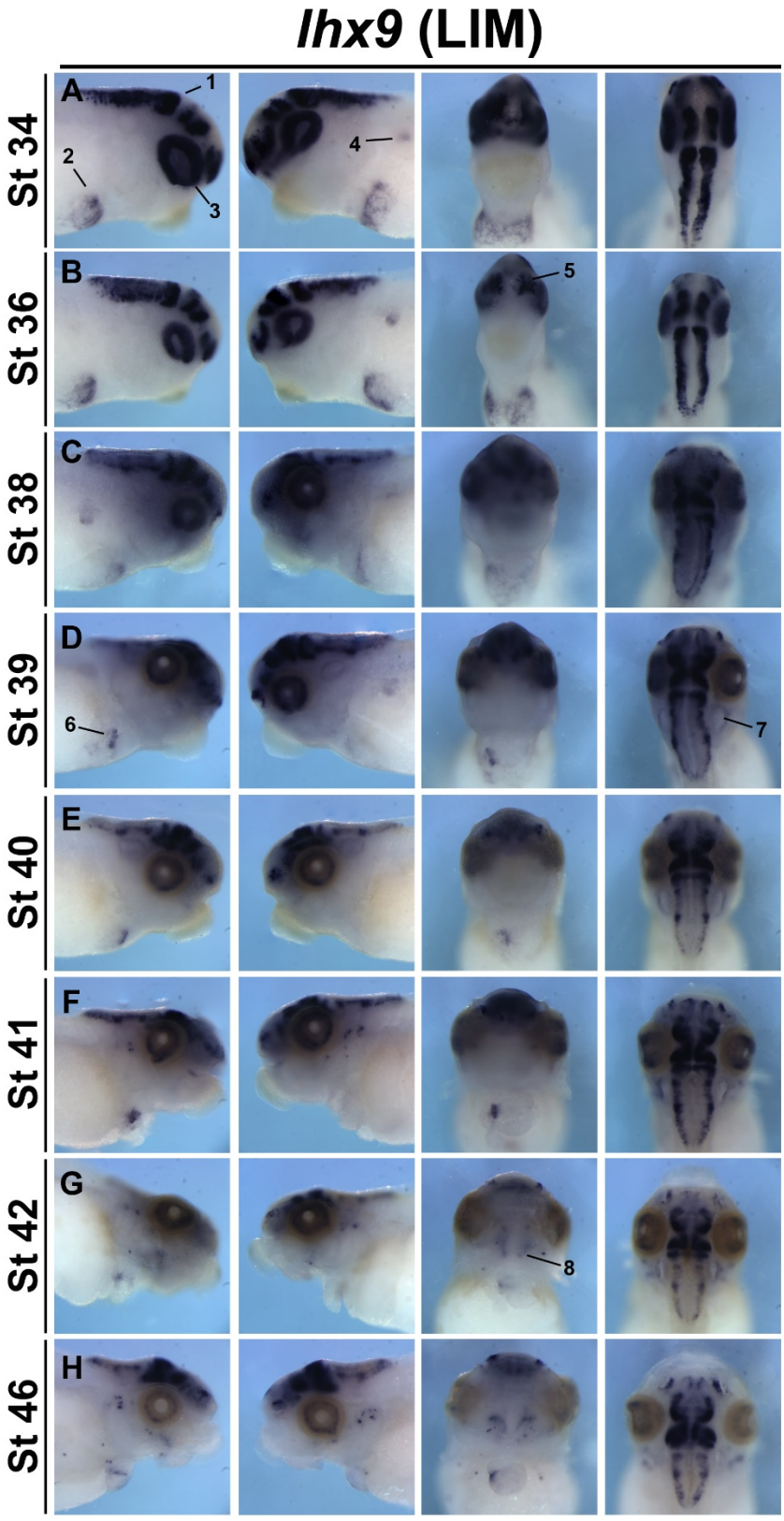

lhx9 (LIM)
St 34
St 36
St 38
St 39
St 40
St 41
St 42
St 46
A
B
C
D
E
F
G
H
1
2
3
4
5
6
7
8

Figure SA2.4. Spatio-temporal analysis of lhx9 during Xenopus embryogenesis. In situ hybridization of whole embryo anterior region using probe specific for the LIM domains of lhx9. Right, left, ventral and dorsal views of wild-type embryos from stage 34 to stage 46. 1; neural tube, 2; septum transversum, 3; retina, 4; kidney, 5; nasal placode, 6; proepicardial cluster, 7; otic placode, 8; jaw cartilage.

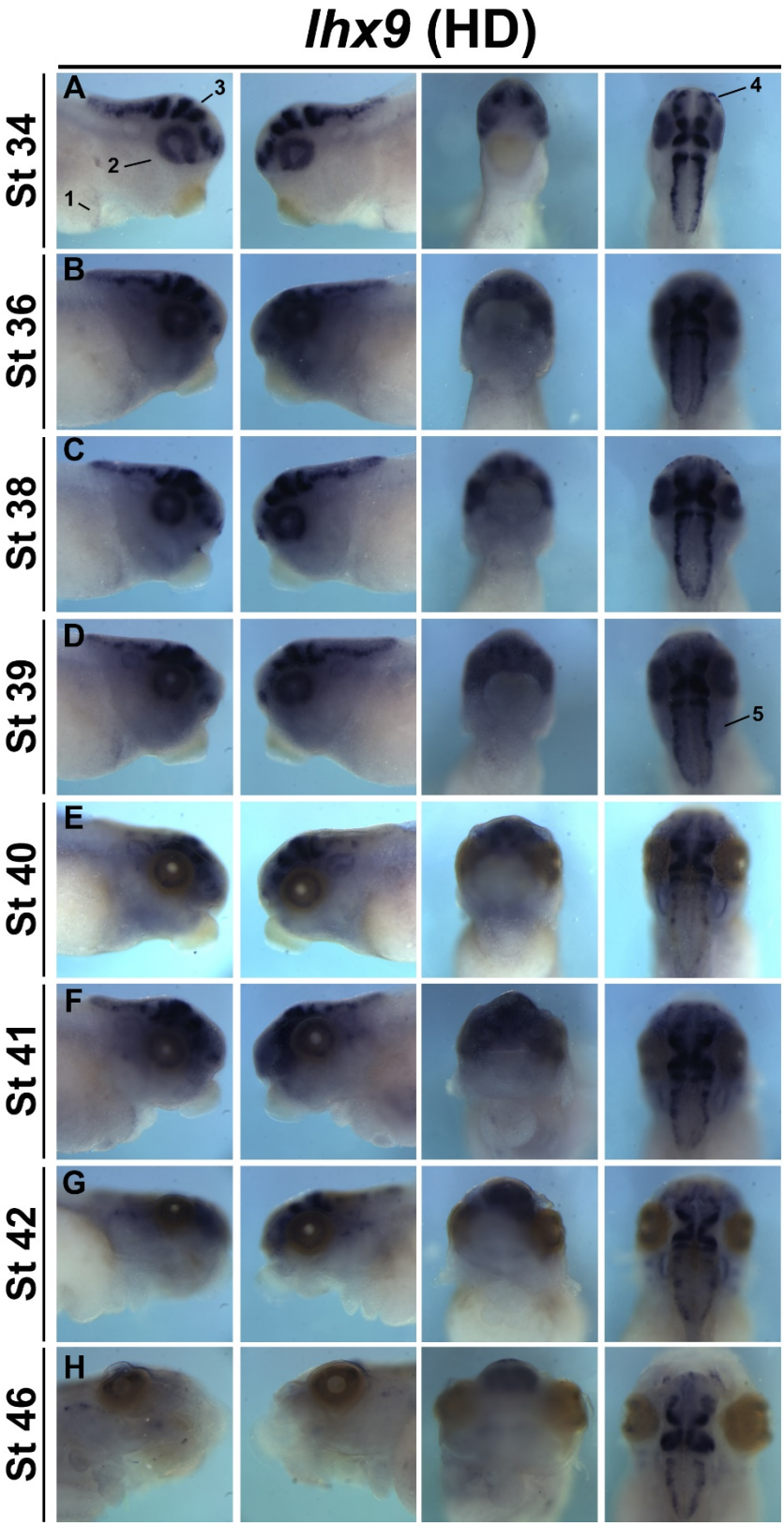

lhx9 (HD)
St 34
St 36
St 38
St 39
St 40
St 41
St 42
St 46
A
B
C
D
E
F
G
H
1
2
3
4
5

Figure SA2.5. Spatio-temporal analysis of lhx9HD during Xenopus embryogenesis. In situ hybridization right, left, ventral and dorsal views of wild-type embryos showing lhx9HD expression of the anterior region over time, from stage 34 to stage 46. 1; septum transversum, 2; retina, 3; neural tube, 4; nasal placode, 5; otic placode.

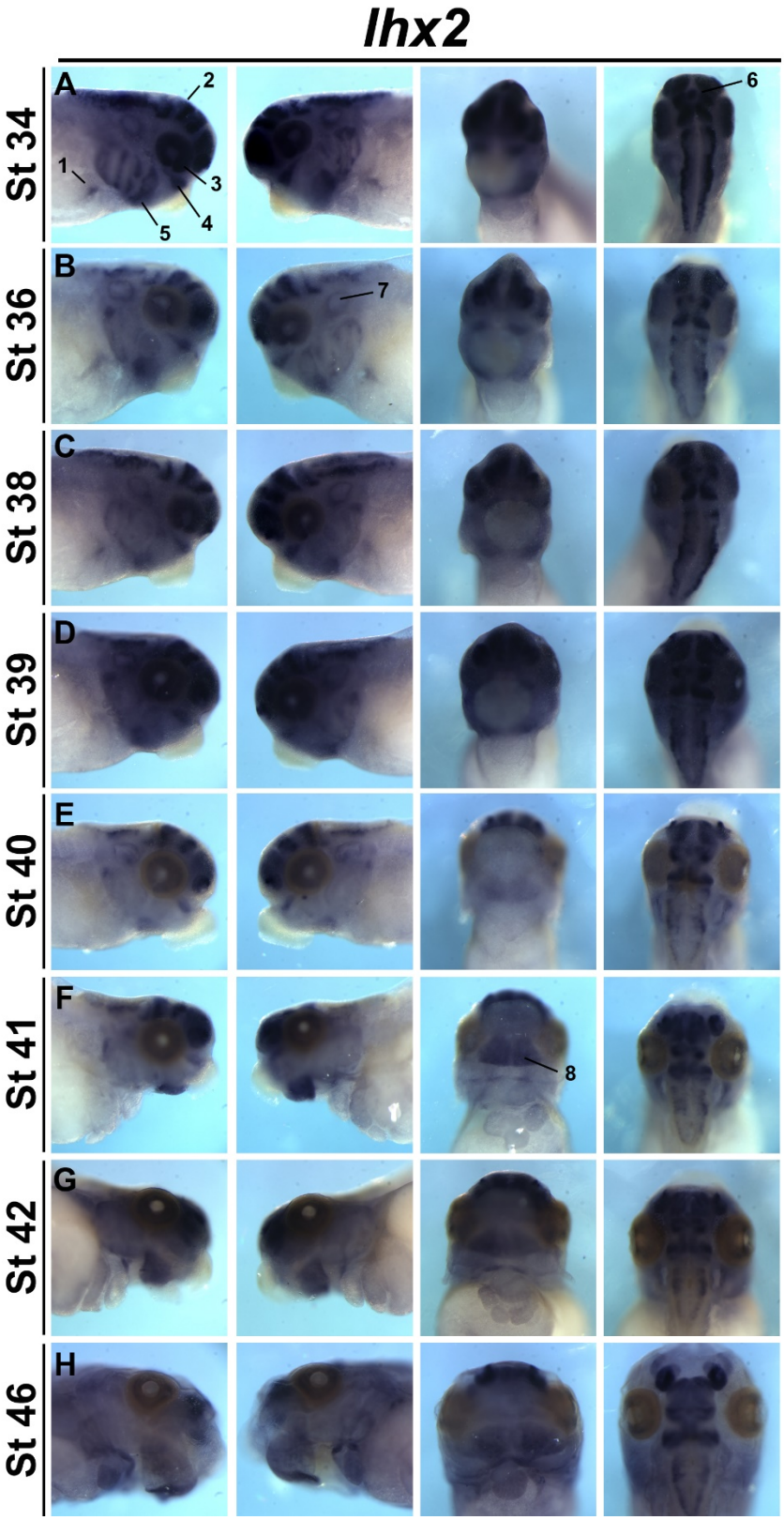

lhx2
St 34
St 36
St 38
St 39
St 40
St 41
St 42
St 46
A
B
C
D
E
F
G
H
1
2
3
4
5
6
7
8

Figure SA2.6. Spatio-temporal analysis of lhx2 during Xenopus embryogenesis. In situ hybridization right, left, ventral and dorsal views of wild-type embryos showing lhx2 expression of the anterior region over time, from stage 34 to stage 46. 1; lung bud, 2; neural tube, 3; retina, 4; mandibular arch, 5; branchial arches, 6; pineal gland, 7; otic placode, 8; lower jaw.

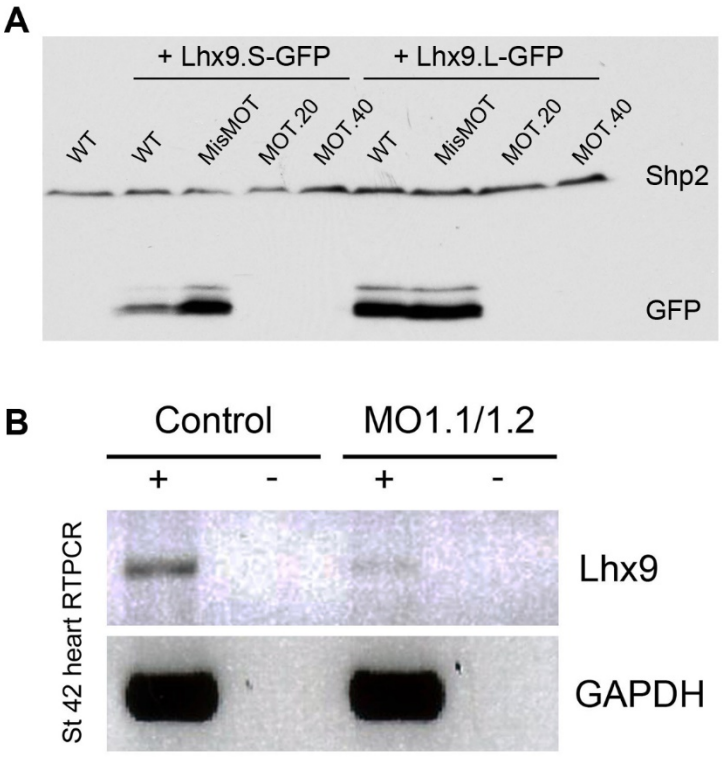


Figure SA2.7. Validation of Lhx9 depletion assays. (A) Validation of MO-specific inhibition of Lhx9 translation by GFP western blot on stage 11 embryos, injected with 1ng Lhx9-5'UTR-GFP RNA and MO at various concentrations (20-40 ng). Shp2 is used as a protein loading control. MOT targets the translational start site from both the short and long genomic versions of X. laevis Lhx9. (B) RT-PCR analysis of cardiac cDNA from stage 42 embryos injected with both MO1 (30ng each) targeting the short and long genomic versions of X. laevis Lhx9. Negative control lanes (-) are without superscript II enzyme, GAPDH PCR as loading control.

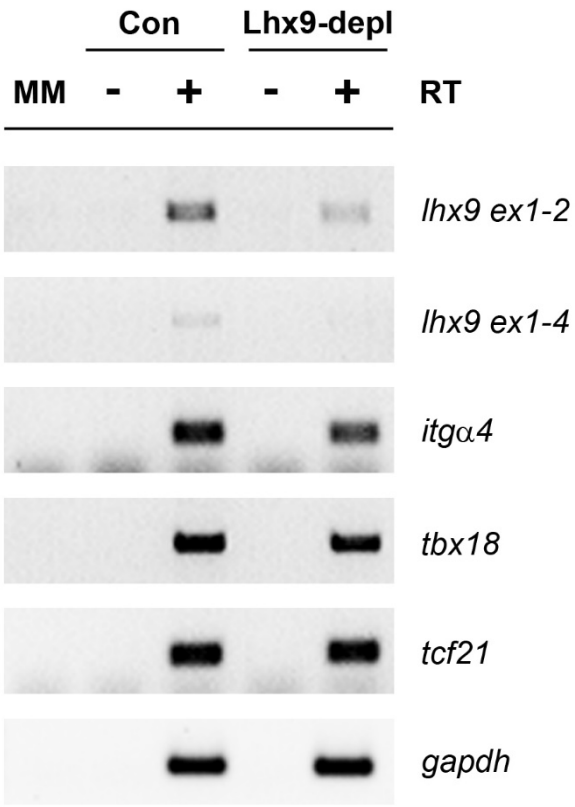


Figure SA2.8. RT-PCR validation of epicardial marker expression in Lhx9-depleted hearts. RT-PCR was performed on equivalent amounts of RNA from control or Lhx9-MO1-depleted hearts from stage 41-42. Lhx9 is significantly depleted by two PCR amplification methods, as well as decreased itg□4. Tbx18 and tcf21 appeared indistinguishable. Gapdh used as loading control.

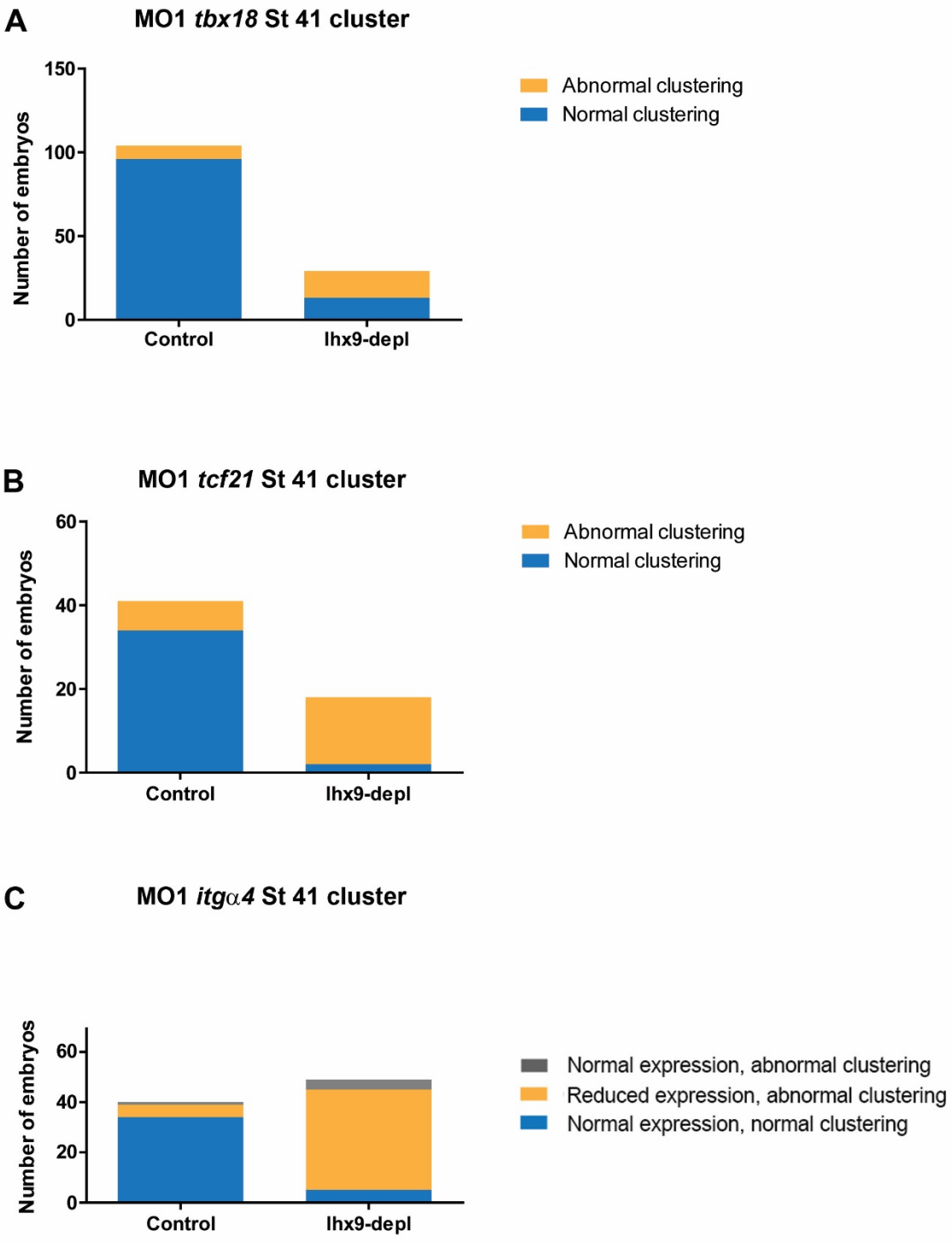

A
MO1 tbx18 St 41 cluster
Number of embryos
150
100
50
0
Control
Ihx9-depl
Abnormal clustering
Normal clustering
B
MO1 tcf21 St 41 cluster
Number of embryos
60
40
20
0
Control
Ihx9-depl
Abnormal clustering
Normal clustering
C
MO1 itgα4 St 41 cluster
Number of embryos
60
40
20
0
Control
Ihx9-depl
Normal expression, abnormal clustering
Reduced expression, abnormal clustering
Normal expression, normal clustering

Figure SA2.9. Lhx9 splice-blocking MO depletion strategy gives comparable PE clustering defects to translation-blocking MO. Clustering defects at stage 41 as assessed by (A) tbx18 and (B) tcf21 whole embryo in situ hybridization expression were present in Lhx9-MO1-depleted embryos (Fishers exact test p = <0.0001). (C) Defects observed in itga4 expression and localization was significant by Chi-square test (p = <0.0001) in Lhx9-MO1-depleted embryos. From three independent experiments.

