THE UNIVERSITY OF CHICAGO

DEFINING GENE REGULATORY NETWORKS OF CARDIAC RHYTHM

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

INTERDISCIPLINARY SCIENTIST TRAINING PROGRAM:

DEVELOPMENT, REGENERATION, AND STEM CELL BIOLOGY

BY

RANGARAJAN D NADADUR

CHICAGO, ILLINOIS

AUGUST 2017

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

First and foremost, I need to thank my adviser, Dr. Ivan Moskowitz. From the day he interviewed me for graduate school, he has been a mentor, role model, and friend. I feel undeserving of his time, support and mentorship.

There are several additional mentors I have had over the last several years. Dr. Elizabeth McNally, who has stayed involved in every aspect of my research and training for the last 4 years. Dr. Mike Broman, who has been a mentor and friend since my first day in the lab. The countless hours we have spent together, in the lab and out, have been a highlight of the last 4 years for me. Dr. Chris Weber, who has been a close collaborator, friend, and mentor for several years. Dr. Rasu Shrestha, Dr. Sant Chawla, Dr. Mansoureh Eghbali, and Dr. Soban Umar, who have been my mentors since long before I knew what I wanted to do, and have been instrumental in leading me down this road.

I would like to thank all the members of the lab, both past and present, for the constant help, support, and friendship. Dr. Stefan Mazurek, who is present in every story worth telling I have over the last 4 years. Special thanks to Dr. Megan Rowton, Alex Guzzetta, Jeff Steimle, Dr. Andrew Hoffmann, Sonja Lazarevic, Mervenaz Koska, and Dr. Kohta Ikegami for being not only my co-workers but also my friends. Dr. Holly Yang, with whom the entirety of Chapter II below is a complete collaboration, and Carlos Perez-Cervantes from her group. The other senior members of the lab over the years who have taught me, including Dr. Ozanna Burnicka-Turek, Dr. Chul Kim, Dr. Junghun Kweon, and Dr. Shuhan Yu.

It has been my privilege to work with three extremely talented, hard-working undergraduate students during my PhD. I have spent too much of the last several years shamelessly accepting credit for their hard work. Jenna Bekeny, who was my constant

companion for almost the entirety of my time in the lab – I am extremely grateful for her assistance, patience, and friendship over the years. More recently, Margaret Gadek and Kaitlyn Shen have been a pleasure to work with. Margaret and Kaitlyn have taken such personal responsibility in the lab that I feel as if I take more direction from them than the other way around. I have doubtless learned more from them than they have learned from me.

I would like to thank all members of the McNally lab. Specifically, my fellow graduate students, Ellis Kim, Andy Vo, and Brandon Gardener. Tharrie Daniels for her willingness to discuss the NBA at all times. Dr. Megan Roy-Puckelwartz, Dr. Eugene Wyatt, Cynthia Harmon, Dr. John Fahrenbach, Dr. Dave Barefield, Dr. Mattia Quattrocelli, Tamari Miller, James Warner, Judy Earley, and Dr. Sasha Bogdanovich.

The other members of my committee and their labs for the expertise and assistance they have provided over the years. Dr. Marcelo Nobrega, Dr. Akira Imamoto, and Dr. Ken Onel.

Also at the University of Chicago, I want to thank Dr. Gene Kim and his lab, especially Tyler Calway and Cynthia Harmon. Dr. James Liao and Pat Ongusaha. Dr. Alex Ruthenburg and from his group Dr. Mike Werner.

I want to thank the University of Chicago MSTP program, especially Dr. Marcus Clark and the members of the class of ~2020.

Much of the work in this thesis was performed in close collaboration with multiple labs throughout the world, and I would like to thank them for their assistance, contributions, and advice over the years. Dr. Vincent Christoffels and his lab in Amsterdam. Dr. James Martin and his lab. Dr. Bas Boukens. Dr. Patrick Ellinor, and Dr. Nate Tucker from his group. Dr. Wouter de Laat, and Carien Hilvering and Valerio Bianchi from his group.

Last and most importantly, all of this work belongs to my family, who can take credit for every word written in this thesis, and everything I do for the rest of my life. I want to thank my parents, Nadadur S. Kumar and Dr. Vatsala Kumar, my siblings Ramanujan and Sandhya Nadadur, and my grandmother Leelamani Ramaswami.