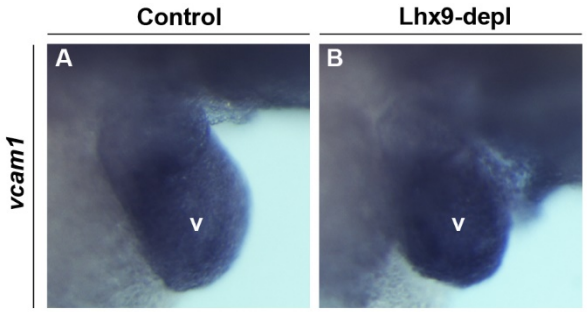


Figure SA2.10. Lhx9 depletion has no obvious effects on vcam1 expression. Lhx9-depletion at stage 41(B) did not significantly alter the expression of vcam1 in the heart compared to controls (A) by in situ hybridization, 6 embryos per condition. v; ventricle.

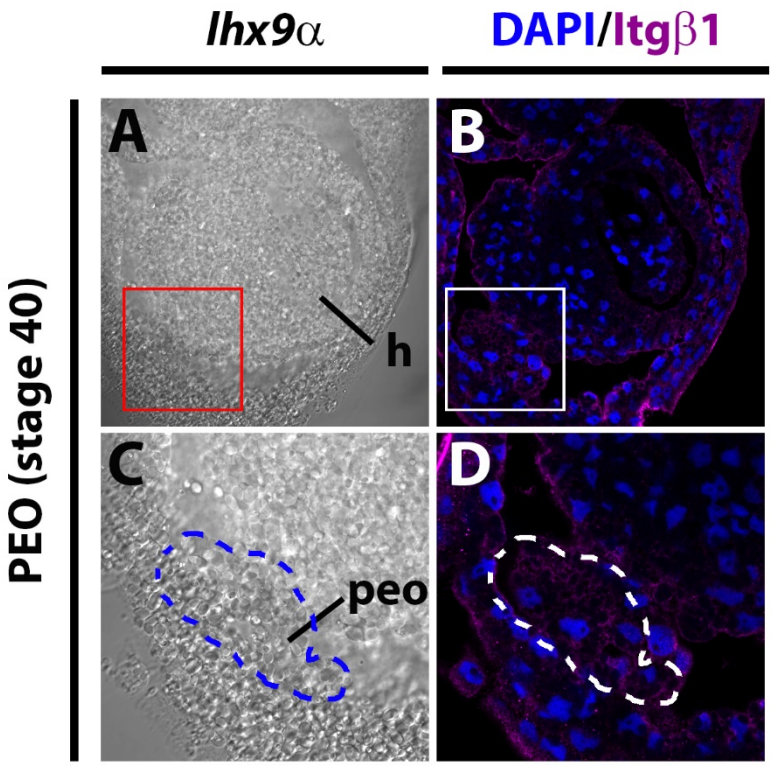


Figure SA2.11. Lhx9α expression correlates with epicardial maker Integrin β1. Transverse agarose sections demonstrate the co-localization of lhx9α in situ hybridization (A, C) with the epicardial cell marker Itgβ1 (B, D, magenta) and DAPI (blue) at stage 40. Magnified images (C, D) shown in boxes (A, B). h, heart; peo, proepicardial organ.

REFERENCES

ACHARYA, A., BAEK, S. T., HUANG, G., ESKIOCAK, B., GOETSCH, S., SUNG, C. Y., BANFI, S., SAUER, M. F., OLSEN, G. S., DUFFIELD, J. S., OLSON, E. N. & TALLQUIST, M. D. 2012. The bHLH transcription factor Tcf21 is required for lineage-specific EMT of cardiac fibroblast progenitors. *Development,* 139**,** 2139-49.

ALON, R., FEIGELSON, S. W., MANEVICH, E., ROSE, D. M., SCHMITZ, J., OVERBY, D. R., WINTER, E., GRABOVSKY, V., SHINDER, V., MATTHEWS, B. D., SOKOLOVSKY-EISENBERG, M., INGBER, D. E., BENOIT, M. & GINSBERG, M. H. 2005. Alpha4beta1-dependent adhesion strengthening under mechanical strain is regulated by paxillin association with the alpha4-cytoplasmic domain. *J Cell Biol,* 171**,** 1073-84.

ALUNNI, A., MENUET, A., CANDAL, E., PENIGAULT, J. B., JEFFERY, W. R. & RETAUX, S. 2007. Developmental mechanisms for retinal degeneration in the blind cavefish Astyanax mexicanus. *J Comp Neurol,* 505**,** 221-33.

ARAB, K., SMITH, L. T., GAST, A., WEICHENHAN, D., HUANG, J. P., CLAUS, R., HIELSCHER, T., ESPINOSA, A. V., RINGEL, M. D., MORRISON, C. D., SCHADENDORF, D., KUMAR, R. & PLASS, C. 2011. Epigenetic deregulation of TCF21 inhibits metastasis suppressor KISS1 in metastatic melanoma. *Carcinogenesis,* 32**,** 1467-73.

ARROYO, A. G., YANG, J. T., RAYBURN, H. & HYNES, R. O. 1999. Alpha4 integrins regulate the proliferation/differentiation balance of multilineage hematopoietic progenitors in vivo. *Immunity,* 11**,** 555-66.

AVRAHAM, O., HADAS, Y., VALD, L., ZISMAN, S., SCHEJTER, A., VISEL, A. & KLAR, A. 2009. Transcriptional control of axonal guidance and sorting in dorsal interneurons by the Lim-HD proteins Lhx9 and Lhx1. *Neural Dev,* 4**,** 21.

BACHY, I., VERNIER, P. & RETAUX, S. 2001. The LIM-homeodomain gene family in the developing Xenopus brain: conservation and divergences with the mouse related to the evolution of the forebrain. *J Neurosci,* 21**,** 7620-9.

BALASUBRAMANIAN, R., BUI, A., DING, Q. & GAN, L. 2014. Expression of LIM-homeodomain transcription factors in the developing and mature mouse retina. *Gene Expr Patterns,* 14**,** 1-8.

BAX, N. A., BLEYL, S. B., GALLINI, R., WISSE, L. J., HUNTER, J., VAN OORSCHOT, A. A., MAHTAB, E. A., LIE-VENEMA, H., GOUMANS, M. J., BETSHOLTZ, C. & GITTENBERGER-DE GROOT, A. C. 2010. Cardiac malformations in Pdgfralpha mutant embryos are associated with increased expression of WT1 and Nkx2.5 in the second heart field. *Dev Dyn,* 239**,** 2307-17.

BELLIS, S. L., MILLER, J. T. & TURNER, C. E. 1995. Characterization of tyrosine phosphorylation of paxillin in vitro by focal adhesion kinase. *J Biol Chem,* 270**,** 17437-41.

BENESH, E. C., MILLER, P. M., PFALTZGRAFF, E. R., GREGA-LARSON, N. E., HAGER, H. A., SUNG, B. H., QU, X., BALDWIN, H. S., WEAVER, A. M. & BADER, D. M. 2013. Bves and NDRG4 regulate directional epicardial cell migration through autocrine extracellular matrix deposition. *Mol Biol Cell,* 24**,** 3496-510.

BERTUZZI, S., PORTER, F. D., PITTS, A., KUMAR, M., AGULNICK, A., WASSIF, C. & WESTPHAL, H. 1999. Characterization of Lhx9, a novel LIM/homeobox gene expressed by the pioneer neurons in the mouse cerebral cortex. *Mech Dev,* 81**,** 193-8.

BIRK, O. S., CASIANO, D. E., WASSIF, C. A., COGLIATI, T., ZHAO, L., ZHAO, Y., GRINBERG, A., HUANG, S., KREIDBERG, J. A., PARKER, K. L., PORTER, F. D. & WESTPHAL, H. 2000. The LIM homeobox gene Lhx9 is essential for mouse gonad formation. *Nature,* 403**,** 909-13.

BLAIR, S. S., BROWER, D. L., THOMAS, J. B. & ZAVORTINK, M. 1994. The role of apterous in the control of dorsoventral compartmentalization and PS integrin gene expression in the developing wing of Drosophila. *Development,* 120**,** 1805-15.

BRODERS-BONDON, F., CHESNEAU, A., ROMERO-OLIVA, F., MAZABRAUD, A., MAYOR, R. & THIERY, J. P. 2007. Regulation of XSnail2 expression by Rho GTPases. *Dev Dyn,* 236**,** 2555-66.

BURRIDGE, K., TURNER, C. E. & ROMER, L. H. 1992. Tyrosine phosphorylation of paxillin and pp125FAK accompanies cell adhesion to extracellular matrix: a role in cytoskeletal assembly. *J Cell Biol,* 119**,** 893-903.

CHARPENTIER, M. S., TANDON, P., TRINCOT, C. E., KOUTLEVA, E. K. & CONLON, F. L. 2015. A distinct mechanism of vascular lumen formation in Xenopus requires EGFL7. *PLoS One,* 10**,** e0116086.

COMBS, M. D., BRAITSCH, C. M., LANGE, A. W., JAMES, J. F. & YUTZEY, K. E. 2011. NFATC1 promotes epicardium-derived cell invasion into myocardium. *Development,* 138**,** 1747-57.

CUI, S., ROSS, A., STALLINGS, N., PARKER, K. L., CAPEL, B. & QUAGGIN, S. E. 2004. Disrupted gonadogenesis and male-to-female sex reversal in Pod1 knockout mice. *Development,* 131**,** 4095-105.

DETTMAN, R. W., PAE, S. H., MORABITO, C. & BRISTOW, J. 2003. Inhibition of alpha4-integrin stimulates epicardial-mesenchymal transformation and alters migration and cell fate of epicardially derived mesenchyme. *Dev Biol,* 257**,** 315-28.

DIAZ, J., MENDOZA, P., ORTIZ, R., DIAZ, N., LEYTON, L., STUPACK, D., QUEST, A. F. & TORRES, V. A. 2014. Rab5 is required in metastatic cancer cells for Caveolin-1-enhanced Rac1 activation, migration and invasion. *J Cell Sci,* 127**,** 2401-6.

DOKIC, D. & DETTMAN, R. W. 2006. VCAM-1 inhibits TGFbeta stimulated epithelial-mesenchymal transformation by modulating Rho activity and stabilizing intercellular adhesion in epicardial mesothelial cells. *Dev Biol,* 299**,** 489-504.

ERVASTI, J. M. & CAMPBELL, K. P. 1993. A role for the dystrophin-glycoprotein complex as a transmembrane linker between laminin and actin. *J Cell Biol,* 122**,** 809-23.

FAILLI, V., ROGARD, M., MATTEI, M. G., VERNIER, P. & RETAUX, S. 2000. Lhx9 and Lhx9alpha LIM-homeodomain factors: genomic structure, expression patterns, chromosomal localization, and phylogenetic analysis. *Genomics,* 64**,** 307-17.

FERAL, C. C., ROSE, D. M., HAN, J., FOX, N., SILVERMAN, G. J., KAUSHANSKY, K. & GINSBERG, M. H. 2006. Blocking the alpha 4 integrin-paxillin interaction selectively impairs mononuclear leukocyte recruitment to an inflammatory site. *J Clin Invest,* 116**,** 715-23.

FRANSEN, M. E. & LEMANSKI, L. F. 1991. Extracellular matrix of the developing heart in normal and cardiac lethal mutant axolotls, Ambystoma mexicanum. *Anat Rec,* 230**,** 387-405.

GARDINER, N. J. 2011. Integrins and the extracellular matrix: key mediators of development and regeneration of the sensory nervous system. *Dev Neurobiol,* 71**,** 1054-72.

GEHLER, S., PONIK, S. M., RICHING, K. M. & KEELY, P. J. 2013. Bi-directional signaling: extracellular matrix and integrin regulation of breast tumor progression. *Crit Rev Eukaryot Gene Expr,* 23**,** 139-57.

GESSERT, S. & KUHL, M. 2009. Comparative gene expression analysis and fate mapping studies suggest an early segregation of cardiogenic lineages in Xenopus laevis. *Dev Biol,* 334**,** 395-408.

GITTENBERGER-DE GROOT, A. C., VRANCKEN PEETERS, M. P., BERGWERFF, M., MENTINK, M. M. & POELMANN, R. E. 2000. Epicardial outgrowth inhibition leads to compensatory mesothelial outflow tract collar and abnormal cardiac septation and coronary formation. *Circ Res,* 87**,** 969-71.

GITTENBERGER-DE GROOT, A. C., WINTER, E. M. & POELMANN, R. E. 2010. Epicardium-derived cells (EPDCs) in development, cardiac disease and repair of ischemia. *J Cell Mol Med,* 14**,** 1056-60.

GUADIX, J. A., CARMONA, R., MUNOZ-CHAPULI, R. & PEREZ-POMARES, J. M. 2006. In vivo and in vitro analysis of the vasculogenic potential of avian proepicardial and epicardial cells. *Dev Dyn,* 235**,** 1014-26.

HARATA, A., MATSUZAKI, T., NISHIKAWA, A. & IHARA, S. 2013. The cell sorting process of Xenopus gastrula cells involves the acto-myosin system and TGF-beta signaling. *In Vitro Cell Dev Biol Anim,* 49**,** 220-9.

HAREL, I., MAEZAWA, Y., AVRAHAM, R., RINON, A., MA, H. Y., CROSS, J. W., LEVIATAN, N., HEGESH, J., ROY, A., JACOB-HIRSCH, J., RECHAVI, G., CARVAJAL, J., TOLE, S., KIOUSSI, C., QUAGGIN, S. & TZAHOR, E. 2012. Pharyngeal mesoderm regulatory network controls cardiac and head muscle morphogenesis. *Proc Natl Acad Sci U S A,* 109**,** 18839-44.

HARLAND, R. M. 1991. In situ hybridization: an improved whole-mount method for Xenopus embryos. *Methods Cell Biol,* 36**,** 685-95.

HIROSE, T., KARASAWA, M., SUGITANI, Y., FUJISAWA, M., AKIMOTO, K., OHNO, S. & NODA, T. 2006. PAR3 is essential for cyst-mediated epicardial development by establishing apical cortical domains. *Development,* 133**,** 1389-98.

HUANG, X. G., CHEN, Y. Z., ZHANG, Z. T., WEI, Y. T., MA, H. Z., ZHANG, T. & ZHANG, S. C. 2013. Rac1 modulates the vitreous-induced plasticity of mesenchymal movement in retinal pigment epithelial cells. *Clin Experiment Ophthalmol,* 41**,** 779-87.

HUMPHRIES, M. J., SHERIDAN, J., MOULD, A. P. & NEWHAM, P. 1995. Mechanisms of VCAM-1 and fibronectin binding to integrin alpha 4 beta 1: implications for integrin function and rational drug design. *Ciba Found Symp,* 189**,** 177-91; discussion 191-9.

HUNG, W. C., CHEN, S. H., PAUL, C. D., STROKA, K. M., LO, Y. C., YANG, J. T. & KONSTANTOPOULOS, K. 2013. Distinct signaling mechanisms regulate migration in unconfined versus confined spaces. *J Cell Biol,* 202**,** 807-24.

HUTTENLOCHER, A. & HORWITZ, A. R. 2011. Integrins in cell migration. *Cold Spring Harb Perspect Biol,* 3**,** a005074.

ISHII, Y., GARRIOCK, R. J., NAVETTA, A. M., COUGHLIN, L. E. & MIKAWA, T. 2010. BMP signals promote proepicardial protrusion necessary for recruitment of coronary vessel and epicardial progenitors to the heart. *Dev Cell,* 19**,** 307-16.

IWAI, L. & KAWASAKI, H. 2009. Molecular development of the lateral geniculate nucleus in the absence of retinal waves during the time of retinal axon eye-specific segregation. *Neuroscience,* 159**,** 1326-37.

IWASAKI, T., NAKATA, A., MUKAI, M., SHINKAI, K., YANO, H., SABE, H., SCHAEFER, E., TATSUTA, M., TSUJIMURA, T., TERADA, N., KAKISHITA, E. & AKEDO, H. 2002. Involvement of phosphorylation of Tyr-31 and Tyr-118 of paxillin in MM1 cancer cell migration. *Int J Cancer,* 97**,** 330-5.

JAHR, M., SCHLUETER, J., BRAND, T. & MANNER, J. 2008. Development of the proepicardium in Xenopus laevis. *Dev Dyn,* 237**,** 3088-96.

KALMAN, F., VIRAGH, S. & MODIS, L. 1995. Cell surface glycoconjugates and the extracellular matrix of the developing mouse embryo epicardium. *Anat Embryol (Berl),* 191**,** 451-64.

KANAI, Y., KANAI-AZUMA, M., TAJIMA, Y., BIRK, O. S., HAYASHI, Y. & SANAI, Y. 2000. Identification of a stromal cell type characterized by the secretion of a soluble integrin-binding protein, MFG-E8, in mouse early gonadogenesis. *Mech Dev,* 96**,** 223-7.

KANG, J., GU, Y., LI, P., JOHNSON, B. L., SUCOV, H. M. & THOMAS, P. S. 2008. PDGF-A as an epicardial mitogen during heart development. *Dev Dyn,* 237**,** 692-701.

KATZ, T. C., SINGH, M. K., DEGENHARDT, K., RIVERA-FELICIANO, J., JOHNSON, R. L., EPSTEIN, J. A. & TABIN, C. J. 2012. Distinct compartments of the proepicardial organ give rise to coronary vascular endothelial cells. *Dev Cell,* 22**,** 639-50.

KLIETSCH, R., ERVASTI, J. M., ARNOLD, W., CAMPBELL, K. P. & JORGENSEN, A. O. 1993. Dystrophin-glycoprotein complex and laminin colocalize to the sarcolemma and transverse tubules of cardiac muscle. *Circ Res,* 72**,** 349-60.

KUMMER, C. & GINSBERG, M. H. 2006. New approaches to blockade of alpha4-integrins, proven therapeutic targets in chronic inflammation. *Biochem Pharmacol,* 72**,** 1460-8.

KUMMER, C., PETRICH, B. G., ROSE, D. M. & GINSBERG, M. H. 2010. A small molecule that inhibits the interaction of paxillin and alpha 4 integrin inhibits accumulation of mononuclear leukocytes at a site of inflammation. *J Biol Chem,* 285**,** 9462-9.

LANGDON, Y., TANDON, P., PADEN, E., DUDDY, J., TAYLOR, J. M. & CONLON, F. L. 2012. SHP-2 acts via ROCK to regulate the cardiac actin cytoskeleton. *Development,* 139**,** 948-57.

LATINKIC, B. V., COOPER, B., TOWERS, N., SPARROW, D., KOTECHA, S. & MOHUN, T. J. 2002. Distinct enhancers regulate skeletal and cardiac muscle-specific expression programs of the cardiac alpha-actin gene in Xenopus embryos. *Dev Biol,* 245**,** 57-70.

LAVINE, K. J., YU, K., WHITE, A. C., ZHANG, X., SMITH, C., PARTANEN, J. & ORNITZ, D. M. 2005. Endocardial and epicardial derived FGF signals regulate myocardial proliferation and differentiation in vivo. *Dev Cell,* 8**,** 85-95.

LEPILINA, A., COON, A. N., KIKUCHI, K., HOLDWAY, J. E., ROBERTS, R. W., BURNS, C. G. & POSS, K. D. 2006. A dynamic epicardial injury response supports progenitor cell activity during zebrafish heart regeneration. *Cell,* 127**,** 607-19.

LEWIS, J. M. & SCHWARTZ, M. A. 1998. Integrins regulate the association and phosphorylation of paxillin by c-Abl. *J Biol Chem,* 273**,** 14225-30.

LI, P., CAVALLERO, S., GU, Y., CHEN, T. H., HUGHES, J., HASSAN, A. B., BRUNING, J. C., PASHMFOROUSH, M. & SUCOV, H. M. 2011. IGF signaling directs ventricular cardiomyocyte proliferation during embryonic heart development. *Development,* 138**,** 1795-805.

LIE-VENEMA, H., ERALP, I., MAAS, S., GITTENBERGER-DE GROOT, A. C., POELMANN, R. E. & DERUITER, M. C. 2005. Myocardial heterogeneity in permissiveness for epicardium-derived cells and endothelial precursor cells along the developing heart tube at the onset of coronary vascularization. *Anat Rec A Discov Mol Cell Evol Biol,* 282**,** 120-9.

LIE-VENEMA, H., VAN DEN AKKER, N. M., BAX, N. A., WINTER, E. M., MAAS, S., KEKARAINEN, T., HOEBEN, R. C., DERUITER, M. C., POELMANN, R. E. & GITTENBERGER-DE GROOT, A. C. 2007. Origin, fate, and function of epicardium-derived cells (EPDCs) in normal and abnormal cardiac development. *ScientificWorldJournal,* 7**,** 1777-98.

LIMANA, F., CAPOGROSSI, M. C. & GERMANI, A. 2010. The epicardium in cardiac repair: From the stem cell view. *Pharmacol Ther*.

LIU, S., THOMAS, S. M., WOODSIDE, D. G., ROSE, D. M., KIOSSES, W. B., PFAFF, M. & GINSBERG, M. H. 1999. Binding of paxillin to alpha4 integrins modifies integrin-dependent biological responses. *Nature,* 402**,** 676-81.

LOBB, R. R. & HEMLER, M. E. 1994. The pathophysiologic role of alpha 4 integrins in vivo. *J Clin Invest,* 94**,** 1722-8.

MANNER, J. 1992. The development of pericardial villi in the chick embryo. *Anat Embryol (Berl),* 186**,** 379-85.

MANNER, J. 1999. Does the subepicardial mesenchyme contribute myocardioblasts to the myocardium of the chick embryo heart? A quail-chick chimera study tracing the fate of the epicardial primordium. *Anat Rec,* 255**,** 212-26.

MANNER, J., PEREZ-POMARES, J. M., MACIAS, D. & MUNOZ-CHAPULI, R. 2001. The origin, formation and developmental significance of the epicardium: a review. *Cells Tissues Organs,* 169**,** 89-103.

MANNINEN, A. 2015. Epithelial polarity - Generating and integrating signals from the ECM with integrins. *Exp Cell Res*.