CHAPTER I: INTRODUCTION

Function of the mammalian heart relies on coordinated electrical activity. Normal electrical conduction in the heart is governed by distinct populations of cardiomyocytes. To initiate the cardiac cycle, cardiomyocytes in the sinus node of the heart autonomously depolarize. This spontaneous electrical activity from the cardiac pacemaker propagates through the atrial chambers, where tightly coupled atrial cardiomyocytes rapidly depolarize, resulting in coordinated atrial contraction, pumping blood from the atria through the mitral and triscuspid valves into the ventricular chambers. The electrical current is slowed at the Atrioventricular (AV) node, allowing blood from the atria to fill the ventricles, before being rapidly propagated through the ventricular conduction system, the Bundle of His and Purkinje fibres, and into ventricular myocardium. Contraction of the ventricular chambers pumps blood through the aorta and pulmonary arteries, concluding the cardiac cycle. The cardiac cycle is remarkably robust, generating more than 2 billion contractions in the average human lifespan. However, the molecular pathways that govern this remarkable fidelity are yet to be fully elucidated.

Disruptions in this normal electrical conduction through the heart result in cardiac arrhythmias, a leading cause of morbidity and mortality in the US. The most common cardiac arrhythmia is Atrial Fibrillation (AF). AF affects more than 33 million individuals throughout the world. One in four Americans above the age of 40 will be diagnosed with AF in his or her lifetime. With an aging population, the prevalence of AF is expected to double in the next 25 years[1-4].

AF can be broadly described as uncoordinated, random electrical activity through the atrial chambers of the heart. The lack of coordinated atrial contraction leads to a multitude of secondary effects. AF accounts for one-fourth of all strokes and is also associated with an

increased risk of dementia, heart failure, and death, even in patients receiving optimal evidence-based care[1,3,4]. A 2014 study by Gladstone et al. in the *New England Journal Medicine* found that up to 20% of cryptogenic ischemic strokes and transient ischemic attacks of unknown cause were due to undiagnosed AF[5]. Even when appropriately diagnosed and managed, the medical and interventional therapies for AF have limited long-term efficacy, and are associated with significant side effects and potential complications. Pharmacologic therapies with anti-arrhythmic drugs have a multitude of side effects, including increasing the risk of arrhythmia burden. More invasive therapies include electrical cardioversion and ablation, both of which carry significant risks and side effects. AF-associated health care costs are estimated at greater than 26 billion dollars annually in the US alone [3,4].

Despite the profound healthcare burden caused by AF, there is a limited understanding of the mechanisms that provoke the arrhythmia in patients. Numerous risk factors have been identified for AF, such as age or heart failure, but these factors have little capacity to predict the onset of AF. Hence, new pathophysiologic insights into AF susceptibility are essential to guide improved diagnostic, therapeutic, and preventive strategies to reduce the burden of stroke, death, and hospitalizations associated with AF.

The Pathophysiology of Atrial Fibrillation

The cellular and molecular mechanisms of AF have been thoroughly reviewed by Heijman et al 2014 [6]. Briefly, AF is characterized by random, chaotic waves of cardiomyocyte depolarization that propagate throughout the atria, suppress sinus node activity and prevent coordinated atrial contraction. AF manifests with irregularly-irregular ventricular complexes on surface ECG accompanied by an undulating baseline with an absence of distinct *p* waves. The

pathophysiology of AF is thought to consist of two arrhythmogenic components: ectopic cellular depolarization events, or “triggers”, and a fibrillogenic substrate that abnormally propagates these triggers throughout the atrial chambers. Current models of AF posit that the presence of these two components is sufficient to sustain the arrhythmia[6].

Cardiomyocytes are electrically excitable cells, which hold a resting membrane potential until a threshold potential is reached. Achieving a threshold potential leads to the firing of an action potential which begins with rapid depolarization of the cell, mediated by voltage-gated sodium channels, notably NaV1.5, encoded by the *Scn5a* gene. This depolarization leads to calcium release from the Sarcoplasmic Reticulum (SR) through the cardiac Ryanodine Receptor (encoded by *Ryr2*). Released calcium binds to the contractile apparatus of the cardiomyocyte, causing cellular contraction. The vast majority of this released calcium (~80%) is taken back into the SR by the SERCA2 pump, encoded by *Atp2a2*, with the remaining calcium extruded from the cell through the sodium calcium exchanger (NCX). Repolarization of cardiomyocytes is governed by a host of repolarizing potassium channels[6-10].

In a normally conducting heart, the only source of depolarization events in the atrium is the sinus node, which shows autonomous phase 4 depolarization mediated by the “funny current” I_f , predominantly governed by the *Hcn4* gene. Depolarization events from any ectopic source in the atrium, form the basis of the triggered activity during AF. These abnormally triggered events can take many forms. Early After Depolarizations (EADs) occur where an action potential is induced during repolarization of the cell, traditionally during abnormal prolongation of cellular repolarization. Delayed After Depolarizations (DADs) occur where an action potential is induced during the refractory window of a cell. Lastly, cells can develop a propensity to gradually

depolarize while holding resting potential, a phenomenon that normally occurs only in autonomous nodal cells in the heart[6-10].

After an action potential, cardiomyocytes will remains refractory for a short period during which an action potential cannot be triggered. This limits re-entrant electrical activity, as a refractory cell is unable to propagate electrical activity necessary to sustain an arrhythmia. A critical component in the progression and maintenance of AF is the ability for triggered depolarizations to inappropriately propagate through normally refractory myocardium, referred to as a fibrillogenic substrate. Canonically, this fibrillogenic substrate is provided by enlargement of the atrial myocardium or fibrosis, which delays electrical propagation, allowing refractory cells to recover fire again. Shortening of the cellular action potential also allows cells to recover more rapidly and propagate arrhythmic triggers. Lastly, slowing of cell/cell conduction through gap junctions has a similar effect on slowing the propagation of atrial activity, resulting in fibrillogenic substrate[6,11].

Defining the Genetic Basis of Atrial Rhythm

The majority of Atrial Fibrillation occurs in the setting of pre-existing cardiac conditions, such as heart failure or hypertension. However, AF does manifest more rarely as a primary disease. “Lone” Atrial Fibrillation generally occurs in patients with early-onset arrhythmia in the absence of other cardiac pathologies and makes up approximately 10% of the diagnosed disease. Historically, the genetic causes of AF have been studied through DNA Sequencing and positional cloning approaches in individuals with early-onset lone AF and in families in which AF segregates as a Mendelian trait [12,13]. The first mutation identified in AF susceptibility was a missense gain of function variant of *KCNQ1*, encoding a cardiac potassium channel subunit [14].

Over the past two decades, DNA sequencing studies in families prone to AF have revealed rare mutations in ion channel subunits, gap junction proteins, and cardiac transcription factors[15,16],[12]. Still, the majority of lone AF cases are unresolved genetically.

Although AF most commonly occurs secondary to other cardiac pathologies, only a subset of patients with these heart diseases develops AF. These findings indicate a genetic predisposition to AF that is uncovered by additional cardiac insult; indeed, AF has high heritability (60%)[17]. Despite the historical studies in heritable AF, a genetic predisposition to common AF has only been more recently uncovered. The advent of Genome Wide Association Studies (GWAS) has allowed the systematic and unbiased identification of disease risk loci. The first GWAS was performed in 2002 and discovered a novel locus on chromosome 6p21 associated with increased risk of myocardial infarction[18]. Over the past 15 years, the number of GWAS studies has grown exponentially, numbering more than 3000 across countless traits and diseases[19]. The first GWAS for Atrial Fibrillation was performed in 2006, and discovered an association at a SNP on chromosome 4q25, upstream of the gene *PITX2*[20]. Since this landmark study, there have been over 30 genetic loci associated with AF[17,21-23] [12,13,24]. Unsurprisingly, the majority of AF risk loci are associated with ion channels and structural heart genes[17]. However, a number of transcriptional regulators have been implicated in AF susceptibility, notably *TBX5* and *PITX2* [12,15-17,20,22-30].

The genetic susceptibility causing increased AF risk in GWAS lies in non-coding regions, and suggests that this variation acts mechanistically by altering *cis*-regulatory function (enhancers) and effector gene expression. These observations posit a transcriptional model for increased AF risk unveiled by AF GWAS: altered transcription factor function at downstream enhancers results in altered effector gene expression and increased AF risk. Defining the specific

transcriptional networks involved will allow prediction of AF risk and a strategy for primary prevention.