MERCER, S. E., ODELBERG, S. J. & SIMON, H. G. 2013. A dynamic spatiotemporal extracellular matrix facilitates epicardial-mediated vertebrate heart regeneration. *Dev Biol,* 382**,** 457-69.

MIKAWA, T. & GOURDIE, R. G. 1996. Pericardial mesoderm generates a population of coronary smooth muscle cells migrating into the heart along with ingrowth of the epicardial organ. *Dev Biol,* 174**,** 221-32.

MILLER, C. L., ANDERSON, D. R., KUNDU, R. K., RAIESDANA, A., NURNBERG, S. T., DIAZ, R., CHENG, K., LEEPER, N. J., CHEN, C. H., CHANG, I. S., SCHADT, E. E., HSIUNG, C. A., ASSIMES, T. L. & QUERTERMOUS, T. 2013. Disease-related growth factor and embryonic signaling pathways modulate an enhancer of TCF21 expression at the 6q23.2 coronary heart disease locus. *PLoS Genet,* 9**,** e1003652.

MILLER, C. L., HAAS, U., DIAZ, R., LEEPER, N. J., KUNDU, R. K., PATLOLLA, B., ASSIMES, T. L., KAISER, F. J., PERISIC, L., HEDIN, U., MAEGDEFESSEL, L., SCHUNKERT, H., ERDMANN, J., QUERTERMOUS, T. & SCZAKIEL, G. 2014. Coronary heart disease-associated variation in TCF21 disrupts a miR-224 binding site and miRNA-mediated regulation. *PLoS Genet,* 10**,** e1004263.

MOLLE, B., PERE, S., FAILLI, V., BACH, I. & RETAUX, S. 2004. Lhx9 and lhx9alpha: differential biochemical properties and effects on neuronal differentiation. *DNA Cell Biol,* 23**,** 761-8.

MONCAUT, N., CROSS, J. W., SILIGAN, C., KEITH, A., TAYLOR, K., RIGBY, P. W. & CARVAJAL, J. J. 2012. Musculin and TCF21 coordinate the maintenance of myogenic regulatory factor expression levels during mouse craniofacial development. *Development,* 139**,** 958-67.

MOORE, A. W., SCHEDL, A., MCINNES, L., DOYLE, M., HECKSHER-SORENSEN, J. & HASTIE, N. D. 1998. YAC transgenic analysis reveals Wilms' tumour 1 gene activity in the proliferating coelomic epithelium, developing diaphragm and limb. *Mech Dev,* 79**,** 169-84.

MORENO, N., BACHY, I., RETAUX, S. & GONZALEZ, A. 2004. LIM-homeodomain genes as developmental and adult genetic markers of Xenopus forebrain functional subdivisions. *J Comp Neurol,* 472**,** 52-72.

MORENO, N., GONZALEZ, A. & RETAUX, S. 2008. Evidences for tangential migrations in Xenopus telencephalon: developmental patterns and cell tracking experiments. *Dev Neurobiol,* 68**,** 504-20.

NAHIRNEY, P. C., MIKAWA, T. & FISCHMAN, D. A. 2003. Evidence for an extracellular matrix bridge guiding proepicardial cell migration to the myocardium of chick embryos. *Dev Dyn,* 227**,** 511-23.

NAKAMURA, K., YANO, H., UCHIDA, H., HASHIMOTO, S., SCHAEFER, E. & SABE, H. 2000. Tyrosine phosphorylation of paxillin alpha is involved in temporospatial regulation of paxillin-containing focal adhesion formation and F-actin organization in motile cells. *J Biol Chem,* 275**,** 27155-64.

NIEUWKOOP, P. D. & FABER, J. 1967. Normal table of *Xenopus laevis (Daudin)*. *Amsterdam: North-Holland Publishing*.

OSHIMA, Y., NOGUCHI, K. & NAKAMURA, M. 2007. Expression of Lhx9 isoforms in the developing gonads of Rana rugosa. *Zoolog Sci,* 24**,** 798-802.

OTTOLENGHI, C., MOREIRA-FILHO, C., MENDONCA, B. B., BARBIERI, M., FELLOUS, M., BERKOVITZ, G. D. & MCELREAVEY, K. 2001. Absence of mutations involving the LIM homeobox domain gene LHX9 in 46,XY gonadal agenesis and dysgenesis. *J Clin Endocrinol Metab,* 86**,** 2465-9.

PAE, S. H., DOKIC, D. & DETTMAN, R. W. 2008. Communication between integrin receptors facilitates epicardial cell adhesion and matrix organization. *Dev Dyn,* 237**,** 962-78.

PENNISI, D. J., BALLARD, V. L. & MIKAWA, T. 2003. Epicardium is required for the full rate of myocyte proliferation and levels of expression of myocyte mitogenic factors FGF2 and its receptor, FGFR-1, but not for transmural myocardial patterning in the embryonic chick heart. *Dev Dyn,* 228**,** 161-72.

PERALTA, M., STEED, E., HARLEPP, S., GONZALEZ-ROSA, J. M., MONDUC, F., ARIZA-COSANO, A., CORTES, A., RAYON, T., GOMEZ-SKARMETA, J. L., ZAPATA, A., VERMOT, J. & MERCADER, N. 2013. Heartbeat-driven pericardiac fluid forces contribute to epicardium morphogenesis. *Curr Biol,* 23**,** 1726-35.

PETIT, V., BOYER, B., LENTZ, D., TURNER, C. E., THIERY, J. P. & VALLES, A. M. 2000. Phosphorylation of tyrosine residues 31 and 118 on paxillin regulates cell migration through an association with CRK in NBT-II cells. *J Cell Biol,* 148**,** 957-70.

PEUKERT, D., WEBER, S., LUMSDEN, A. & SCHOLPP, S. 2011. Lhx2 and Lhx9 determine neuronal differentiation and compartition in the caudal forebrain by regulating Wnt signaling. *PLoS Biol,* 9**,** e1001218.

PINCO, K. A., LIU, S. & YANG, J. T. 2001. alpha4 integrin is expressed in a subset of cranial neural crest cells and in epicardial progenitor cells during early mouse development. *Mech Dev,* 100**,** 99-103.

PLAVICKI, J. S., HOFSTEEN, P., YUE, M. S., LANHAM, K. A., PETERSON, R. E. & HEIDEMAN, W. 2014. Multiple modes of proepicardial cell migration require heartbeat. *BMC Dev Biol,* 14**,** 18.

POMBAL, M. A., CARMONA, R., MEGIAS, M., RUIZ, A., PEREZ-POMARES, J. M. & MUNOZ-CHAPULI, R. 2008. Epicardial development in lamprey supports an evolutionary origin of the vertebrate epicardium from an ancestral pronephric external glomerulus. *Evol Dev,* 10**,** 210-6.

RATAJSKA, A., CZARNOWSKA, E. & CISZEK, B. 2008. Embryonic development of the proepicardium and coronary vessels. *Int J Dev Biol,* 52**,** 229-36.

RETAUX, S., ROGARD, M., BACH, I., FAILLI, V. & BESSON, M. J. 1999. Lhx9: a novel LIM-homeodomain gene expressed in the developing forebrain. *J Neurosci,* 19**,** 783-93.

RICHARDS, K. L., ZHANG, B., SUN, M., DONG, W., CHURCHILL, J., BACHINSKI, L. L., WILSON, C. D., BAGGERLY, K. A., YIN, G., HAYES, D. N., WISTUBA, II & KRAHE,

R. 2011. Methylation of the candidate biomarker TCF21 is very frequent across a spectrum of early-stage nonsmall cell lung cancers. *Cancer,* 117**,** 606-17.

RICHARDSON, A., MALIK, R. K., HILDEBRAND, J. D. & PARSONS, J. T. 1997. Inhibition of cell spreading by expression of the C-terminal domain of focal adhesion kinase (FAK) is rescued by coexpression of Src or catalytically inactive FAK: a role for paxillin tyrosine phosphorylation. *Mol Cell Biol,* 17**,** 6906-14.

ROGER, V. L., GO, A. S., LLOYD-JONES, D. M., BENJAMIN, E. J., BERRY, J. D., BORDEN, W. B., BRAVATA, D. M., DAI, S., FORD, E. S., FOX, C. S., FULLERTON, H. J., GILLESPIE, C., HAILPERN, S. M., HEIT, J. A., HOWARD, V. J., KISSELA, B. M., KITTNER, S. J., LACKLAND, D. T., LICHTMAN, J. H., LISABETH, L. D., MAKUC, D. M., MARCUS, G. M., MARELLI, A., MATCHAR, D. B., MOY, C. S., MOZAFFARIAN, D., MUSSOLINO, M. E., NICHOL, G., PAYNTER, N. P., SOLIMAN, E. Z., SORLIE, P. D., SOTOODEHNIA, N., TURAN, T. N., VIRANI, S. S., WONG, N. D., WOO, D. & TURNER, M. B. 2012. Heart disease and stroke statistics--2012 update: a report from the American Heart Association. *Circulation,* 125**,** e2-e220.

RONGISH, B. J., HINCHMAN, G., DOTY, M. K., BALDWIN, H. S. & TOMANEK, R. J. 1996. Relationship of the extracellular matrix to coronary neovascularization during development. *J Mol Cell Cardiol,* 28**,** 2203-15.

ROSE, D. M. 2006. The role of the alpha4 integrin-paxillin interaction in regulating leukocyte trafficking. *Exp Mol Med,* 38**,** 191-5.

ROSE, D. M., LIU, S., WOODSIDE, D. G., HAN, J., SCHLAEPFER, D. D. & GINSBERG, M. H. 2003. Paxillin binding to the alpha 4 integrin subunit stimulates LFA-1 (integrin alpha L beta 2)-dependent T cell migration by augmenting the activation of focal adhesion kinase/proline-rich tyrosine kinase-2. *J Immunol,* 170**,** 5912-8.

SACHDEV, S., BU, Y. & GELMAN, I. H. 2009. Paxillin-Y118 phosphorylation contributes to the control of Src-induced anchorage-independent growth by FAK and adhesion. *BMC Cancer,* 9**,** 12.

SCHLUETER, J. & BRAND, T. 2009. A right-sided pathway involving FGF8/Snai1 controls asymmetric development of the proepicardium in the chick embryo. *Proc Natl Acad Sci U S A,* 106**,** 7485-90.

SCHLUETER, J. & BRAND, T. 2013. Subpopulation of proepicardial cells is derived from the somatic mesoderm in the chick embryo. *Circ Res,* 113**,** 1128-37.

SCHULTE, I., SCHLUETER, J., ABU-ISSA, R., BRAND, T. & MANNER, J. 2007. Morphological and molecular left-right asymmetries in the development of the proepicardium: a comparative analysis on mouse and chick embryos. *Dev Dyn,* 236**,** 684-95.

SENGBUSCH, J. K., HE, W., PINCO, K. A. & YANG, J. T. 2002. Dual functions of [alpha]4[beta]1 integrin in epicardial development: initial migration and long-term attachment. *J Cell Biol,* 157**,** 873-82.

SERLUCA, F. C. 2008. Development of the proepicardial organ in the zebrafish. *Dev Biol,* 315**,** 18-27.

SMAGULOVA, F. O., MANUYLOV, N. L., LEACH, L. L. & TEVOSIAN, S. G. 2008. GATA4/FOG2 transcriptional complex regulates Lhx9 gene expression in murine heart development. *BMC Dev Biol,* 8**,** 67.

SMALHEISER, N. R. & SCHWARTZ, N. B. 1987. Cranin: a laminin-binding protein of cell membranes. *Proc Natl Acad Sci U S A,* 84**,** 6457-61.

STRAIGHT, A. F., CHEUNG, A., LIMOUZE, J., CHEN, I., WESTWOOD, N. J., SELLERS, J. R. & MITCHISON, T. J. 2003. Dissecting temporal and spatial control of cytokinesis with a myosin II Inhibitor. *Science,* 299**,** 1743-7.

STUCKMANN, I. & LASSAR, A. B. 2002. Erythropoietin and retinoic acid signaling in the epicardium is required for cardiac myocyte proliferation. *Cold Spring Harb Symp Quant Biol,* 67**,** 45-8.

TANDON, P., MITEVA, Y. V., KUCHENBROD, L. M., CRISTEA, I. M. & CONLON, F. L. 2013. Tcf21 regulates the specification and maturation of proepicardial cells. *Development,* 140**,** 2409-21.

TANDON, P., SHOWELL, C., CHRISTINE, K. & CONLON, F. L. 2012. Morpholino injection in Xenopus. *Methods Mol Biol,* 843**,** 29-46.

TURNER, C. E. 1991. Paxillin is a major phosphotyrosine-containing protein during embryonic development. *J Cell Biol,* 115**,** 201-7.

TURNER, C. E. 1998. Paxillin. *Int J Biochem Cell Biol,* 30**,** 955-9.

TZCHORI, I., DAY, T. F., CAROLAN, P. J., ZHAO, Y., WASSIF, C. A., LI, L., LEWANDOSKI, M., GORIVODSKY, M., LOVE, P. E., PORTER, F. D., WESTPHAL, H. & YANG, Y. 2009. LIM homeobox transcription factors integrate signaling events that control three-dimensional limb patterning and growth. *Development,* 136**,** 1375-85.

UNO, Y., NISHIDA, C., TAKAGI, C., UENO, N. & MATSUDA, Y. 2013. Homoeologous chromosomes of Xenopus laevis are highly conserved after whole-genome duplication. *Heredity (Edinb),* 111**,** 430-6.

VICZIAN, A. S., BANG, A. G., HARRIS, W. A. & ZUBER, M. E. 2006. Expression of Xenopus laevis Lhx2 during eye development and evidence for divergent expression among vertebrates. *Dev Dyn,* 235**,** 1133-41.

VIRAGH, S., GITTENBERGER-DE GROOT, A. C., POELMANN, R. E. & KALMAN, F. 1993. Early development of quail heart epicardium and associated vascular and glandular structures. *Anat Embryol (Berl),* 188**,** 381-93.

VLADIMIROVA, V., MIKESKA, T., WAHA, A., SOERENSEN, N., XU, J., REYNOLDS, P. C. & PIETSCH, T. 2009. Aberrant methylation and reduced expression of LHX9 in malignant gliomas of childhood. *Neoplasia,* 11**,** 700-11.

VON GISE, A. & PU, W. T. 2012. Endocardial and epicardial epithelial to mesenchymal transitions in heart development and disease. *Circ Res,* 110**,** 1628-45.

WALLINGFORD, J. B. 2010. Preparation of fixed Xenopus embryos for confocal imaging. *Cold Spring Harb Protoc,* 2010**,** pdb prot5426.

WANG, J., KARRA, R., DICKSON, A. L. & POSS, K. D. 2013. Fibronectin is deposited by injury-activated epicardial cells and is necessary for zebrafish heart regeneration. *Dev Biol,* 382**,** 427-35.

WEI, K., DIAZ-TRELLES, R., LIU, Q., DIEZ-CUNADO, M., SCIMIA, M. C., CAI, W., SAWADA, J., KOMATSU, M., BOYLE, J. J., ZHOU, B., RUIZ-LOZANO, P. & MERCOLA, M. 2015. Developmental origin of age-related coronary artery disease. *Cardiovasc Res,* 107**,** 287-94.

WEISS, D., STOCKMANN, C., SCHRODTER, K. & RUDACK, C. 2013. Protein expression and promoter methylation of the candidate biomarker TCF21 in head and neck squamous cell carcinoma. *Cell Oncol (Dordr),* 36**,** 213-24.

WILSON, S. I., SHAFER, B., LEE, K. J. & DODD, J. 2008. A molecular program for contralateral trajectory: Rig-1 control by LIM homeodomain transcription factors. *Neuron,* 59**,** 413-24.

WINTER, E. M. & GITTENBERGER-DE GROOT, A. C. 2007. Epicardium-derived cells in cardiogenesis and cardiac regeneration. *Cell Mol Life Sci,* 64**,** 692-703.

WINTER, E. M., VAN OORSCHOT, A. A., HOGERS, B., VAN DER GRAAF, L. M., DOEVENDANS, P. A., POELMANN, R. E., ATSMA, D. E., GITTENBERGER-DE GROOT, A. C. & GOUMANS, M. J. 2009. A new direction for cardiac regeneration therapy: application of synergistically acting epicardium-derived cells and cardiomyocyte progenitor cells. *Circ Heart Fail,* 2**,** 643-53.

WOLFENSON, H., LAVELIN, I. & GEIGER, B. 2013. Dynamic regulation of the structure and functions of integrin adhesions. *Dev Cell,* 24**,** 447-58.

WU, C., FIELDS, A. J., KAPTEIJN, B. A. & MCDONALD, J. A. 1995. The role of alpha 4 beta 1 integrin in cell motility and fibronectin matrix assembly. *J Cell Sci,* 108 (Pt 2)**,** 821-9.

YANG, J. T., RAYBURN, H. & HYNES, R. O. 1995. Cell adhesion events mediated by alpha 4 integrins are essential in placental and cardiac development. *Development,* 121**,** 549-60.

YANG, Y. & WILSON, M. J. 2015. Lhx9 gene expression during early limb development in mice requires the FGF signalling pathway. *Gene Expr Patterns*.

YANG, Z., LI, D. M., XIE, Q. & DAI, D. Q. 2014. Protein expression and promoter methylation of the candidate biomarker TCF21 in gastric cancer. *J Cancer Res Clin Oncol*.

YE, Y. W., JIANG, Z. M., LI, W. H., LI, Z. S., HAN, Y. H., SUN, L., WANG, Y., XIE, J., LIU, Y. C., ZHAO, J., TANG, A. F., LI, X. X., GUAN, Z. C., GUI, Y. T. & CAI, Z. M. 2012. Down-regulation of TCF21 is associated with poor survival in clear cell renal cell carcinoma. *Neoplasma,* 59**,** 599-605.

ZAIDEL-BAR, R., MILO, R., KAM, Z. & GEIGER, B. 2007. A paxillin tyrosine phosphorylation switch regulates the assembly and form of cell-matrix adhesions. *J Cell Sci,* 120**,** 137-48.

ZHANG, H., GUO, Y., SHANG, C., SONG, Y. & WU, B. 2012. miR-21 downregulated TCF21 to inhibit KISS1 in renal cancer. *Urology,* 80**,** 1298-302 e1.

APPENDIX 3: A Diverged Cardiac Program in Vertebrates

Introduction

The pathologies and ultimately treatment of human heart disease has relied on vertebrate model systems with a dual circulatory system including *Xenopus laevis*, *Xenopus tropicalis*, *Mus musculus*, and *Sus scrofa*. However, which protein and protein pathways are conserved and which have diverged between these species and humans has not been established. Here we conduct an evolutionary comparative analysis of these cardiac proteomes to that of human. We find that all species with a pulmonary and systemic circulatory system contain a conserved set of protein and protein pathways that correlate with a common cardiac set of metabolic pathways. We further find mammalian and amphibian specific as well as species specific enriched proteins and protein pathways. Surprisingly, these include pathways that are conserved between individual species and those involved in human heart disease. Integration of these finding with those proteins causative to cardiac disease will therefore facilitate the optimization of species specific models for distinct disease states.

Results

Identification of a vertebrate core cardiac program

To define those protein and protein pathways that are conserved and diverged in abundance between *Mus musculus, Sus scrofa, Xenopus laevis,* and *Xenopus tropicalis* and human, we performed a differential extraction protocol of heart tissue (Fig A3.1A)(Fig A3.1B, Fig SA3.1). The resulting three extracts were fractionated by SDS

PAGE, generating 15 fractions per biological replicates and analyzed them by LC-MS/MS. To ensure assignment to different species, identified proteins were assembled and clustered in Scaffold. In total, 9067 protein clusters were observed, with approximately 4500 proteins identified for each species in each replicate (Fig SA3.2). Quantitation was performed using summed abundance-based precursor data with median normalized across replicates. Principle component analysis (PCA) (Fig A3.1C, Fig SA3.2) revealed *M. musculus* and *S. scrofa* samples separated while the two *Xenopus* species remained clustered. Multiscatter plot analysis confirmed that the two *Xenopus* species had the highest degree of similarity (Pearson correlation of 0.83, Fig SA3.4). Strikingly, signatures of the proteomes were still observed when the PCA was conducted using different extract samples as inputs verses summed data from all extracts (Fig SA3.3) To identify shared protein and protein pathways**,** we removed all remaining redundancies (i.e., where products of the same gene might exist in two or more clusters for the different species) (Fig A3.1D, Fig A3.2A, Fig SA3.5) resulting in 4710 quantifiable proteins (Fig A3.2A) with 1770 (38%) found in all species. We find the core cardiac program shows an increase in abundance of proteins ($p \leq 0.05$) in pathways that include metabolism, translation, and cellular organization (Fig A3.2C).

Mammalian and Xenopus hearts pathways that have diverged

In addition to the identification of a core cardiac program, our results identified differential species' enrichments (Fig A3.2B). This analysis revealed sets of proteins shared between mammals (cluster 3) and a second set shared between *Xenopus* species (cluster 4). The cluster showing commonalities among mammalian proteomes

contained enrichments in metabolic processes related to lipid and carbohydrate metabolism, while the *Xenopus* proteomes revealed enrichments in cell cycle proteins, intracellular signal transduction, and cell communication (Fig A3.2C). To determine functional differences of mammalian verses *Xenopus* cardiac proteomes we used ClueGO, (Fig A3.3B) which revealed 63 terms to be significantly enriched (adjusted p ≤0.01). The results of this analysis show increased numbers of metabolic and mitochondrial proteins for the mammals, consistent with energy production being more critical and more complex four-chambered hearts versed the three chambered heart of frogs.

Species specific pathways

Our analysis revealed a high degree of species specific proteins and protein pathways. For example, *M. musculus*-driven cluster was enriched in vesicle mediated transport functions, the *S. scrofa* enriched for mitochondrial and transcriptional categories, the *X. tropicalis* proteome was enriched in RNA processing, while *X. laevis* had enrichment in cell adhesion proteins (Fig A3.3A). Analysis further revealed species-specific abundance of protein and protein pathways (Fig A3.3A, clusters in Fig A3.2B). For example, the *M. musculus* proteome had enrichments for immune signaling pathways, while *X. laevis* had further enrichments for RNA related processes.

Cell cycle enrichment in Xenopus laevis

One surprising finding is protein and protein pathways have diverged between *X. tropicalis* and *X. laevis*. One of the most striking was a *X. laevis* enrichment in cell cycle

proteins (Fig A3.4A) which fell into eleven different GO categories all being significantly higher in *X. laevis* across all species in three pairwise comparisons (Fig A3.4B, top, Fig SA3.6). These proteins were found to be in a diverse set of cell cycle pathways including cell cycle checkpoints, G2 and M transition, cell cycle regulation and cell cycle signaling. Though enrichment in *X. laevis* was consistent and the most prominent difference by GSEA (Fig SA3.7) we find not all proteins were uniformly higher in *X. laevis* (Fig A3.4B, bottom). For example, mitochondrial proteins were enriched in *S. scrofa* compared to the other species. Thus changes in cardiac protein abundance in a species specific manner with the most pronounced being an increase in cell cycle proteins in *X. laevis*.

Targeted mass spectrometry validation of cell cycle protein enrichment

To validate our discovery of enriched cell cycle proteins in *X. laevis*, we performed an orthogonal assay based on quantitative targeted mass spectrometry. Our analysis focused on cell cycle proteins identified as enriched in *X laevis* by at least 5 fold in comparison to other species. To minimize sample processing and improve accuracy of quantification, we developed a parallel reaction monitoring (PRM) assay (Fig A3.5A). A key component of this assay was the presence of one or more tryptic peptides shared across all four species. Following an iterative optimization process, we generated a final list of 33 high-confidence cell cycle protein targets. The scheduled PRM assays were run on a new set of tissues in three biological replicates, and loading was normalized based on the average MS1 signals. The resulting peak areas for each protein are illustrated as grouped by functional categories (Fig A3.5B). The

quantification graphs for individual peptides for each protein are illustrated in Fig SA3.8. Two-thirds of the proteins targeted had a higher average peak area in *X. laevis* (shown in bold) than the other three species, confirming our finding of enriched cell cycle proteins.

Choice of species in human disease modeling

Given the observation that the cardiac proteome of vertebrates has diverged in the abundance of a subset of protein pathways, we queried which pathways are conserved between *Xenopus laevis*, *Xenopus tropicalis*, *Mus musculus*, and *Sus scrofa* and those associated with human disease (Fig A3.6). Our data shows an interconnectedness of heart disease proteins and further, allows for identification of hub proteins that are important in multiple disease states, i.e. TNNT2 (Troponin T2).

This work was performed in collaboration with Joel D. Federspiel, Laura E. Herring, Samvida Venkatesh, Lauren Wasson, Ileana M. Cristea, and Frank L. Conlon.

Material and Methods

Sample preparation for data dependent analysis (DDA)

Heart tissues were frozen in liquid nitrogen and cryogenically homogenized using a mortar and pestle. The *Sus scrofa* hearts were received (Pel-Freez) on wet ice, diced into pieces, snap frozen in liquid nitrogen and then blended in a stainless-steel Waring blender. For each species, frozen powder from the following number of hearts were combined equally for each biological replicate: *Mus musculus*: 3, *Xenopus laevis*: 3, *Xenopus tropicalis*: 7, *Sus scrofa*: 2. The mixed tissue powder was then dissolved in

homogenization buffer (10 mM Tris pH 7.4, 250 mM sucrose, 1 mM EDTA, 0.15% NP-40, 10 mM sodium butyrate, 1 mM PMSF, 1X protease inhibitors, and 1X phosphatase inhibitors). Insoluble material was pelleted out and the remaining supernatant (Cytoplasmic fraction) was put aside for further processing. The pellet (Nuclear fraction) was resuspended in lysis buffer (20 mM K-HEPES pH 7.4, 110 mM KOAc, 2 mM $MgCl_2$, 0.1% Tween-20, 1 µM $ZnCl_2$, 1 mM $CaCl_2$, 150 mM NaCl, 0.5% Triton X100 1X protease inhibitors, and 1X phosphatase inhibitors) and then homogenized via Polytron (Kinematica). Remaining insoluble material (Nuclear insoluble) was pelleted out and the supernatant (Nuclear Soluble) was subjected to acetone precipitation to concentrate the protein. The resulting acetone pellet was resuspended in lysis buffer (as above). The Nuclear Insoluble pellet was resuspended in lysis buffer (as above). Protein yields of all three subcellular fractions were assessed by Bradford assay and 25 µg of protein from each fraction was reduced with 25 mM iodoacetimde at RT in the dark for 1h. LDS buffer and 3 µL of reducing agent (Novex) was added to each sample and boiled for 10 min at 70°C. Each sample was then resolved by 4-12% Bis-Tris SDS PAGE in MOPS buffer for a length of 1.5 cm into the gel. The gels were then Coomassie stained, cut into 5 fractions and then digested overnight with trypsin (125 ng/gel slice) in 25 mM ammonium bicarbonate, pH 8.0. Peptides were extracted from the gel using 3 consecutive washes of 60% acetonitrile (ACN), 1% formic acid (FA), 39% H_2O with 20 min incubations. The extracted peptides were dried *in vacuo* and resuspended in 10 µL of 1% FA, 1% ACN, 98% H_2O for LC-MS/MS analysis.

Sample preparation for Parallel Reaction Monitoring (PRM) targeted analysis.

Heart tissues were thawed on ice, minced into small pieces, and then homogenized in 50 mM Tris-HCl, pH 8.0, 100 mM NaCl, 0.5 mM EDTA by douncing in a Tenbroeck homogenizer. The homogenate was then mixed 1:1 with the same buffer containing 4% SDS for a final SDS concentration of 2%. One whole heart per replicate was used apart from Sus scrofa hearts where 300 mg of ground powder (as above for the data dependent samples) from a single heart was used instead of intact heart tissue. The homogenized tissues in the SDS buffer were then heated at 95°C for 5 min and then sonicated in a cup horn sonicator (1s pulses, medium power) for 20s. The heating and sonication process was repeated three times until little or no insoluble material was visible. Samples were centrifuged to pellet any remaining insoluble material and a BCA assay (Pierce) was performed to assess protein concentration. 50 µg of protein from each sample was simultaneously reduced and alkylated with 20 mM tris(2-carboxyethyl)phosphine (Pierce) and chloroacetamide respectively for 20 min at 70°C. Protein samples were then cleaned up by methanol-chloroform precipitation. Briefly, LC-MS grade methanol, chloroform, and water (at a 4:1:3 ratio) were added to the sample with vortexing following each addition. The samples were spun at 2,000 × g for 5 min at room temperature and the top phase was removed. Three volumes of cold methanol were then added, and the samples were spun at 9,000 × g for 2 min at 4°C. All liquid was removed, and the protein pellets were washed with five volumes of cold methanol and then spun at 9,000 × g for 2 min at 4°C. All liquid was removed again, and the dried protein pellets were resuspended in 50 mM HEPES pH 8.5 at a 0.5 µg/µL concentration. Trypsin (Pierce) was added at a 1:50 protease:protein ratio and the

samples were incubated at 37°C overnight. Peptides were cleaned up via SDB-RPS StageTip as described (Sauls, et al., 2018), briefly, peptides were acidified, bound to StageTips, washed with 0.5% trifluoroacetic acid in H_2O, and eluted directly into autosampler vials with 5% ammonium hydroxide, 80% ACN, 15% H_2O. Peptides were dried *in vacuo* and resuspended in 1% ACN, 1% FA, 98% H_2O and peptide concentrations were assessed by A205 assay on a Nanodrop One (Thermo Scientific). Samples were adjusted to equal concentrations and analyzed by LC-MS/MS.

Mass spectrometry analyses for DDA

Peptides (6µL) were analyzed by LC-MS/MS using an EasyNano nLC1000 UPLC (Thermo Scientific) coupled online to an EASYSpray ion source and Q Exactive HF (Thermo Scientific). Peptides were separated on an EASYSpray C18 column (75µm x 50cm) heated to 50°C using a gradient of solvents A (0.1% FA in H_2O) and B (0.1% FA, 2.9% H_2O in ACN) at a flow rate of 250 nL/min as follows: 5% B to 20% B over 40 min, then to 32% B in 11 min, then to 90% B in 2 min, and hold at 90% B for 12 min. Peptides were ionized at 2.5 kV and an MS1 survey scan was performed from 400 to 1600 m/z at 120,000 resolution with an automatic gain control (AGC) setting of 3e6 and a maximum injection time (MIT) of 100ms recorded in profile. The top 15 precursors were then selected for fragmentation and MS2 scans were acquired at a resolution of 15,000 with an AGC setting of 2e4, a MIT of 50ms, an isolation window of 1.6 m/z, normalized collision energy of 27, intensity threshold of 2e4, peptide match set to preferred, and a dynamic exclusion of 30s recorded in profile. MS/MS data were analyzed by Proteome Discoverer (Thermo Fisher Scientific, v2.1.1.21). A fully tryptic

Sequest HT search against a Uniprot database containing human, *Mus musculus*, *Xenopus laevis*, *Xenopus tropicalis*, and *Sus scrofa* sequences appended with common contaminants (141,645 entries, downloaded 2016_12) was conducted. Sequest settings required peptides to have 5ppm accuracy on the precursor and 0.02Da accuracy on the fragments. Allowed peptide modifications included static carbamidomethyl modifications to cysteine, dynamic oxidation of methionine, dynamic deamidation of asparagine, and dynamic methionine loss and acetylation of protein n-termini. The precursor ions area detector node in Proteome Discoverer was used to allow MS1 quantitation. The resulting peptide spectrum matches were imported into Scaffold (Proteome Software, v4.7.3) with an X!Tandem rescore with Phosphorylations on STY included as additional modifications. The protein clustering algorithm in Scaffold was used and the data was assembled at a 2 peptide/protein level with a 99.9% protein probability and 96% peptide probability. This resulted in a final protein and peptide FDR of 0.9% and 0.11% respectively. The total (summed abundance) precursor intensity quantitation in Scaffold was utilized for all subsequent analysis.

Targeted quantification by PRM

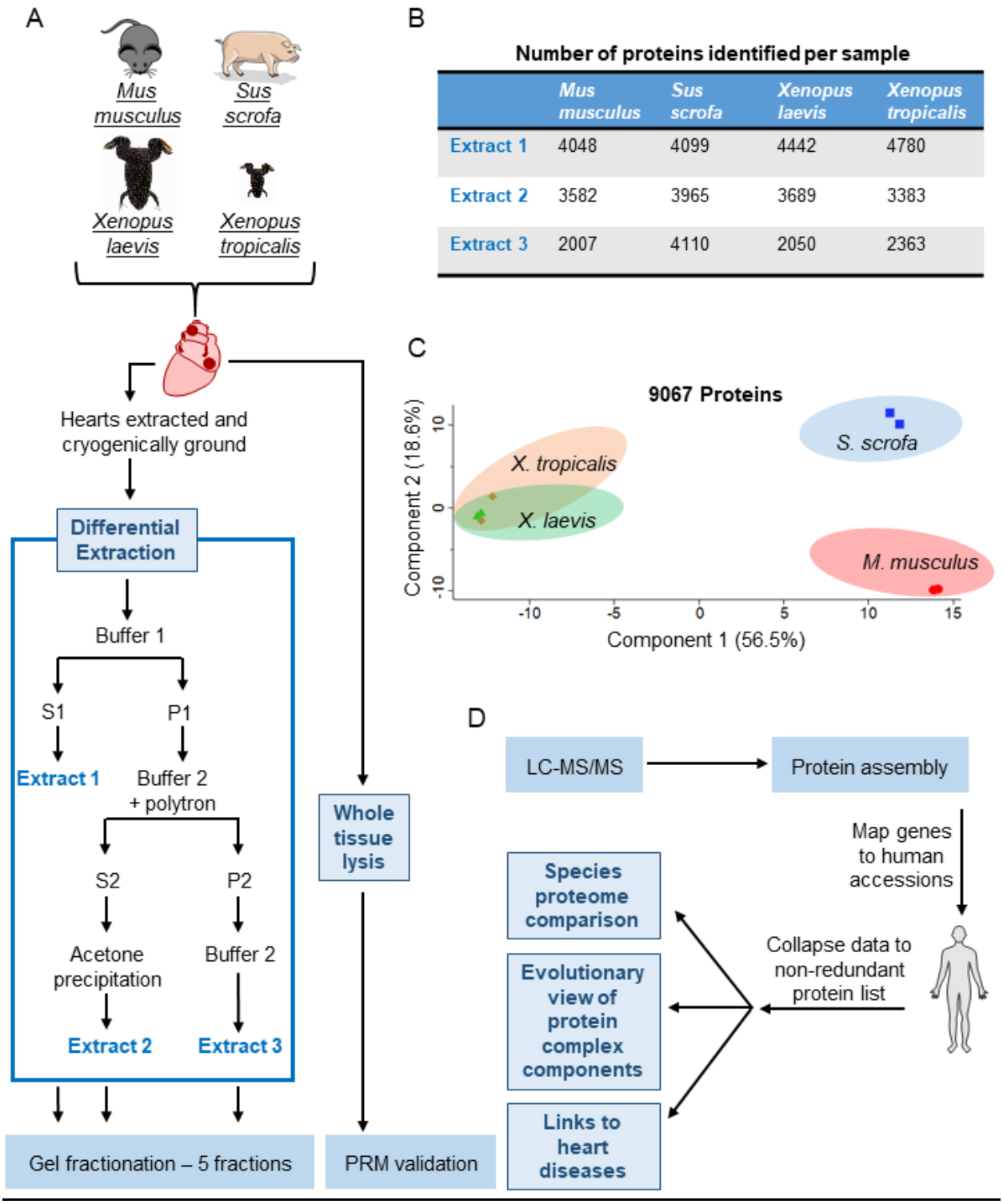
Peptides (2µL) were analyzed by LC-MS/MS using a Dionex Ultimate 3000 UPLC (Thermo Scientific) coupled online to an EASYSpray ion source and Q Exactive HF. Peptides were separated on an EASYSpray C18 column (75µm x 25cm) heated to 50°C at a flow rate of 250nL/min as follows: 1% B to 4% B over 9 min, then to 14% B over 40 min, then to 25% B over 20 min, then to 50% B over 4 min, then to 97% B over 0.5 min, then hold at 97% B for 6 min, then to 70% B over 0.5 min. A peptide inclusion

list was generated with retention time windows of 5 minutes used to isolate precursors for fragmentation. Peptides were ionized at 1.7 kV and a scheduled PRM method was run where MS2 scans were acquired at a resolution of 30,000 with an AGC setting of 2e5, a MIT of 150ms, an isolation window of 0.8 m/z, fixed first mass of 125.0 m/z, and normalized collision energy of 27 recorded in profile. After every 15 MS2 scans, a single MS1 scan was performed from 380-1800 Da at a resolution of 15,000 and an AGC setting of 3e6. Label-free targeted quantitation of peptides specific for a protein of interest and with identical sequences between all four species was designed and analyzed using the Skyline software. Summed area under the curve of 3-4 transitions per peptide was used for quantitation. The peptide peak areas were normalized by the average MS1 intensity of that sample as reported by RawMeat (Vast Scientific).

Statistical analysis and data visualization

Biological duplicates of each species were analyzed by LC-MS/MS. Identified proteins were exported from Scaffold and median normalization of the MS1 abundance values was performed in Excel. The identified proteins were mapped to human accession numbers using Blast2GO (BioBam, v5.0.13) (Gotz, et al., 2008) by performing a local blastp search against a human proteome file containing 20,205 entries using default settings. The resulting mapped proteins contained multiple entries that mapped to the same human accession number and these entries were averaged together to attain a single set of MS1 abundance values for each mapped human protein. Differential proteins were analyzed via over representation analysis and gene set enrichment analysis using www.pantherdb.org (Mi, et al., 2016) for associated gene

ontology enrichments. Example proteins of different classes were graphed in GraphPad Prism v5.04 and heatmaps of protein classes were generated using Morpheus (Broad Institute). Principle component analysis was performed in Perseus (Tyanova, et al., 2016). Network diagrams were created in Cytoscape (Shannon, et al., 2003) using the Reactome (Wu, et al., 2010) and ClueGO (Bindea, et al., 2009) plugins. Protein complex clustering was performed by mapping the identified proteins to human protein complexes in CORUM and then calculating a mean complex score (the mean value of the precursor area in all quantified components). The scored complexes were then clustered in R using the hclust algorithm with the ward method. The resulting dendrogram was cut at a level that generated 7 clusters based on evaluation of the clustering via PCA. The clustered complexes were then assessed for functional enrichments.



	Mus musculus	*Sus scrofa*	*Xenopus laevis*	*Xenopus tropicalis*
Extract 1	4048	4099	4442	4780
Extract 2	3582	3965	3689	3383
Extract 3	2007	4110	2050	2363

Figure A3.1. Generating multispecies proteomes. Heart tissue from *M. musculus, S. scrofa, X. laevis,* and *X. tropicalis* were collected and subjected to differential protein extraction as detailed. The resulting three extracts were fractionated by SDS PAGE. (B)

Number of proteins identified in the five gel fractions for each extract per species. These numbers include proteins found in two or more extracts. (C) PCA analysis of MS1 proteome data demonstrates that the mammalian sample separate from each other and the *Xenopus* samples but the *Xenopus* samples maintain a tight association. (D) Data analysis workflow.

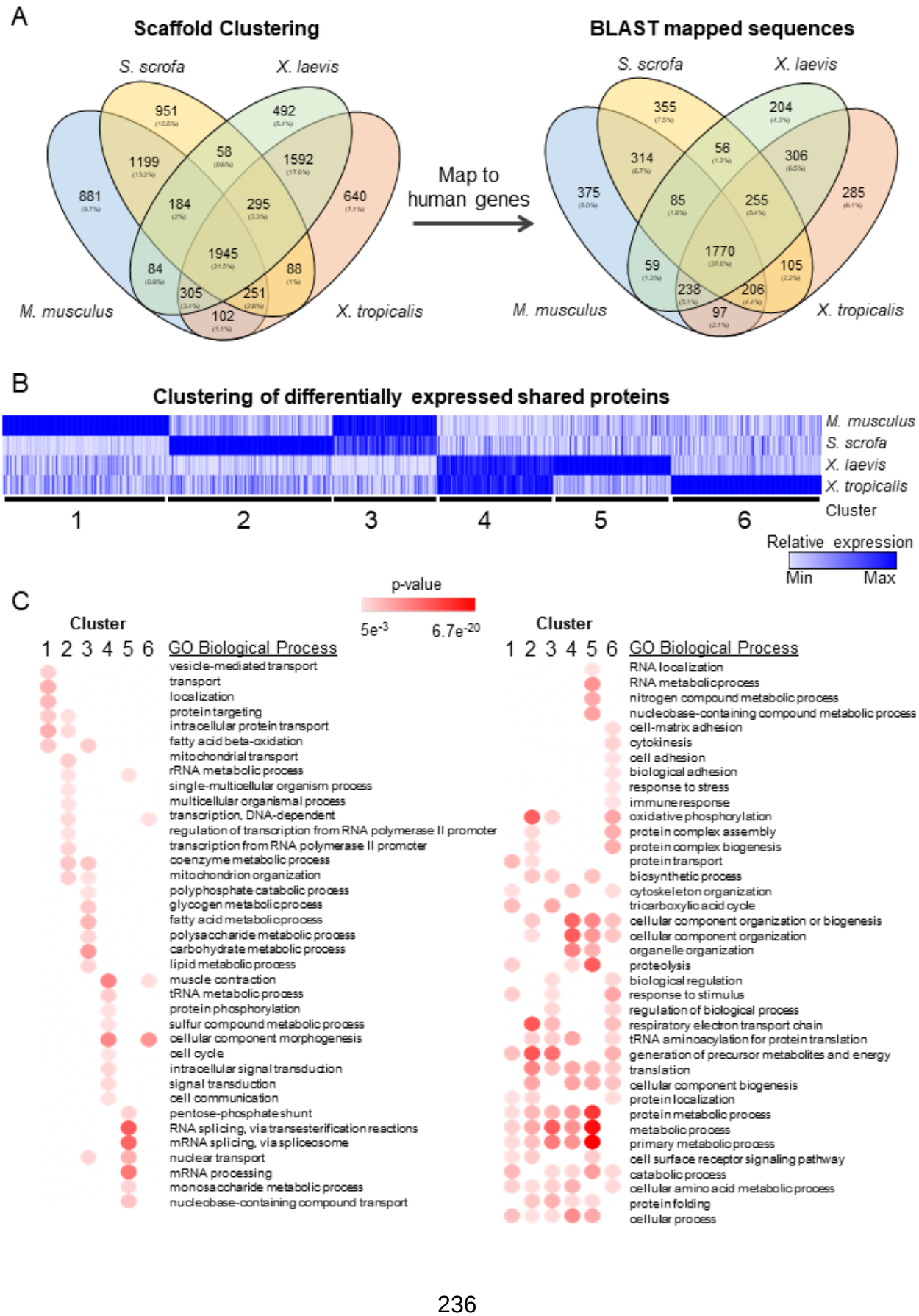

A
Scaffold Clustering
S. scrofa
X. laevis
951
492
58
1199
1592
881
184
295
640
1945
84
88
305
251
102
M. musculus
X. tropicalis
Map to human genes
BLAST mapped sequences
S. scrofa
X. laevis
355
204
314
56
306
375
85
255
285
1770
59
105
238
206
97
M. musculus
X. tropicalis
B
Clustering of differentially expressed shared proteins
M. musculus
S. scrofa
X. laevis
X. tropicalis
1
2
3
4
5
6
Cluster
Relative expression
Min
Max
C
p-value
5e-3
6.7e-20
Cluster
1 2 3 4 5 6
GO Biological Process
vesicle-mediated transport
transport
localization
protein targeting
intracellular protein transport
fatty acid beta-oxidation
mitochondrial transport
rRNA metabolic process
single-multicellular organism process
multicellular organismal process
transcription, DNA-dependent
regulation of transcription from RNA polymerase II promoter
transcription from RNA polymerase II promoter
coenzyme metabolic process
mitochondrion organization
polyphosphate catabolic process
glycogen metabolic process
fatty acid metabolic process
polysaccharide metabolic process
carbohydrate metabolic process
lipid metabolic process
muscle contraction
tRNA metabolic process
protein phosphorylation
sulfur compound metabolic process
cellular component morphogenesis
cell cycle
intracellular signal transduction
signal transduction
cell communication
pentose-phosphate shunt
RNA splicing, via transesterification reactions
mRNA splicing, via spliceosome
nuclear transport
mRNA processing
monosaccharide metabolic process
nucleobase-containing compound transport
Cluster
1 2 3 4 5 6
GO Biological Process
RNA localization
RNA metabolic process
nitrogen compound metabolic process
nucleobase-containing compound metabolic process
cell-matrix adhesion
cytokinesis
cell adhesion
biological adhesion
response to stress
immune response
oxidative phosphorylation
protein complex assembly
protein complex biogenesis
protein transport
biosynthetic process
cytoskeleton organization
tricarboxylic acid cycle
cellular component organization or biogenesis
cellular component organization
organelle organization
proteolysis
biological regulation
response to stimulus
regulation of biological process
respiratory electron transport chain
tRNA aminoacylation for protein translation
generation of precursor metabolites and energy
translation
cellular component biogenesis
protein localization
protein metabolic process
metabolic process
primary metabolic process
cell surface receptor signaling pathway
catabolic process
cellular amino acid metabolic process
protein folding
cellular process

Figure A3.2. Analysis of shared protein expression. (A) Identified proteins were mapped to human accession numbers using Blast2GO. (B) The 1770 proteins that are shared across all four species were clustered using k means with k=6 revealing clusters driven by higher protein expression in one or more species. Each of the clusters from the heatmap were analyzed for overrepresentation of GO biological process terms and the p-values of the significantly enriched terms were shown in a heatmap.