This thesis will first define a two tiered transcriptional network underlying AF susceptibility. This transcriptional hierarchy, structured as an incoherent feed-forward loop, confers robustness to atrial rhythm. The second chapter of this thesis will address the general challenges of defining cis-regulatory elements in gene regulatory networks. We propose a novel genomic approach, using transcription of regulatory elements as noncoding RNAs, to define transcriptional networks, and test this approach in the context of defining the networks underlying susceptibility to Atrial Fibrillation.

References

1. Andrade, J., Khairy, P., Dobrev, D. & Nattel, S. The clinical profile and pathophysiology of atrial fibrillation: relationships among clinical features, epidemiology, and mechanisms. *Circulation research* **114**, 1453-1468 (2014).
2. Woods, C.E. & Olgin, J. Atrial fibrillation therapy now and in the future: drugs, biologicals, and ablation. *Circulation research* **114**, 1532-1546 (2014).
3. Camm, A.J., *et al.* 2012 focused update of the ESC Guidelines for the management of atrial fibrillation: an update of the 2010 ESC Guidelines for the management of atrial fibrillation. Developed with the special contribution of the European Heart Rhythm Association. *Eur Heart J* **33**, 2719-2747 (2012).
4. Nishida, K. & Nattel, S. Atrial fibrillation compendium: historical context and detailed translational perspective on an important clinical problem. *Circulation research* **114**, 1447-1452 (2014).
5. Gladstone, D.J., *et al.* Atrial fibrillation in patients with cryptogenic stroke. *The New England journal of medicine* **370**, 2467-2477 (2014).
6. Heijman, J., Voigt, N., Nattel, S. & Dobrev, D. Cellular and molecular electrophysiology of atrial fibrillation initiation, maintenance, and progression. *Circulation research* **114**, 1483-1499 (2014).
7. Ginsburg, K.S., Weber, C.R. & Bers, D.M. Control of maximum sarcoplasmic reticulum Ca load in intact ferret ventricular myocytes. Effects Of thapsigargin and isoproterenol. *J Gen Physiol* **111**, 491-504 (1998).
8. Ginsburg, K.S., Weber, C.R. & Bers, D.M. Cardiac Na+-Ca2+ exchanger: dynamics of Ca2+-dependent activation and deactivation in intact myocytes. *J Physiol* **591**, 2067-2086 (2013).
9. Glukhov, A.V., Flagg, T.P., Fedorov, V.V., Efimov, I.R. & Nichols, C.G. Differential K(ATP) channel pharmacology in intact mouse heart. *J Mol Cell Cardiol* **48**, 152-160 (2010).
10. Limberg, S.H., *et al.* TASK-1 channels may modulate action potential duration of human atrial cardiomyocytes. *Cell Physiol Biochem* **28**, 613-624 (2011).
11. Christophersen, I.E., *et al.* Rare variants in GJA5 are associated with early-onset lone atrial fibrillation. *Can J Cardiol* **29**, 111-116 (2013).
12. Fatkin, D., Santiago, C.F., Huttner, I.G., Lubitz, S.A. & Ellinor, P.T. Genetics of Atrial Fibrillation: State of the Art in 2017. *Heart, lung & circulation* (2017).
13. Christophersen, I.E. & Ellinor, P.T. Genetics of atrial fibrillation: from families to genomes. *J Hum Genet* (2015).
14. Chen, Y.H., *et al.* KCNQ1 gain-of-function mutation in familial atrial fibrillation. *Science* **299**, 251-254 (2003).
15. Mahida, S., Lubitz, S.A., Rienstra, M., Milan, D.J. & Ellinor, P.T. Monogenic atrial fibrillation as pathophysiological paradigms. *Cardiovascular research* **89**, 692-700 (2011).
16. Olesen, M.S., Nielsen, M.W., Haunso, S. & Svendsen, J.H. Atrial fibrillation: the role of common and rare genetic variants. *Eur J Hum Genet* **22**, 297-306 (2014).
17. Tucker, N.R. & Ellinor, P.T. Emerging directions in the genetics of atrial fibrillation. *Circulation research* **114**, 1469-1482 (2014).

18. Ozaki, K., *et al.* Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. *Nature genetics* **32**, 650-654 (2002).
19. Manolio, T.A. In Retrospect: A decade of shared genomic associations. *Nature* **546**, 360-361 (2017).
20. Gudbjartsson, D.F., *et al.* Variants conferring risk of atrial fibrillation on chromosome 4q25. *Nature* **448**, 353-357 (2007).
21. Ellinor, P.T., *et al.* Meta-analysis identifies six new susceptibility loci for atrial fibrillation. *Nature genetics* **44**, 670-675 (2012).
22. Lubitz, S.A., *et al.* Novel genetic markers associate with atrial fibrillation risk in Europeans and Japanese. *Journal of the American College of Cardiology* **63**, 1200-1210 (2014).
23. Sinner, M.F., *et al.* Integrating genetic, transcriptional, and functional analyses to identify 5 novel genes for atrial fibrillation. *Circulation* **130**, 1225-1235 (2014).
24. Christophersen, I.E., *et al.* Large-scale analyses of common and rare variants identify 12 new loci associated with atrial fibrillation. *Nature genetics* **49**, 946-952 (2017).
25. Kirchhof, P., *et al.* PITX2c is expressed in the adult left atrium, and reducing Pitx2c expression promotes atrial fibrillation inducibility and complex changes in gene expression. *Circ Cardiovasc Genet* **4**, 123-133 (2011).
26. McDermott, D.A., Hatcher, C.J. & Basson, C.T. Atrial Fibrillation and Other Clinical Manifestations of Altered TBX5 Dosage in Typical Holt-Oram Syndrome. *Circulation research* **103**, e96 (2008).
27. Olesen, M.S., *et al.* Genetic loci on chromosomes 4q25, 7p31, and 12p12 are associated with onset of lone atrial fibrillation before the age of 40 years. *Can J Cardiol* **28**, 191-195 (2012).
28. Patel, C., Silcock, L., McMullan, D., Brueton, L. & Cox, H. TBX5 intragenic duplication: a family with an atypical Holt-Oram syndrome phenotype. *Eur J Hum Genet* **20**, 863-869 (2012).
29. Postma, A.V., *et al.* A gain-of-function TBX5 mutation is associated with atypical Holt-Oram syndrome and paroxysmal atrial fibrillation. *Circulation research* **102**, 1433-1442 (2008).
30. Zhou, Y.M., Zheng, P.X., Yang, Y.Q., Ge, Z.M. & Kang, W.Q. A novel PITX2c lossoffunction mutation underlies lone atrial fibrillation. *Int J Mol Med* **32**, 827-834 (2013).

CHAPTER II: *PITX2* MODULATES A *TBX5*-DEPENDENT GENE REGULATORY NETWORK TO MAINTAIN ATRIAL RHYTHM

Abstract

Cardiac rhythm is extremely robust, generating two billion contraction cycles during the average human lifespan. Transcriptional control of cardiac rhythm is poorly understood. We found that removal of the transcription factor gene *Tbx5* from the adult mouse caused primary spontaneous and sustained atrial fibrillation (AF). Atrial cardiomyocytes from the *Tbx5*-mutant mice exhibited action-potential abnormalities, including spontaneous depolarizations, that were rescued by chelating free calcium. We identified a multi-tiered transcriptional network that linked seven previously defined AF risk loci: TBX5 directly activated PITX2, and TBX5 and PITX2 antagonistically regulated membrane effector genes *Scn5a, Gja1, Ryr2, Dsp,* and *Atp2a2*. In addition, reduced *Tbx5* dose by adult-specific haploinsufficiency caused decreased target gene expression, myocardial automaticity, and AF inducibility, which were all rescued by *Pitx2* haploinsufficiency in mice. These results defined a transcriptional architecture for atrial rhythm control organized as an incoherent feed-forward loop, driven by TBX5 and modulated by PITX2. TBX5/PITX2 interplay provides tight control of atrial rhythm effector gene expression, and perturbation of the co-regulated network caused AF susceptibility. This work provides a model for the molecular mechanisms underpinning the genetic implication of multiple AF GWAS loci and will contribute to future efforts to stratify patients for AF risk based on genotype.

Introduction

The transcriptional architecture that confers robustness to cardiac rhythm must tightly control cardiac channel genes expression, as increased or decreased channel expression can cause cardiac arrhythmias (*1*). Atrial fibrillation (AF), the most common human arrhythmia, is an irregularly irregular pattern of atrial depolarization resulting in uncoordinated atrial contraction. AF affects more than 33 million people worldwide and represents a growing cause of morbidity and mortality (*2*). Although most AF presents in the context of concomitant cardiac pathology, such as hypertension and heart failure, genome wide association studies (GWAS) have identified a genetic predisposition underlying AF.