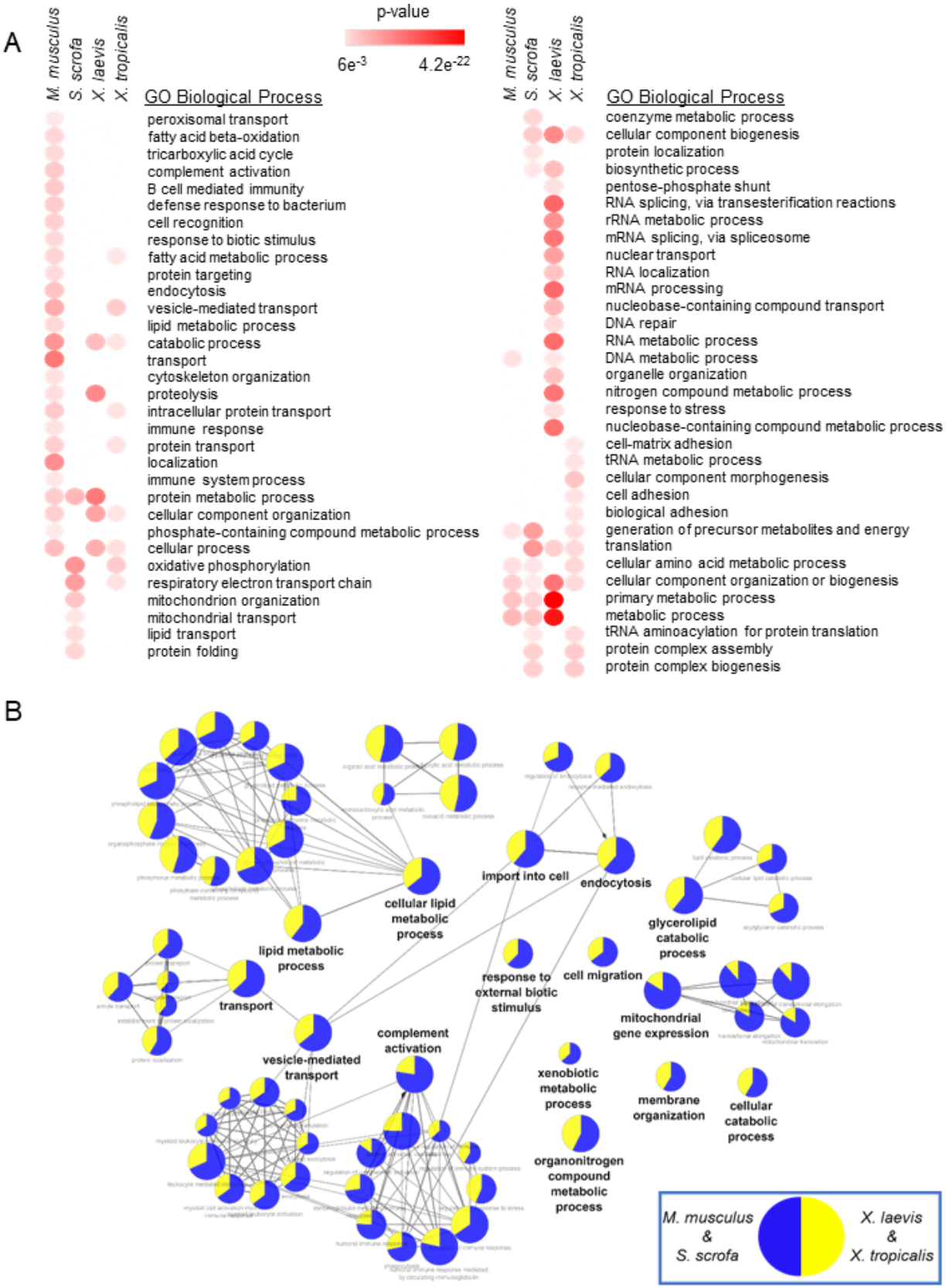

A
M. musculus
S. scrofa
X. laevis
X. tropicalis
p-value
6e-3
4.2e-22
GO Biological Process
peroxisomal transport
fatty acid beta-oxidation
tricarboxylic acid cycle
complement activation
B cell mediated immunity
defense response to bacterium
cell recognition
response to biotic stimulus
fatty acid metabolic process
protein targeting
endocytosis
vesicle-mediated transport
lipid metabolic process
catabolic process
transport
cytoskeleton organization
proteolysis
intracellular protein transport
immune response
protein transport
localization
immune system process
protein metabolic process
cellular component organization
phosphate-containing compound metabolic process
cellular process
oxidative phosphorylation
respiratory electron transport chain
mitochondrion organization
mitochondrial transport
lipid transport
protein folding
GO Biological Process
coenzyme metabolic process
cellular component biogenesis
protein localization
biosynthetic process
pentose-phosphate shunt
RNA splicing, via transesterification reactions
rRNA metabolic process
mRNA splicing, via spliceosome
nuclear transport
RNA localization
mRNA processing
nucleobase-containing compound transport
DNA repair
RNA metabolic process
DNA metabolic process
organelle organization
nitrogen compound metabolic process
response to stress
nucleobase-containing compound metabolic process
cell-matrix adhesion
tRNA metabolic process
cellular component morphogenesis
cell adhesion
biological adhesion
generation of precursor metabolites and energy
translation
cellular amino acid metabolic process
cellular component organization or biogenesis
primary metabolic process
metabolic process
tRNA aminoacylation for protein translation
protein complex assembly
protein complex biogenesis
B
cellular lipid metabolic process
lipid metabolic process
transport
vesicle-mediated transport
complement activation
import into cell
endocytosis
response to external biotic stimulus
cell migration
glycerolipid catabolic process
mitochondrial gene expression
xenobiotic metabolic process
membrane organization
cellular catabolic process
organonitrogen compound metabolic process
M. musculus & S. scrofa
X. laevis & X. tropicalis

Figure A3. 3. Analysis of species unique proteins. (A) Proteins found uniquely in one species only were combined with the cluster of proteins that were more highly expressed in that species and analyzed for overrepresentation of GO biological process terms. The p-values of the significantly enriched terms are shown in a heatmap. (B) The proteins uniquely identified in the mammalian and *Xenopus* species were used to examine evolutionary-driven protein expression differences in these organisms using ClueGO.

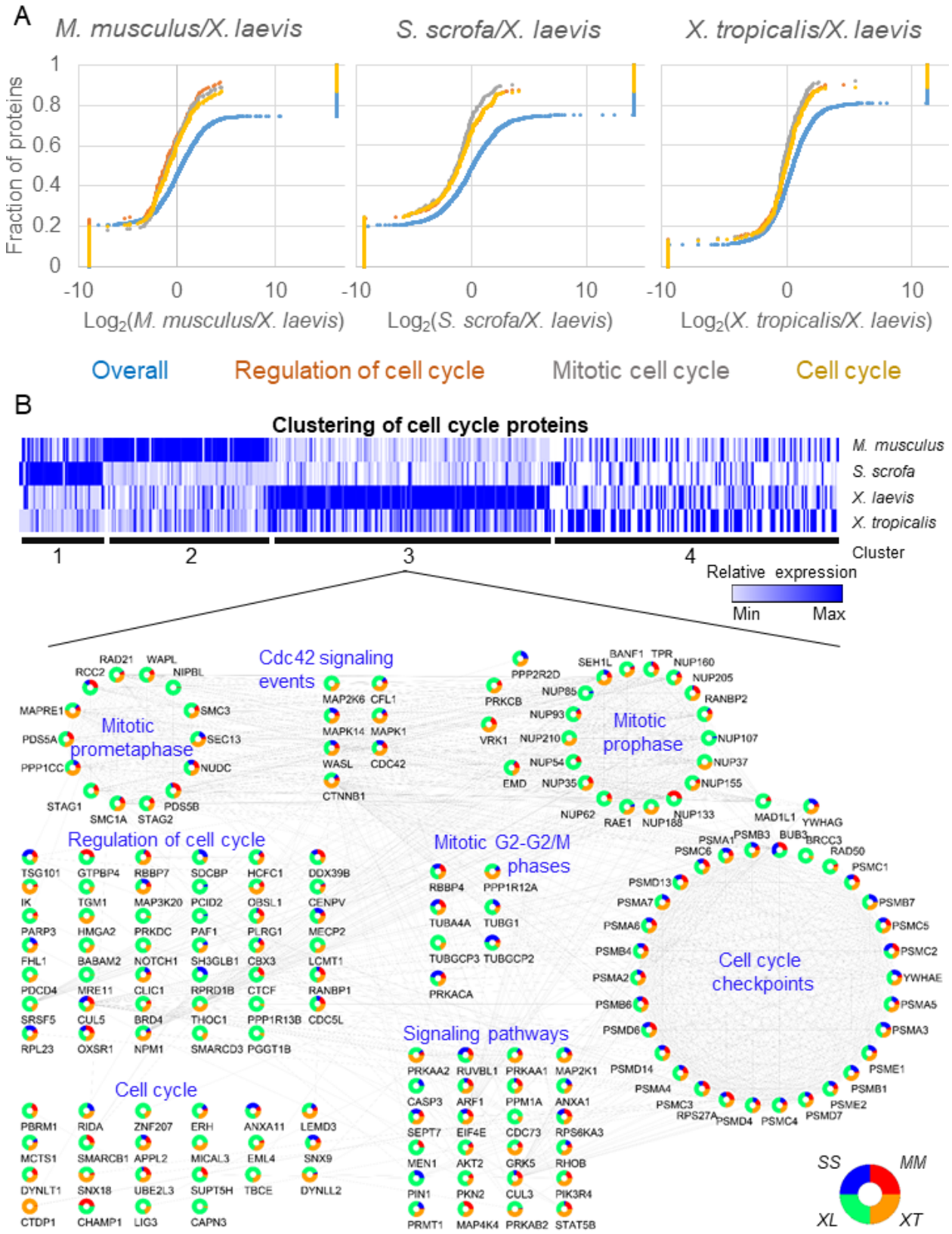

A
M. musculus/X. laevis
S. scrofa/X. laevis
X. tropicalis/X. laevis
Fraction of proteins
1
0.8
0.6
0.4
0.2
0
-10
0
10
Log2(M. musculus/X. laevis)
Log2(S. scrofa/X. laevis)
Log2(X. tropicalis/X. laevis)
Overall
Regulation of cell cycle
Mitotic cell cycle
Cell cycle
B
Clustering of cell cycle proteins
M. musculus
S. scrofa
X. laevis
X. tropicalis
1
2
3
4
Cluster
Relative expression
Min
Max
Cdc42 signaling events
Mitotic prometaphase
Mitotic prophase
Regulation of cell cycle
Mitotic G2-G2/M phases
Cell cycle checkpoints
Signaling pathways
Cell cycle
SS
MM
XL
XT

Figure A3.4. Gene set enrichment analysis reveals an increased presence of cell cycle proteins in *Xenopus laevis*. (A) Pairwise GSEA was carried out between *X. laevis* and the other three species which revealed an enrichment in cell cycle related proteins in *X. laevis*. (B) A heatmap of all proteins found in any enriched cell cycle related category by GSEA and clustered using kmeans. Relative protein expression is shown. The proteins found in cluster 2 were further analyzed in Cytoscape and grouped by function to show the interrelated nature of the proteins enriched in this cluster.

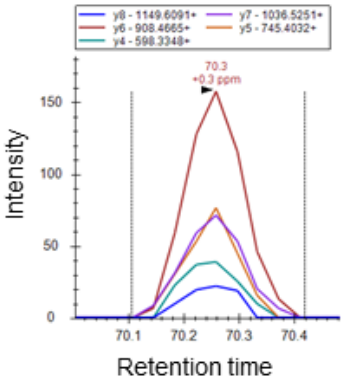
A

PRM validation of DDA results

33 cell cycle proteins from cluster 2 in Fig 4b with FC>5

Shared tryptic peptide(s) present in all four species

Peptide inclusion list



Detected signature fragment ions

Normalize to average MS1 peak area

Scale each peptide to the average peak area across all replicates

B

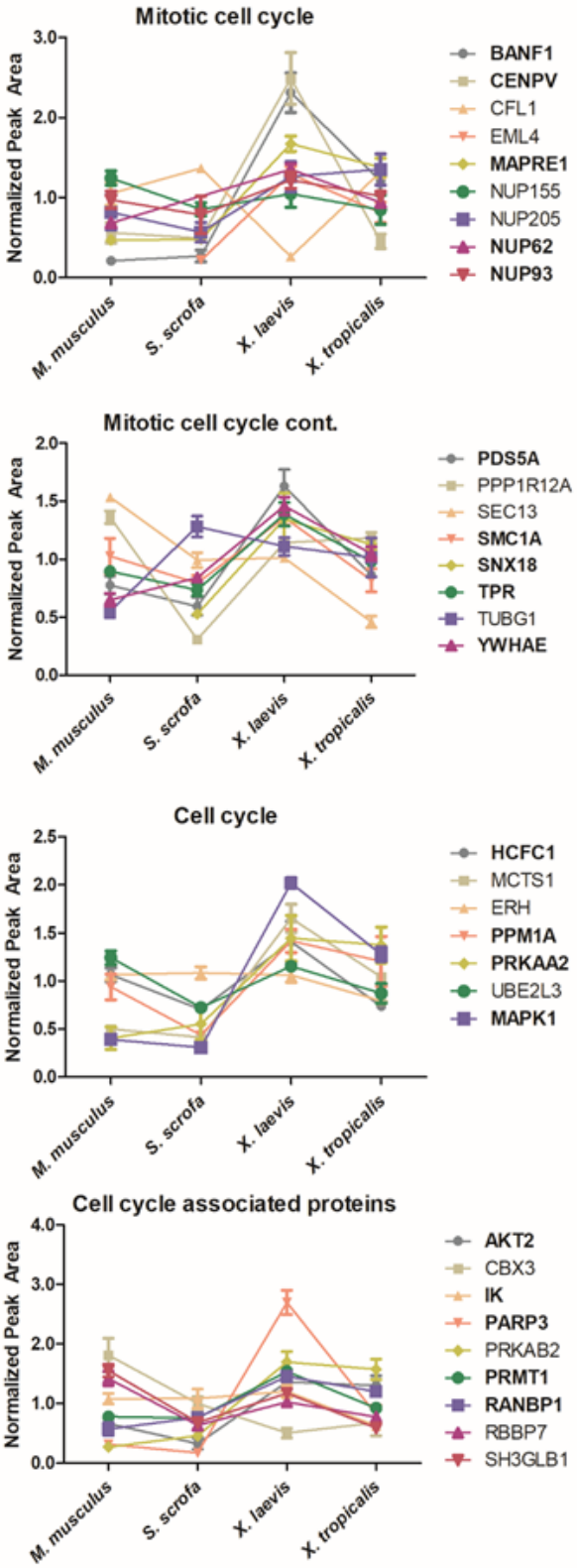
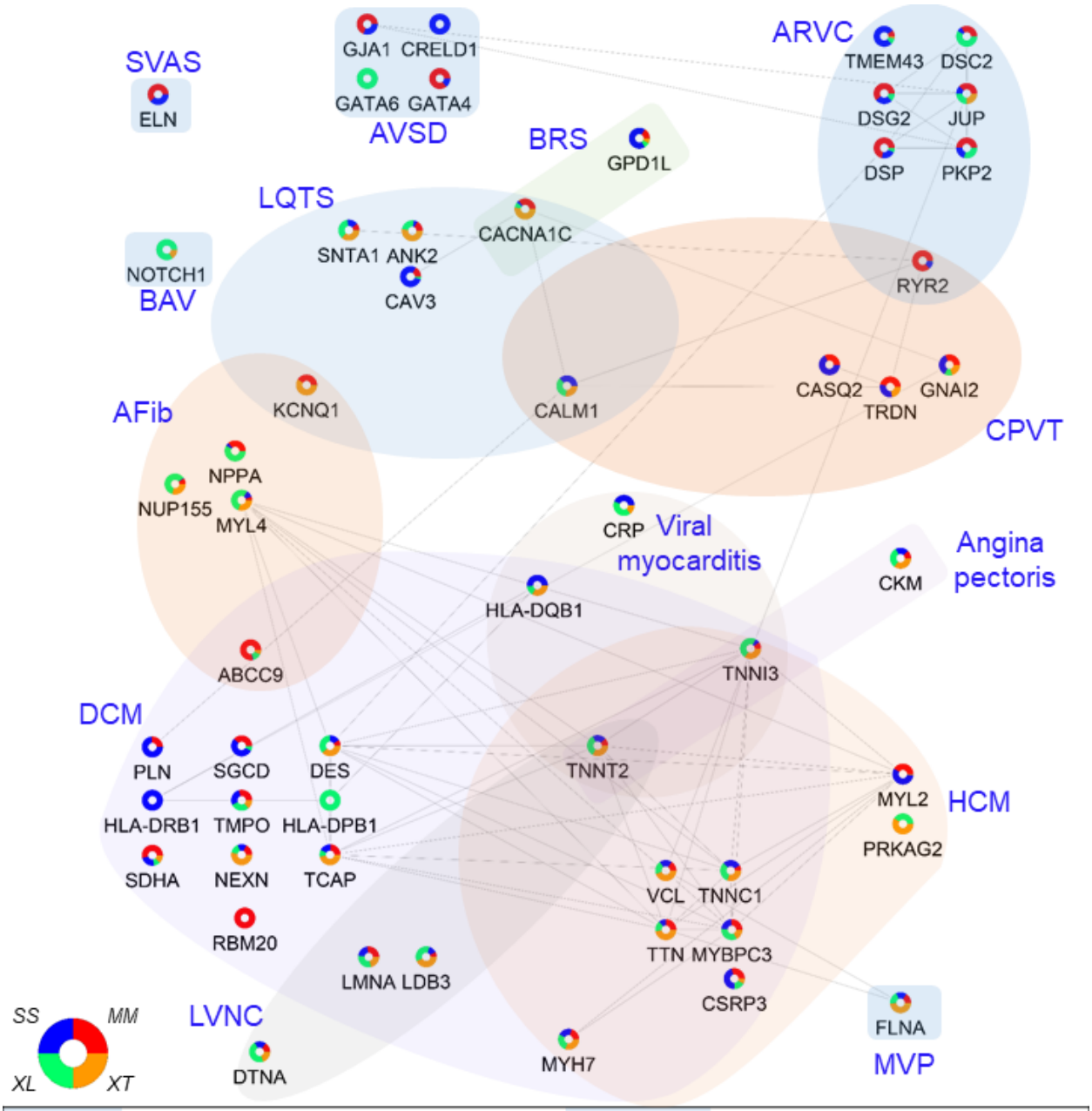


Figure A3.5. PRM validation of cell cycle enrichment in *X. laevis*. (A) Workflow of target protein and peptide selection for PRM validation. (B) Averaged protein expression with SEM in each species by PRM assay are shown. The data were grouped by protein functional categories and proteins graphed in blue have a higher average expression in *X. laevis* than all other species.



BAV	Bicuspid aortic valve	SVAS	Congenital supravalvar aortic stenosis
CPVT	Catecholaminergic polymorphic ventricular tachycardia	ARVC	Arrhythmogenic right ventricular cardiomyopathy
AVSD	Atrioventricular septal defect	MVP	Mitral valve prolapse
LQTS	Long QT syndrome	HCM	Hypertrophic cardiomyopathy
AFib	Atrial fibrillation	DCM	Dilated cardiomyopathy
BRS	Brugada syndrome	LVNC	Left ventricular noncompaction

Figure A3.6. Links to human heart disease. A diagram of detected proteins with known links to human heart disease was created in Cytoscape showing both relative protein expression data across the four species examined here as well as known protein-protein interactions between these heart disease related proteins.