Human AF GWAS have implicated multiple transcription factors, including *TBX5* and *PITX2* in AF, raising the possibility that perturbations of a gene regulatory network for atrial rhythm control may underlie some AF susceptibility (*3*). Dominant mutations in the T-box transcription factor *TBX5* cause Holt-Oram Syndrome, characterized by disrupted heart and limb development (*4-6*), and increased AF risk (*6*). GWAS has linked common sequence variants close to or intronic within TBX5 to prolongation of the PR-interval—the time interval between the electrocardiogram P-wave and QRS, an AF risk factor—and more recently to increased AF risk itself (*7-10*). *PITX2*, a paired-like homeodomain transcription factor, plays a critical role in heart development and adult rhythm control. *PITX2* is the most frequently reported AF-susceptibility locus (*3*). Adult-specific *Pitx2* deletion in mice causes AF susceptibility and increased expression of ion channels linked to AF (*11*).

AF pathophysiology requires a cellular trigger event, which initiates inappropriate depolarization, and a fibrillogenic substrate, or abnormal myocardium that propagates the trigger. Most genetic loci implicated in AF encode ion channels that affect trigger frequency or substrate,

but not both (*3*). Animal models based on disruption of these channels do not exhibit spontaneous AF in the absence of concomitant cardiac pathophysiology, and no mouse model of primary AF has been reported (*12*). We tested the hypothesis that adult-specific removal of *Tbx5* in the mouse may cause atrial gene regulatory network dysfunction and AF. This removal generated primary, spontaneous, sustained AF in the absence of other cardiac pathology. We identified a regulatory element at *PITX2* that was modulated by a canonical TBX5 binding site. *Tbx5*-haploinsufficiency–induced transcriptional and myocyte abnormalities and AF susceptibility were rescued by *Pitx2*-haploinsufficiency. We have therefore uncovered a multi-level gene regulatory network for atrial rhythm homeostasis, driven by TBX5 and repressed by PITX2, organized as a type 1 incoherent feed-forward loop (*13*). This transcriptional architecture links 7 AF GWAS loci in a single network, providing a molecular model for their genetic implication in AF risk.

Results

Adult-specific *Tbx5* deletion causes rapid onset atrial fibrillation

Tbx5 was deleted from the adult mouse by combining a *Tbx5* floxed allele ($Tbx5^{fl}$) (*5*) with a tamoxifen-inducible Cre recombinase allele at the Rosa26(R26) locus (*14*). Mice homozygous for both alleles ($Tbx5^{fl/fl};R26^{CreERt2}$) were treated with tamoxifen at 6 weeks-of-age, removing the T-box DNA binding region and ablating TBX5 expression (*15*) (*4, 5*). *Tbx5*-deleted mice developed an irregularly irregular heart rate 1-2 weeks after TM-treatment. Electrocardiograms (ECGs) of ambulatory TM-treated $Tbx5^{fl/fl};R26^{CreERt2}$ mice showed a disorganized pattern of atrial activity compared to TM-treated $R26^{CreERt`2}$ controls (Fig 2.1A). Absence of P-waves in $Tbx5^{fl/fl};R26^{CreERt2}$ mice but not controls was observed by signal-averaging ~1000 heartbeats (Fig 1B). Poincaré analysis of the heartbeat, comparing successive beat lengths (utilizing R-R intervals, the time interval between sequential ECG 'R' peaks) demonstrated stable intervals in controls but tremendous instability in $Tbx5^{fl/fl};R26^{CreERt2}$ mice, indicative of an irregularly irregular heartbeat and AF (Fig. 1C). Furthermore, a single atrial depolarization overlapped the surface P-wave in $R26^{CreERt2}$ controls but multifocal irregular depolarizations were observed in $Tbx5^{fl/fl};R26^{CreERt2}$ mice, by *in vivo* intracardiac electrograms 2 weeks after TM treatment (Fig 1D).