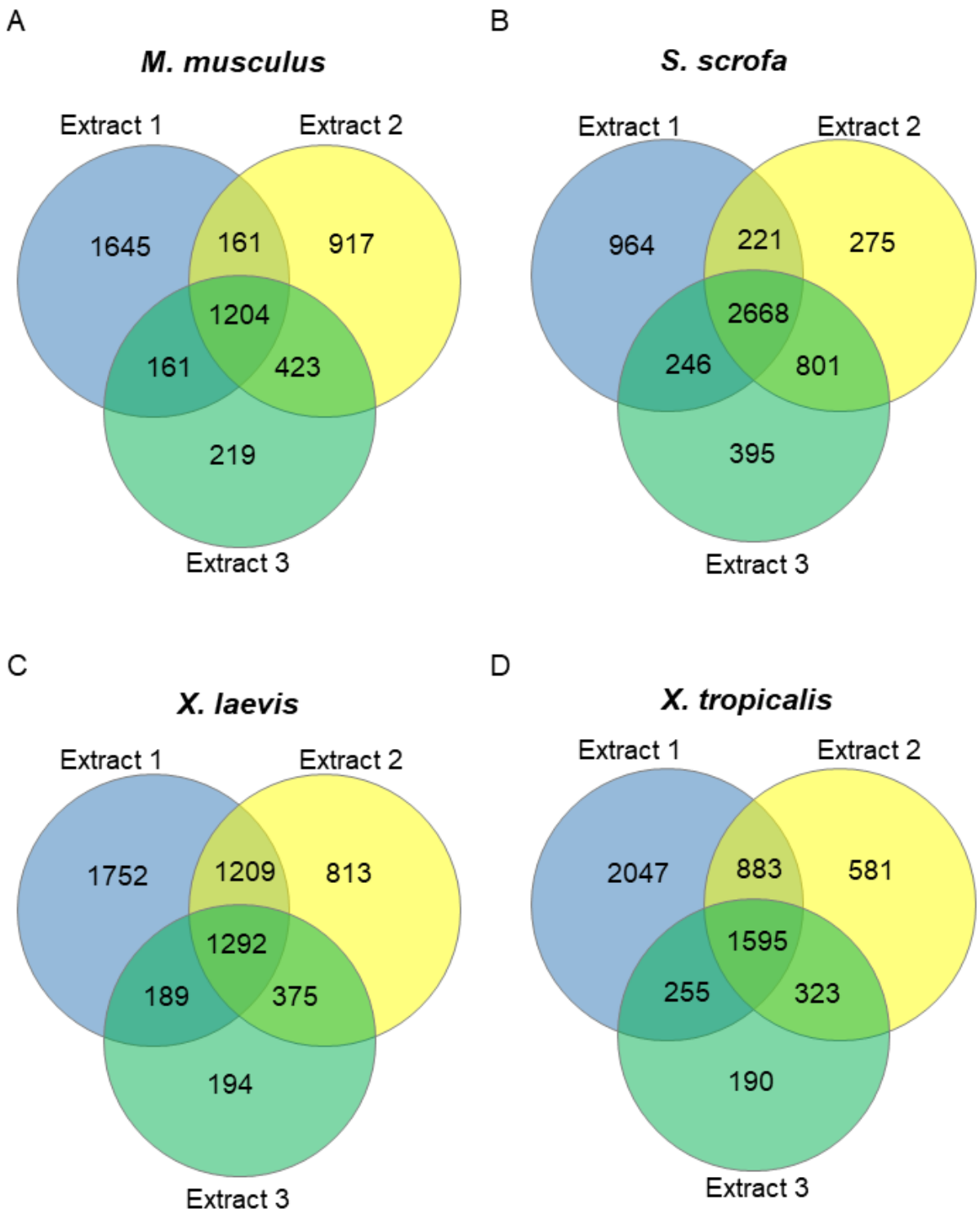


Figure SA3.1. Total proteins identified. Number of proteins identified in each extract for (A) M. musculus (B) S. scrofa (C) X. laevis (D) X. tropicalis.

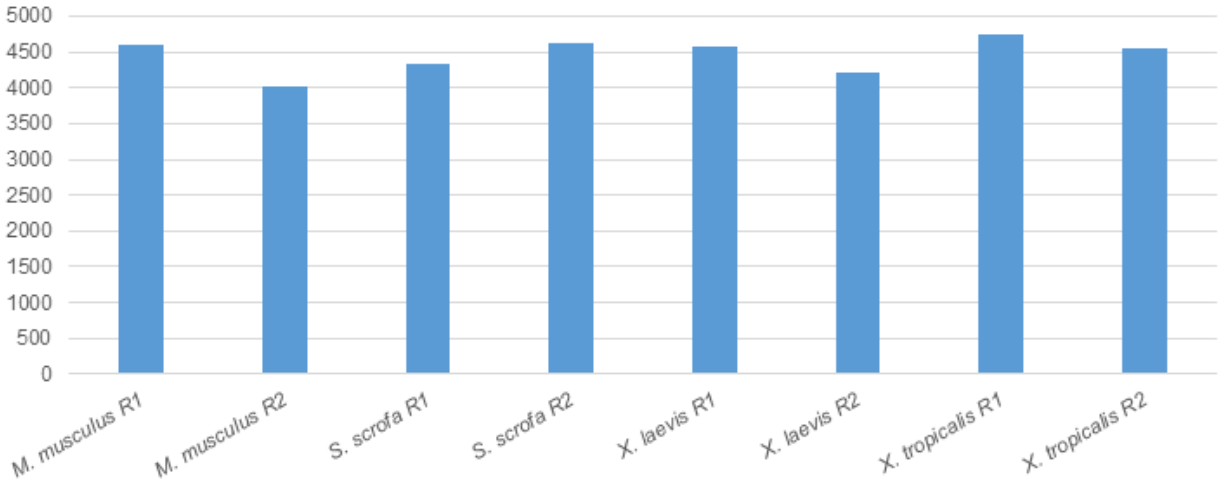


Figure SA3.2. Number of proteins identified in each replicate by species.

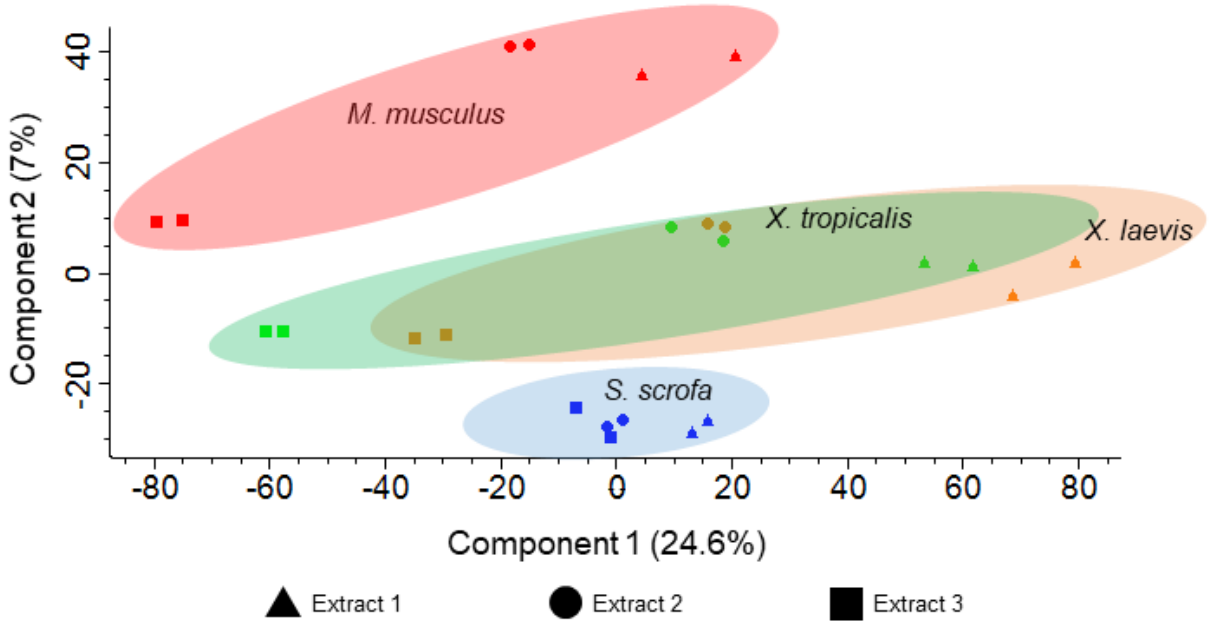


Figure SA3.3. PCA analysis of proteins identified in each extract from each species.

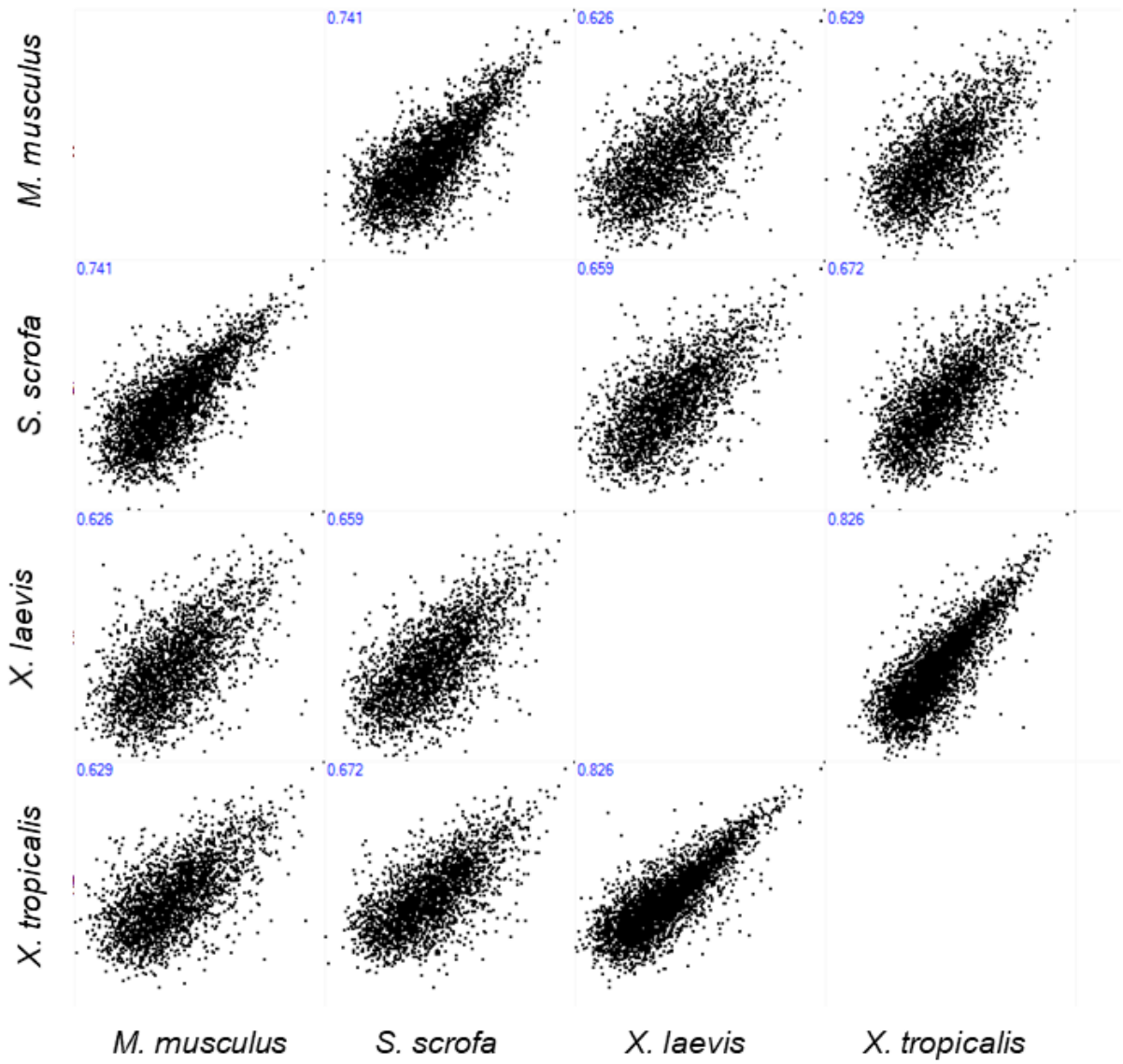


Figure SA3.4. Multiscatter plot of median-normalized precursor values across the four species. The two Xenopus species exhibit the greatest degree of similarity.

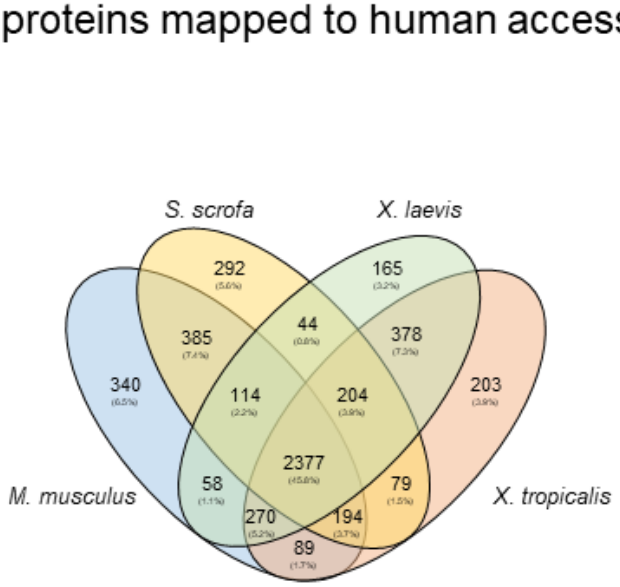


Figure SA3.5. All identified proteins that could be mapped to human entries. The data in figure 2 includes quantified proteins only, whereas this figure includes any identified but not quantified proteins.

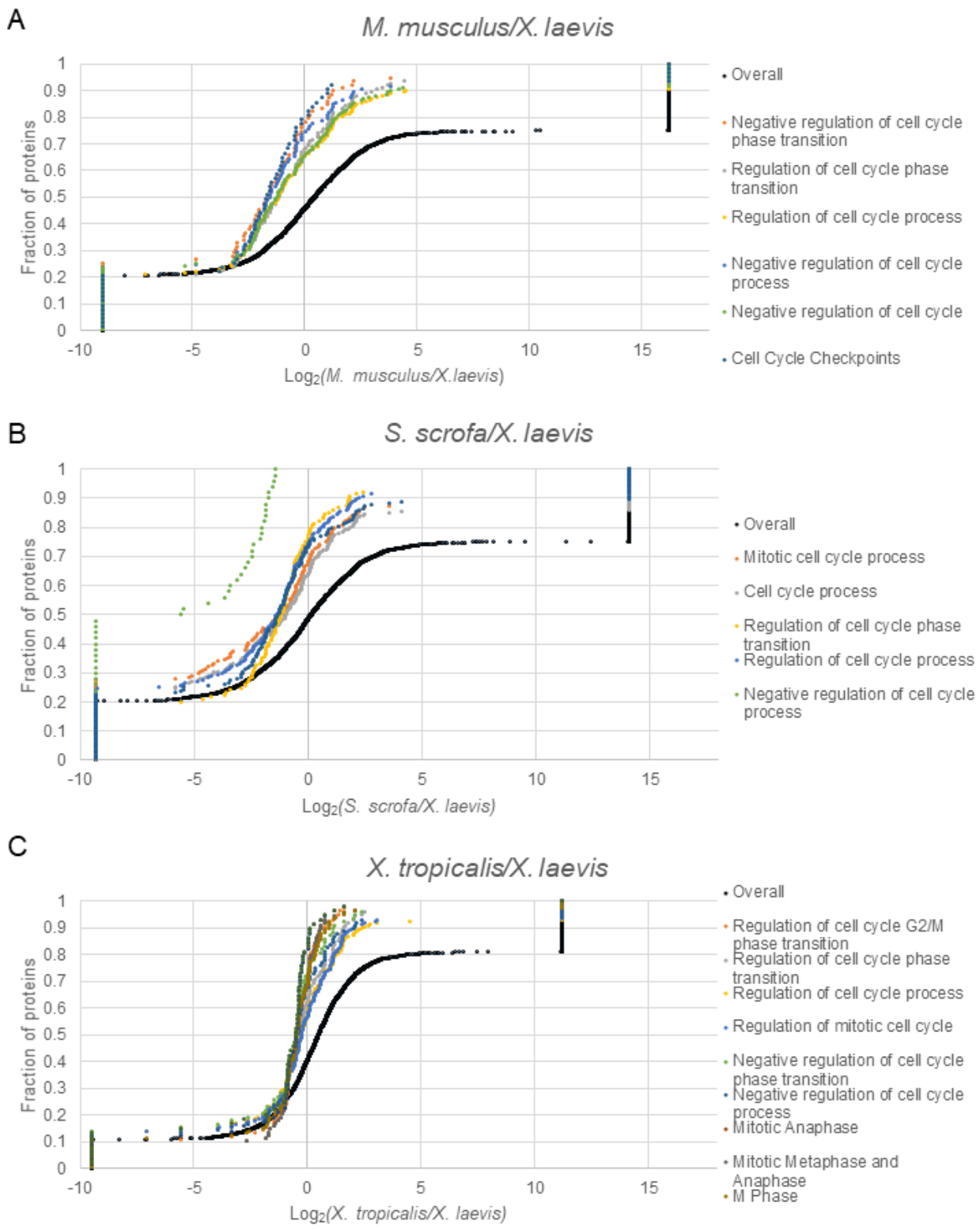
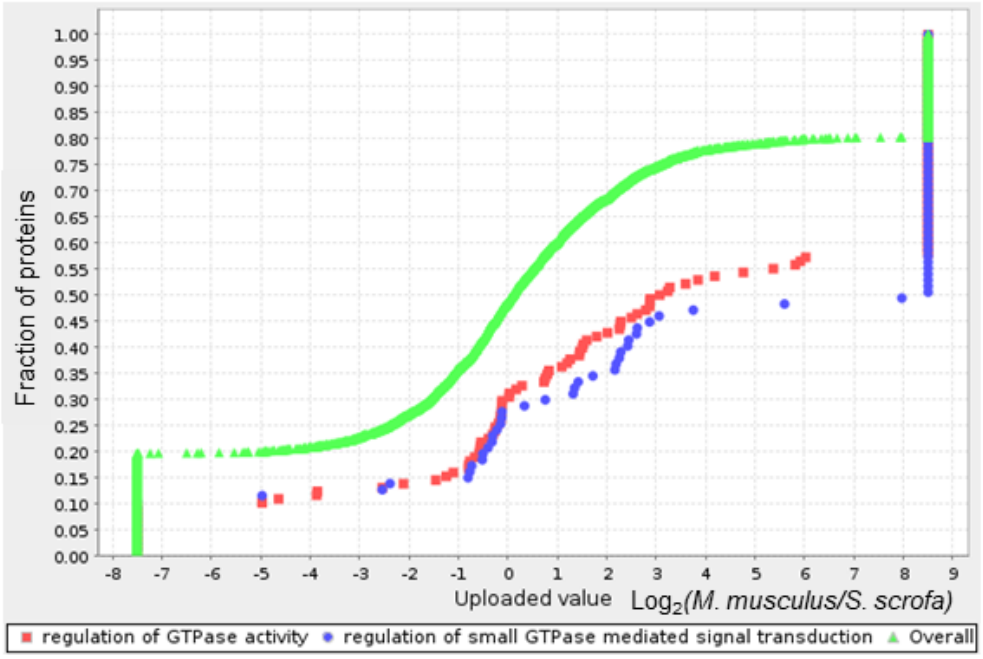


Figure SA3.6. Cell cycle GO enrichments in every pairwise comparison. X. laevis. (A) M. musculus / X. laevis (B) S. scrofa / X. laevis (C) X. tropicalis / X. laevis.

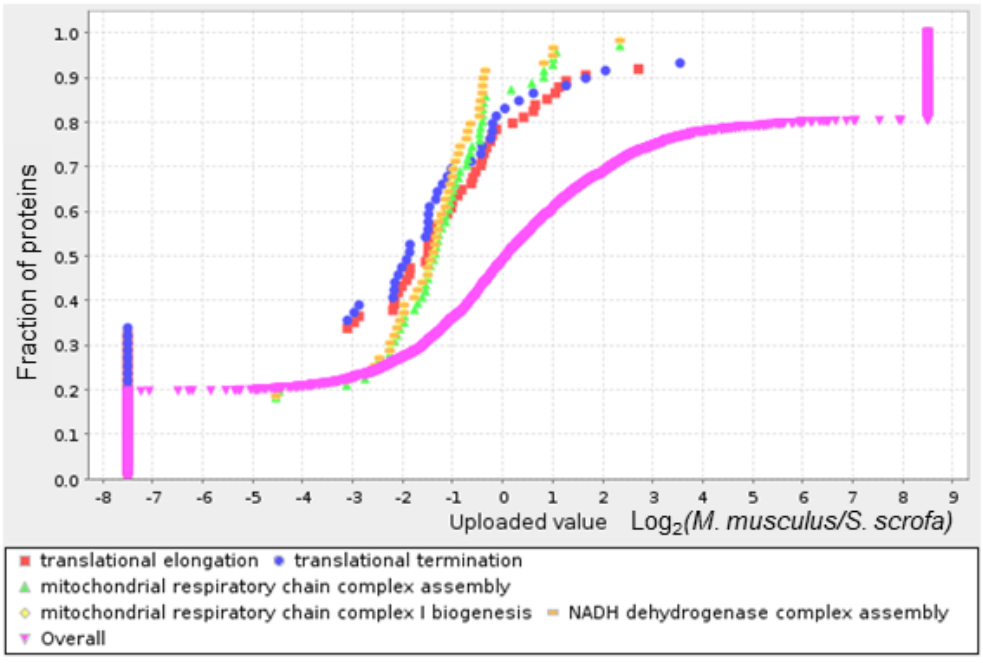
A

M. musculus compared to *S. scrofa*

Higher in
M. musculus



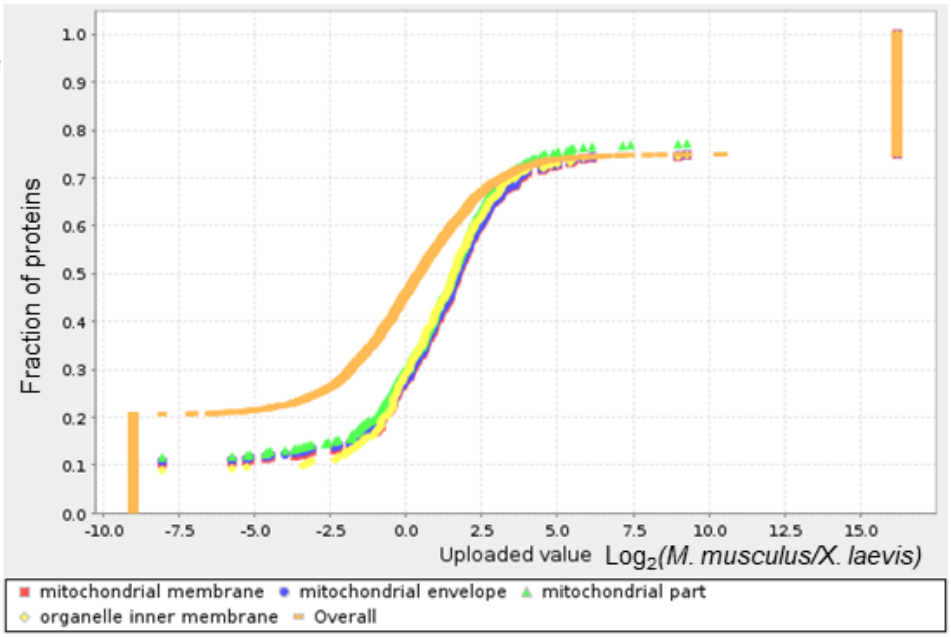
Higher in
S. scrofa



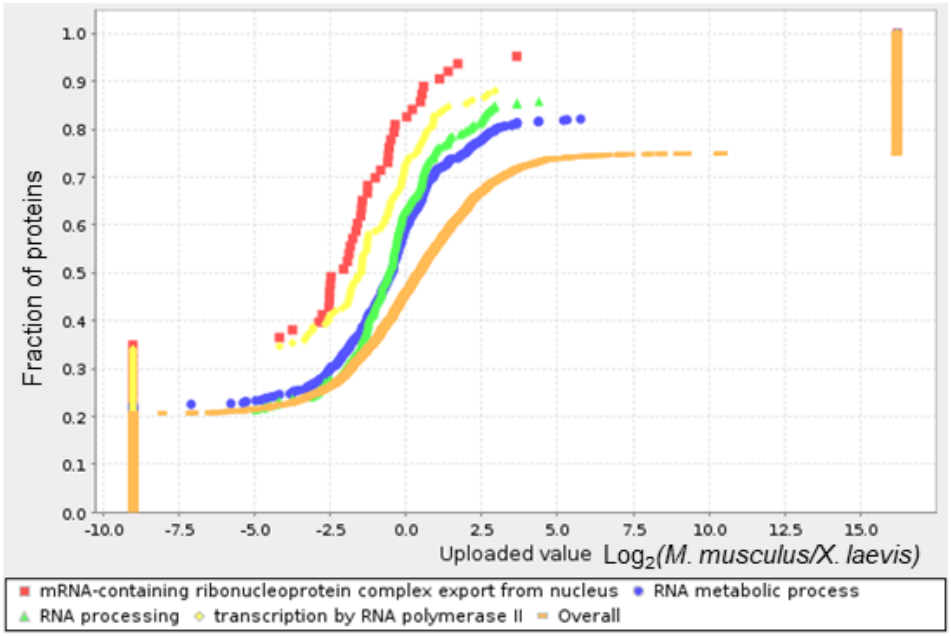
B

M. musculus compared to *X. laevis*

Higher in *M. musculus*



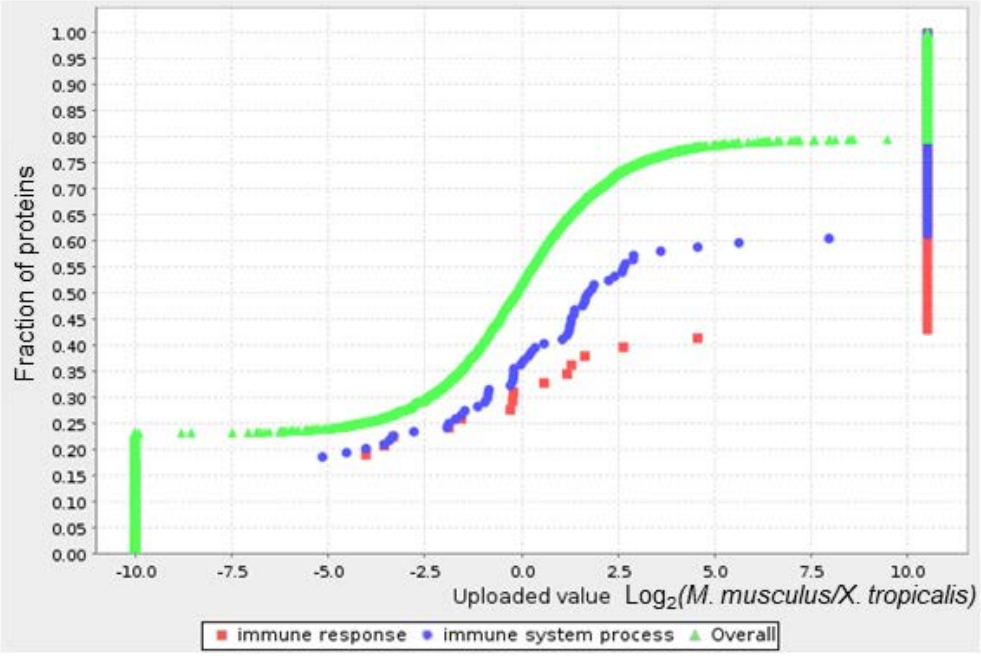
Higher in *X. laevis*

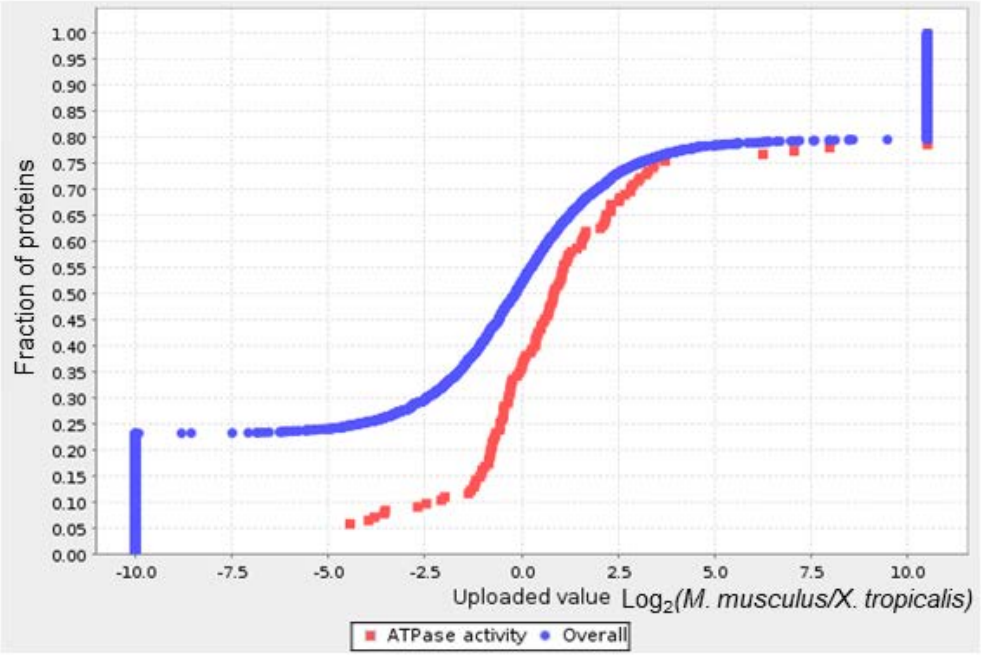


C

M. musculus compared to *X. tropicalis*

Higher in
M. musculus

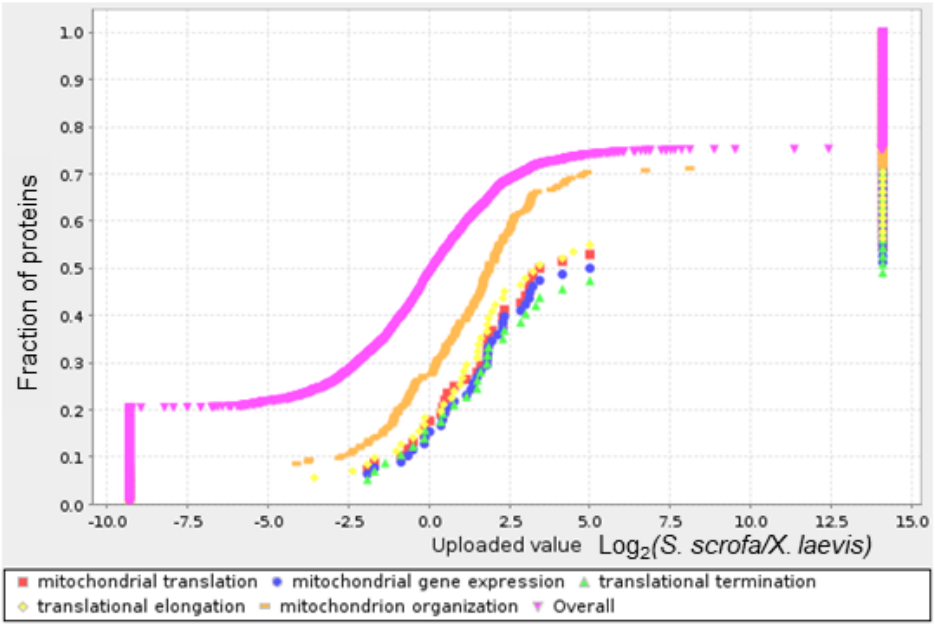




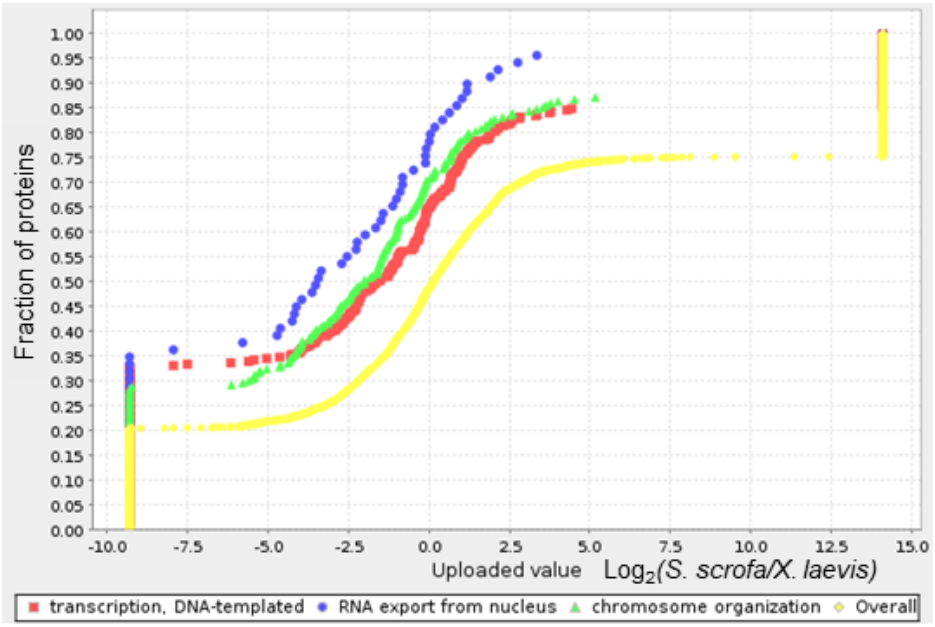
D

S. scrofa compared to *X. laevis*

Higher in
S. scrofa



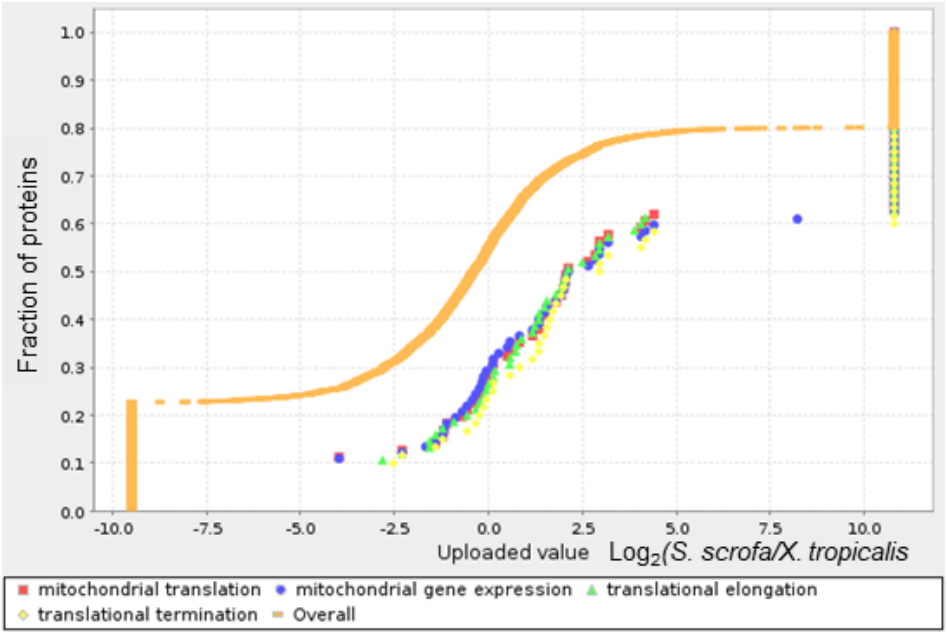
Higher in
X. laevis



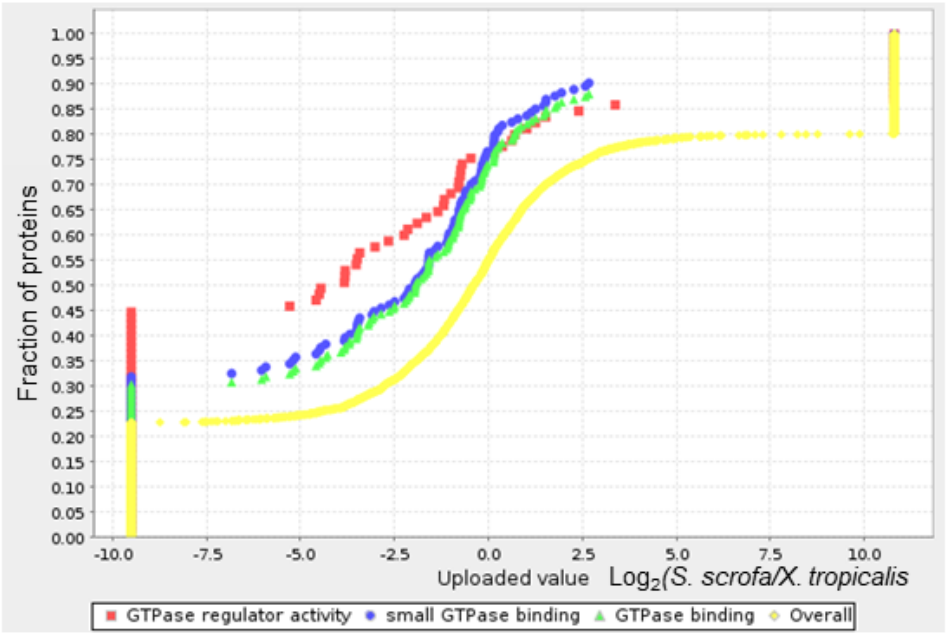
E

S. scrofa compared to *X. tropicalis*

Higher in *S. scrofa*



Higher in *X. tropicalis*



F

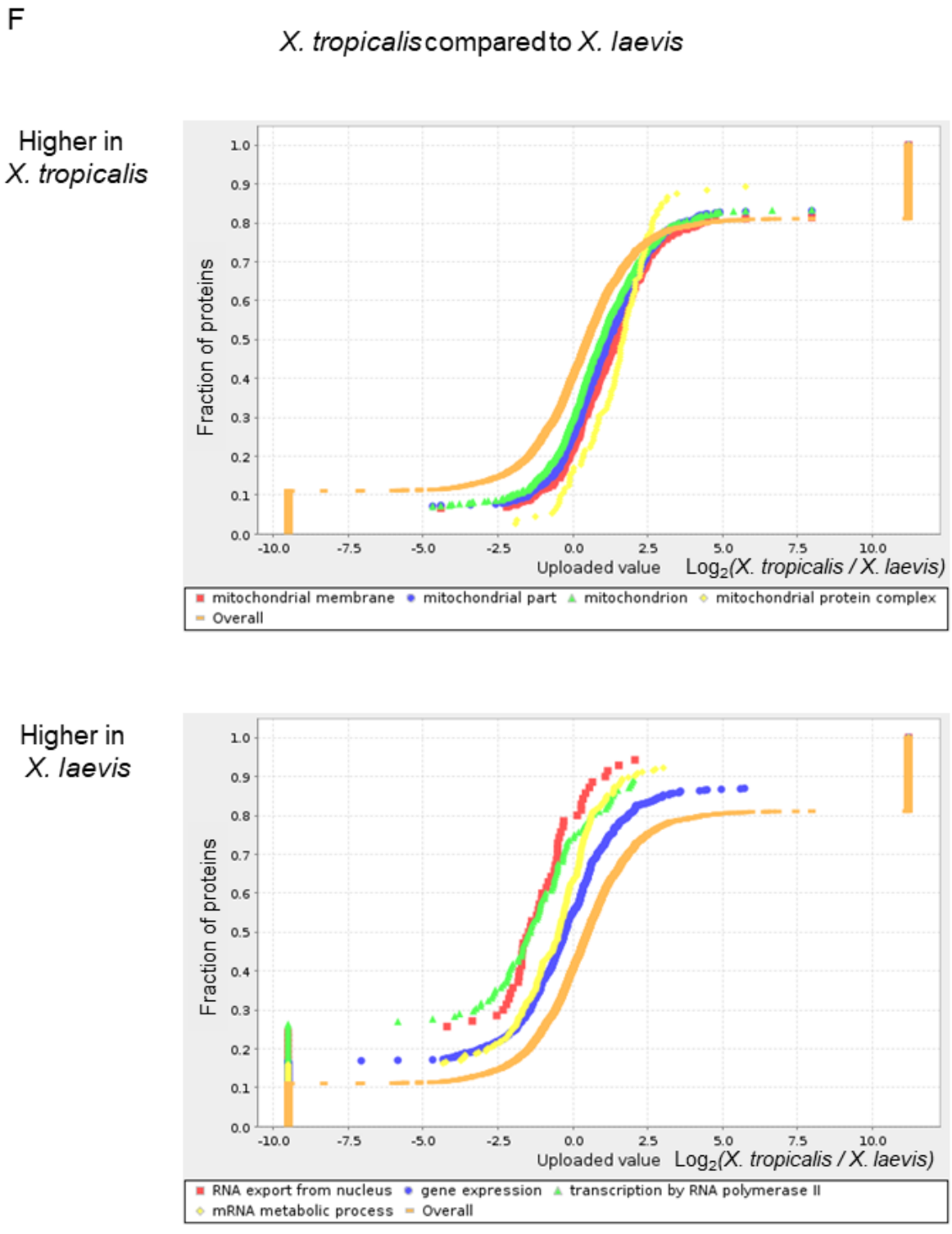
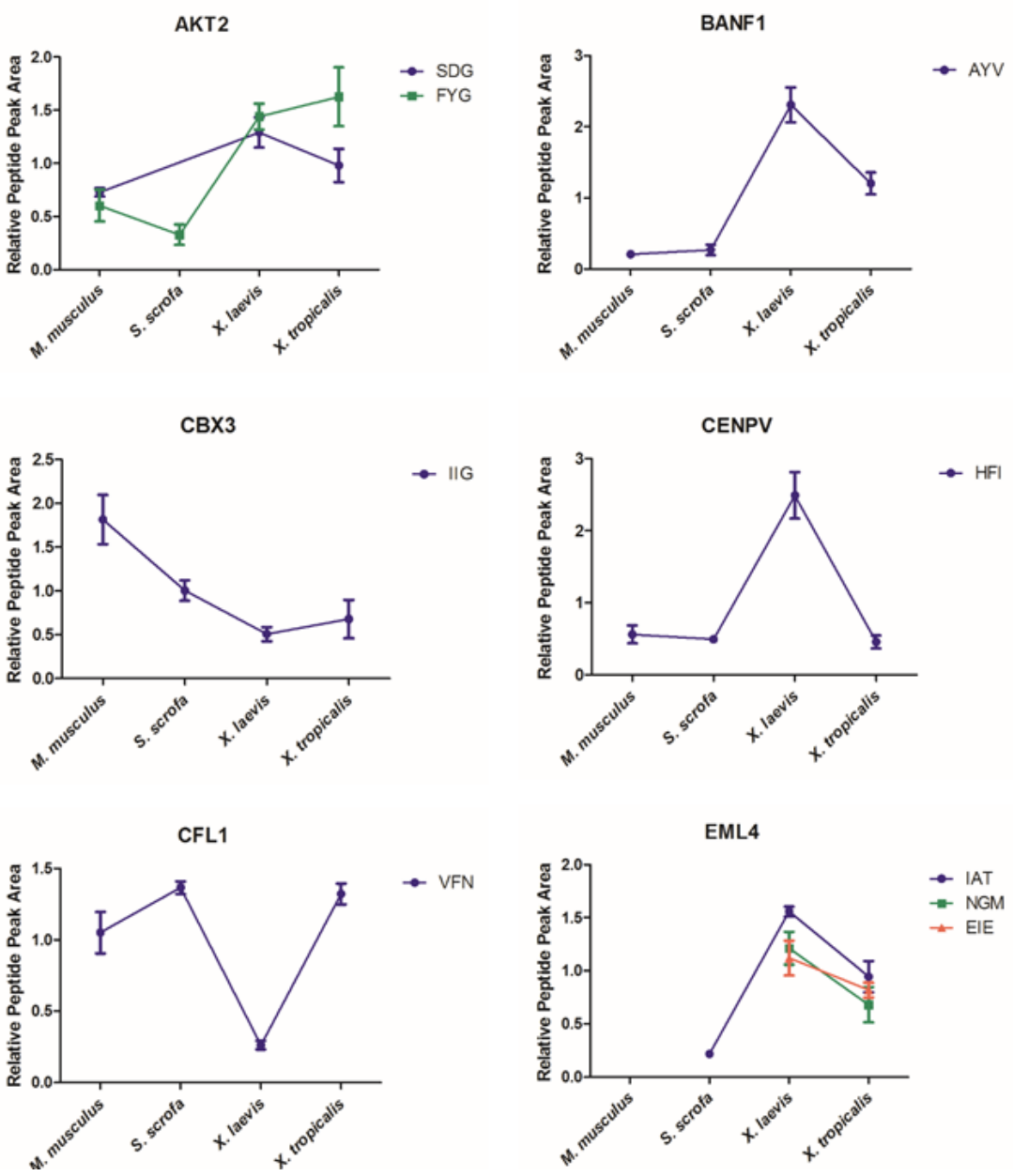
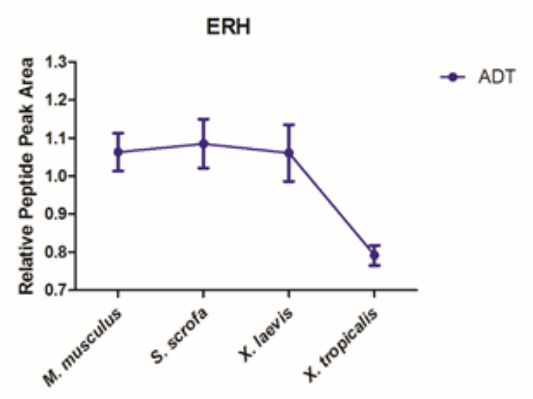
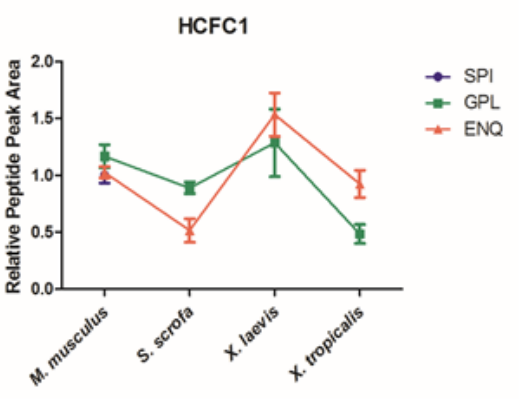