We further interrogated the atrial arrhythmia by examining trans-mitral valve blood flow by pulsed wave Doppler echocardiography. Trans-mitral valve blood flow demonstrated 2 phases: an "E-wave" followed by "A-wave" in control $R26^{CreERt2}$ mice, but the "A-wave" was absent in $Tbx5^{fl/fl};R26^{CreERt2}$ mice (Fig 2.1E). The absence of "A-wave" flow across the mitral valve is indicative of a lack of coordinated atrial contraction, a characteristic of AF. We observed

AF by these metrics in 24/24 *Tbx5*$^{fl/fl}$*;R26*^{CreERt2} mice and 0/10 *R26*^{CreERt2} mice ($p=7.6x10^{-9}$, Fisher's Exact Test). These findings demonstrate spontaneous sustained AF in adult-specific *Tbx5*-mutant mice.

Cardiac function of *Tbx5*$^{fl/fl}$*;R26*^{CreERt2} mice at the onset of AF was unchanged compared to controls., with no difference in left ventricular ejection fraction (LVEF) observed in *Tbx5*$^{fl/fl}$*;R26*^{CreERt2} compared to *R26*^{CreERt2} mice at day-14 post-TM (fig 2.S1). Ventricular conduction abnormalities, including monomorphic ventricular tachycardia (fig S2A), polymorphic ventricular tachycardia (fig S2B), and prolonged sinus pauses (fig S2C) were observed; however, these arrhythmias were not present at the time of AF onset in *Tbx5*$^{fl/fl}$*;R26*^{CreERt2} mice and only manifested well after AF was observed in these animals, beginning ~3 weeks post-tamoxifen. Removal of *Tbx5* from the adult mouse therefore causes primary AF in the absence of left ventricular dysfunction or ventricular conduction abnormalities.

To assess atrial action-potential propagation, we performed *ex vivo* optical mapping of intact adult mouse hearts (*16*). Adult-specific *Tbx5*-mutant mouse hearts demonstrated progressively slowed atrial conduction prior to AF (Fig 2.1F). Atrial activation time was 13 ms at day 7, 20 ms at day 12, and >40 ms at day 14 post-tamoxifen TM *in Tbx5*$^{fl/fl}$*; R26*^{CreERt2} mice (Fig 2.1F, movies S1-S3). Action potential duration (APD) increased from 44±4 to 53±3 ms (120-ms pacing interval) at day-14 post-tamoxifen (fig S3-S4). Macro-reentrant arrhythmias were observed in *Tbx5*$^{fl/fl}$*;R26*^{CreERt2} mice with paroxysmal and sustained fibrillating patterns by day 14 (Fig 2.1G). In contrast, *R26*^{CreERt2} control mice showed a uniform atrial activation pattern with no conduction slowing 14 days after the tamoxifen regimen (Fig 2.1F). No paroxysmal or sustained atrial fibrillation was observed. Thus, removal of *Tbx5* slowed trans-atrial conduction

4-fold between the first and second week after deletion, setting the stage for reentrant arrhythmias—an organ-level mechanism for AF.

Disrupted calcium flux causes abnormal cellular depolarizations in *Tbx5*-mutant atrial myocytes

Removal of *Tbx5* caused prolongation of action potentials and abnormal autonomous depolarizations in murine atrial cardiomyocytes. Action potentials of atrial myocytes isolated from $Tbx5^{fl/fl};R26^{CreERt2}$ mice 7 days after tamoxifen treatment and paced at 0.5 Hz were significantly prolonged during phases 2-3 of the AP, compared to $R26^{CreERt2}$ atrial myocytes (Fig 2.2A); time to 90% repolarization (APD90) and 50% repolarization (APD50) were both prolonged. Early after depolarizations (EADs), delayed after depolarizations (DADs), and spontaneous triggered action potentials were all observed frequently in $Tbx5^{fl/fl};R26^{CreERt2}$ myocytes but never in $R26^{CreERt2}$ myocytes isolated 7 days post-TM and paced at 0.5 Hz (Fig 2.2B). These triggers are consistent with the initiation of paroxysmal AF, the most common form of AF (*2, 17, 18*). The aberrant depolarizations observed in isolated atrial myocytes provide evidence for an arrhythmic trigger after *Tbx5* removal.