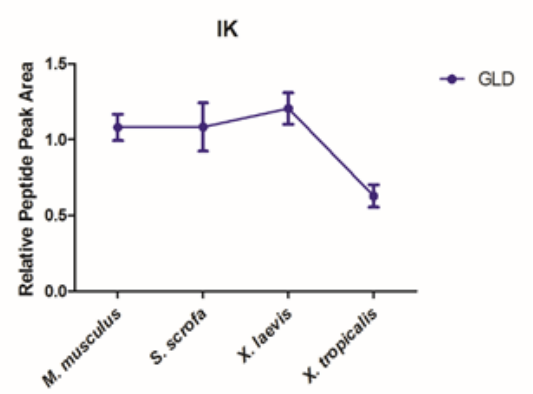
Figure SA3.7. Pairwise GSEA analysis of heart proteomes. *M. musculus* compared to *S. scrofa* (A), *M. musculus* compared to *X. laevis* (B), *M. musculus* compared to *X.*

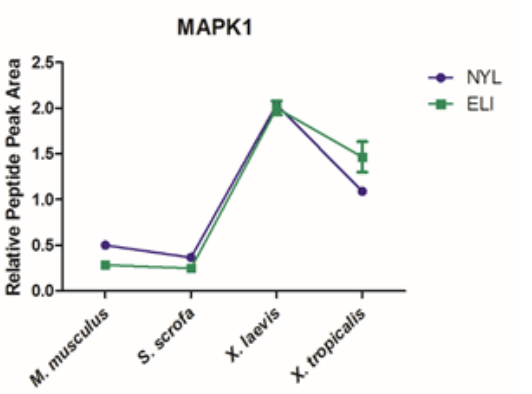
tropicalis (C), *S. scrofa* compared to *X. laevis* (D), *S. scrofa* compared to *X. tropicalis* (E), *X. tropicalis* compared to *X. laevis* (F).

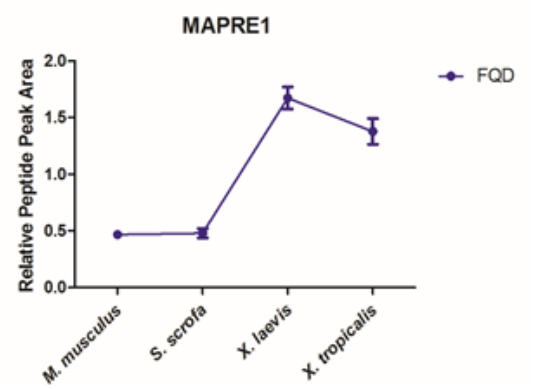

AKT2
Relative Peptide Peak Area
2.0
1.5
1.0
0.5
0.0
SDG
FYG
M. musculus
S. scrofa
X. laevis
X. tropicalis
BANF1
Relative Peptide Peak Area
3
2
1
0
AYV
M. musculus
S. scrofa
X. laevis
X. tropicalis
CBX3
Relative Peptide Peak Area
2.5
2.0
1.5
1.0
0.5
0.0
IIG
M. musculus
S. scrofa
X. laevis
X. tropicalis
CENPV
Relative Peptide Peak Area
3
2
1
0
HFI
M. musculus
S. scrofa
X. laevis
X. tropicalis
CFL1
Relative Peptide Peak Area
1.5
1.0
0.5
0.0
VFN
M. musculus
S. scrofa
X. laevis
X. tropicalis
EML4
Relative Peptide Peak Area
2.0
1.5
1.0
0.5
0.0
IAT
NGM
EIE
M. musculus
S. scrofa
X. laevis
X. tropicalis

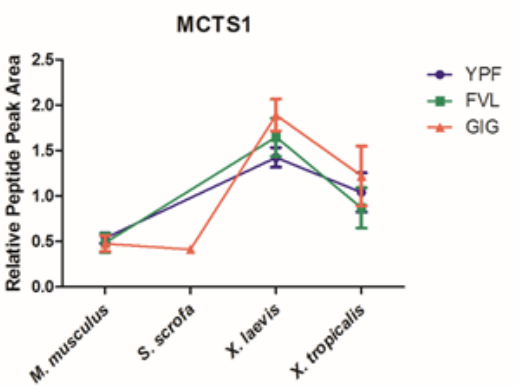

ERH
Relative Peptide Peak Area
1.3
1.2
1.1
1.0
0.9
0.8
0.7
ADT
M. musculus
S. scrofa
X. laevis
X. tropicalis

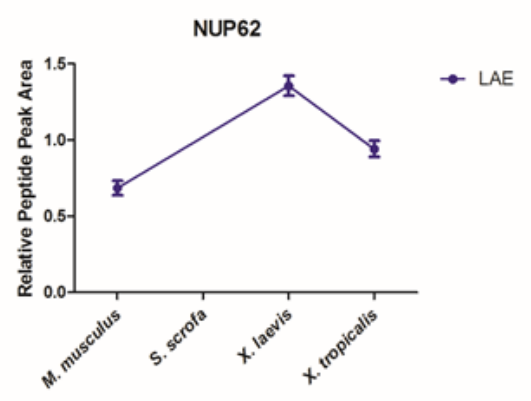

HCFC1
Relative Peptide Peak Area
2.0
1.5
1.0
0.5
0.0
SPI
GPL
ENQ
M. musculus
S. scrofa
X. laevis
X. tropicalis

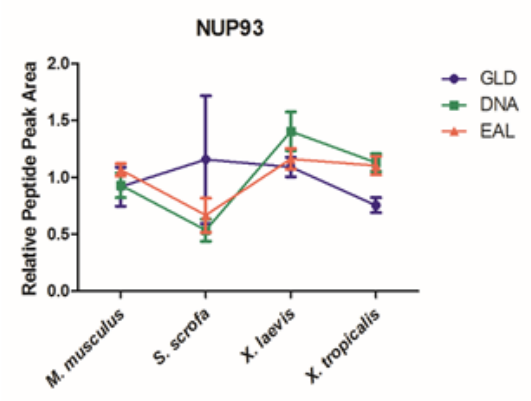

IK
Relative Peptide Peak Area
1.5
1.0
0.5
0.0
GLD
M. musculus
S. scrofa
X. laevis
X. tropicalis

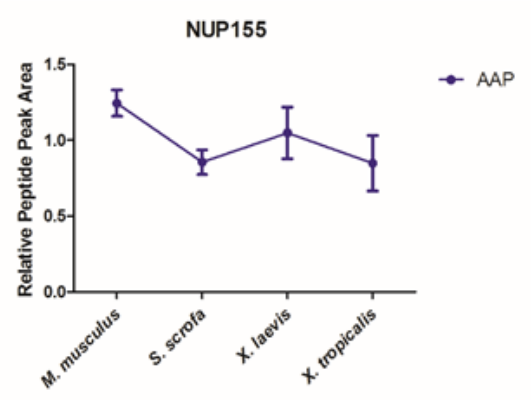

MAPK1
Relative Peptide Peak Area
2.5
2.0
1.5
1.0
0.5
0.0
NYL
ELI
M. musculus
S. scrofa
X. laevis
X. tropicalis

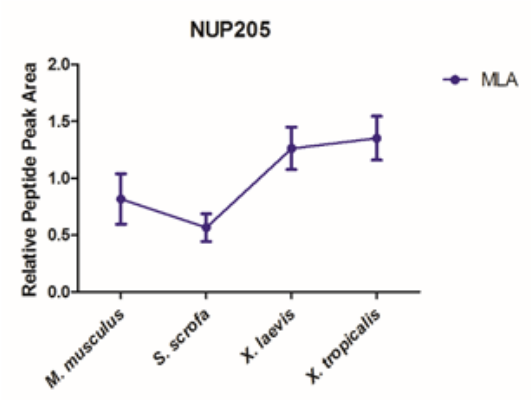

MAPRE1
Relative Peptide Peak Area
2.0
1.5
1.0
0.5
0.0
FQD
M. musculus
S. scrofa
X. laevis
X. tropicalis

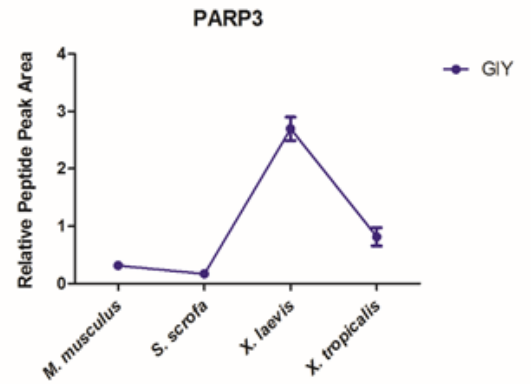

MCTS1
Relative Peptide Peak Area
2.5
2.0
1.5
1.0
0.5
0.0
YPF
FVL
GIG
M. musculus
S. scrofa
X. laevis
X. tropicalis

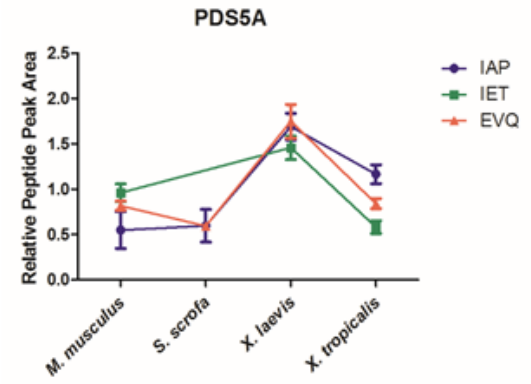

NUP62
Relative Peptide Peak Area
1.5
1.0
0.5
0.0
LAE
M. musculus
S. scrofa
X. laevis
X. tropicalis

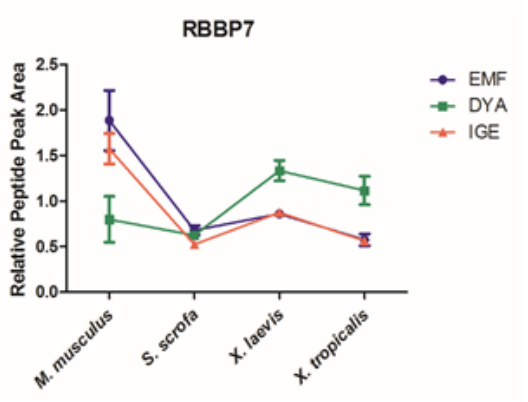

NUP93
Relative Peptide Peak Area
2.0
1.5
1.0
0.5
0.0
GLD
DNA
EAL
M. musculus
S. scrofa
X. laevis
X. tropicalis

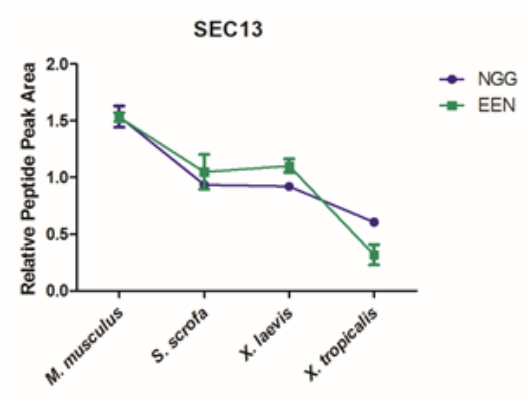

NUP155
Relative Peptide Peak Area
1.5
1.0
0.5
0.0
AAP
M. musculus
S. scrofa
X. laevis
X. tropicalis

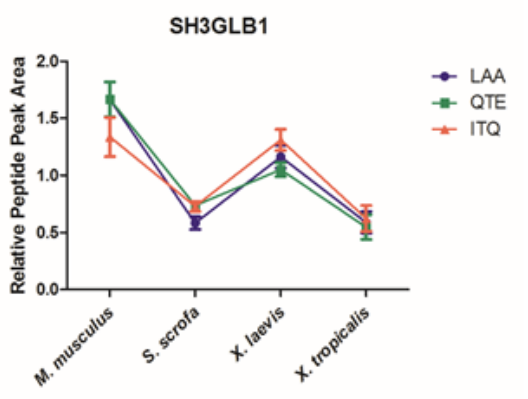

NUP205
Relative Peptide Peak Area
2.0
1.5
1.0
0.5
0.0
MLA
M. musculus
S. scrofa
X. laevis
X. tropicalis

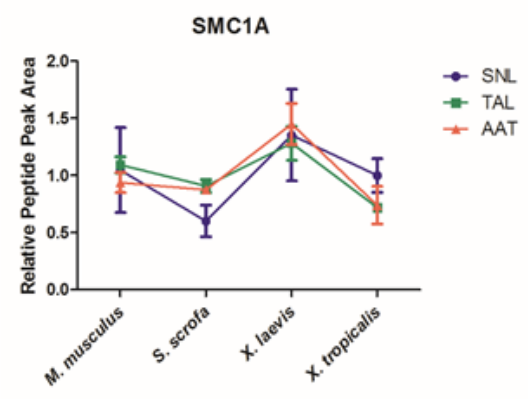

PARP3
Relative Peptide Peak Area
4
3
2
1
0
GIY
M. musculus
S. scrofa
X. laevis
X. tropicalis

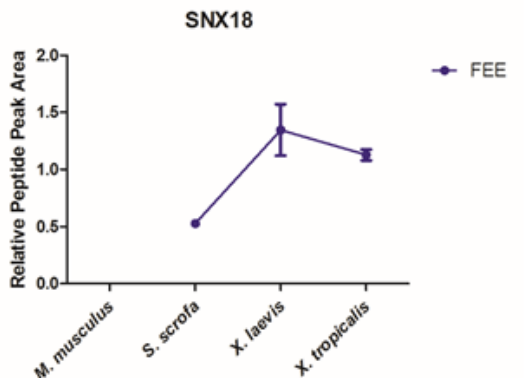

PDS5A
Relative Peptide Peak Area
2.5
2.0
1.5
1.0
0.5
0.0
IAP
IET
EVQ
M. musculus
S. scrofa
X. laevis
X. tropicalis

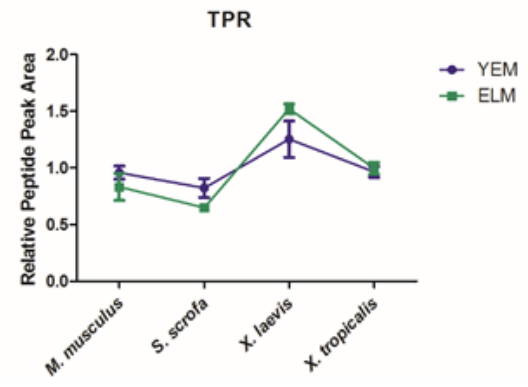

RBBP7
Relative Peptide Peak Area
2.5
2.0
1.5
1.0
0.5
0.0
EMF
DYA
IGE
M. musculus
S. scrofa
X. laevis
X. tropicalis


SEC13
Relative Peptide Peak Area
2.0
1.5
1.0
0.5
0.0
NGG
EEN
M. musculus
S. scrofa
X. laevis
X. tropicalis


SH3GLB1
Relative Peptide Peak Area
2.0
1.5
1.0
0.5
0.0
LAA
QTE
ITQ
M. musculus
S. scrofa
X. laevis
X. tropicalis


SMC1A
Relative Peptide Peak Area
2.0
1.5
1.0
0.5
0.0
SNL
TAL
AAT
M. musculus
S. scrofa
X. laevis
X. tropicalis


SNX18
Relative Peptide Peak Area
2.0
1.5
1.0
0.5
0.0
FEE
M. musculus
S. scrofa
X. laevis
X. tropicalis


TPR
Relative Peptide Peak Area
2.0
1.5
1.0
0.5
0.0
YEM
ELM
M. musculus
S. scrofa
X. laevis
X. tropicalis

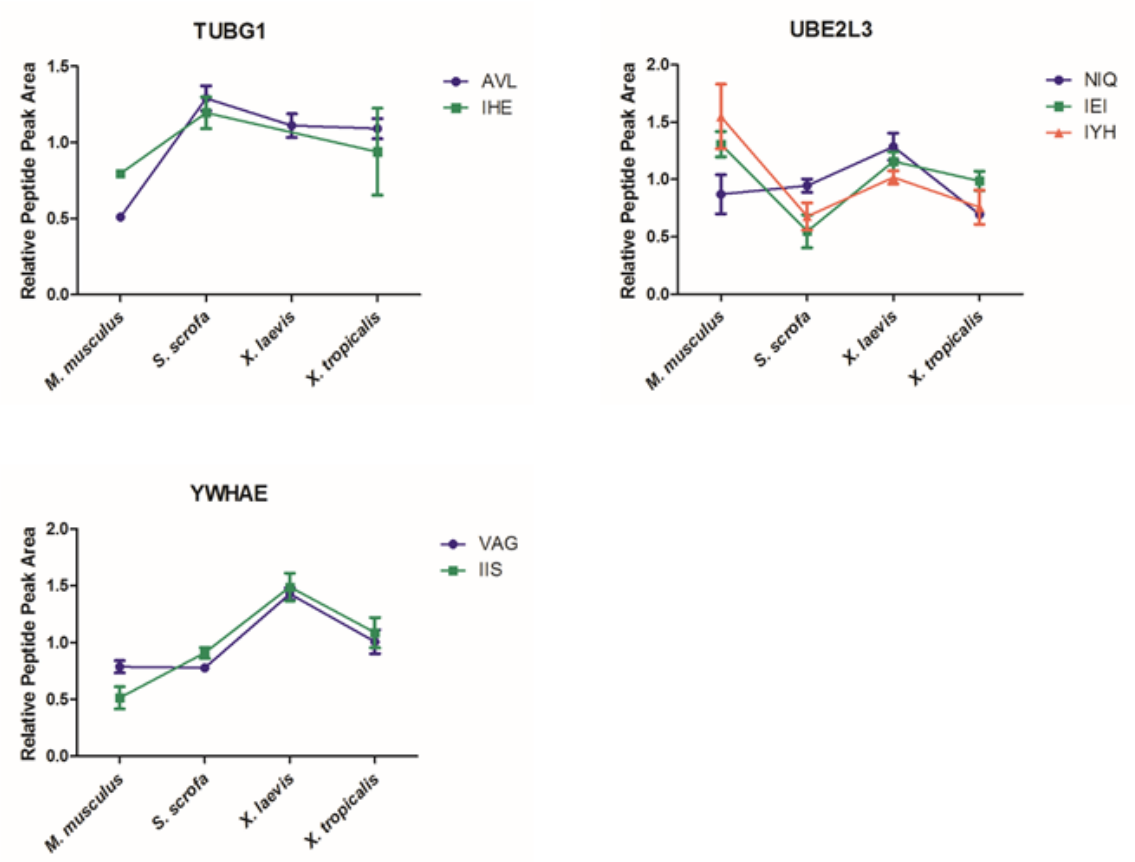


Figure SA3.8. Normalized peptide peak areas from PRM assays targeting cell cycle proteins enriched in X. laevis. Gene names are listed above the graphs and the first three amino acids for each peptide are listed to the right of the graph.

REFERENCES

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.H., Pages, F., Trajanoski, Z., and Galon, J. (2009). ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics 25, 1091-3.

Gotz, S., Garcia-Gomez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Robles, M., Talon, M., Dopazo, J., and Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res 36, 3420-35.

McCauley, M.D., and Wehrens, X.H. (2009). Animal models of arrhythmogenic cardiomyopathy. Dis Model Mech 2, 563-70.

Mi, H., Poudel, S., Muruganujan, A., Casagrande, J.T., and Thomas, P.D. (2016). PANTHER version 10: expanded protein families and functions, and analysis tools. Nucleic Acids Res 44, D336-42.

Sauls, K., Greco, T.M., Wang, L., Zou, M., Villasmil, M., Qian, L., Cristea, I.M., and Conlon, F.L. (2018). Initiating Events in Direct Cardiomyocyte Reprogramming. Cell Rep 22, 1913-1922.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13, 2498-504.

Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., and Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat Methods 13, 731-40.

Wu, G., Feng, X., and Stein, L. (2010). A human functional protein interaction network and its application to cancer data analysis. Genome Biol 11, R53.