Calcium-mediated inappropriate depolarizations of atrial myocardium have been implicated as an AF mechanism (*1*). Cytosolic calcium transients were prolonged in atrial myocytes isolated from $Tbx5^{fl/fl};R26^{CreERt2}$ mice but not $R26^{CreERt2}$ controls 7 days after tamoxifen treatment (Fig 2.2C). The decay constant, time to 50% decay, and time to peak were significantly prolonged in $Tbx5^{fl/fl};R26^{CreERt2}$ compared with $R26^{CreERt2}$ atrial myocytes. To assess whether prolonged calcium transients were the cause of prolonged action potentials in $Tbx5^{fl/fl};R26^{CreERt2}$ atrial myocytes, we current-clamped isolated atrial myocytes after including

calcium-chelator BAPTA (5 mM) in the intracellular solution. BAPTA rescued AP prolongation in *Tbx5*$^{fl/fl}$*;R26*^{CreERt2} atrial myocytes at APD50 and APD90 (Fig 2D). Calcium chelation by BAPTA also rescued abnormal depolarizations in *Tbx5*$^{fl/fl}$*;R26*^{CreERt2} atrial myocytes (p=1.0 vs *R26*^{CreERt2}, p=1x10^{-4} vs *Tbx5*$^{fl/fl}$*;R26*^{CreERt2} without BAPTA, by two-tailed Fisher Exact Test) (Fig 2D). Thus, disrupted calcium handling causes AP abnormalities in adult-specific *Tbx5* mutant mice.

Disrupted expression of AF-susceptibility genes in adult-specific *Tbx5*-mutant atria

To explore the molecular mechanism underlying AF, we identified TBX5-dependent transcripts by RNA-Seq transcriptional profiling left atria isolated from *Tbx5*$^{fl/fl}$*;R26*^{CreERt2} and *R26*^{CreERt2} 1 week after tamoxifen treatment, prior to the onset of AF. Numerous genes critical to atrial rhythm and approximately half the genes previously linked to AF (*3*) were significantly downregulated in *Tbx5*$^{fl/fl}$*;R26*^{CreERt2} left atria by RNA-seq (Fig 3A-C, Table 1). By RNA-Seq, we noted significant downregulation of numerous genes required for calcium handling, including *Ryr2*, *Atp2a2*, and *Sln* (Table 2.1). There was a modest and significant upregulation of calmodulins *Calm2* and *Calm3*, but no change in expression of CaV1.2 (*Cacna1c*), CamkII (*Camk2a/b/d/g*), or NCX (*Slc8a1/Slc8a2/Slc8a3*) (Table 2.1).

We validated the dysregulated genes critical to atrial conduction by RT-PCR, including the sarcoplasmic reticulum calcium handling genes *Ryr2*, *Sln* and *Atp2a2* (SERCA2), connexins *Gja1* and *Gja5*, sodium channels *Scn5a* and *Hcn4*,potassium channel subunits *Kcnj3*, *Kcnd3*, *Kcnn2*, *Kcnj8*, *Kcna5*, *Kcnj2*, *Kcnj5*, and *Abcc9*, and desmoplakin (*Dsp*) (Fig 3A-C; Table 2.1). Expression of some channels previously linked to AF, including *Kcnq1*, *Kcne1*, *Kcnh2*, and *Scn1b*, were unchanged, indicating specific gene expression changes after *Tbx5* deletion. These

data establish a TBX5-directed gene regulatory network (GRN) for atrial rhythm.

A *PITX2* cis-regulatory element is modulated by TBX5

PITX2 is the most frequently reported human AF GWAS locus and was significantly downregulated in the left atrium of $Tbx5^{fl/fl};R26^{CreERt2}$ mice (Fig 2.3C, Table 2.1). We hypothesized that the genomic region comprising the AF GWAS signal at *PITX2,* 100 kbp upstream of the *PITX2* promoter (*3, 7, 19, 20*), may harbor TBX5-dependent *cis*-regulatory elements (CREs). To identify candidate regulatory elements, we performed Assay for Transposase Accessible Chromatin (ATAC) sequencing (*21*), which reveals open chromatin, indicative of CRE activity. ATAC-Seq was performed on cardiac cells derived from human induced pluripotent stem (iPS) cells. A strong signal indicative of open chromatin was located in a single region upstream of *PITX2*, defining a candidate CRE (hg19 chr4:111,711,915-111,716,751) (Fig 3D). This region, located 150-155 kb 5' to the *PITX2* transcriptional start site, also harbored marks for DNase-I hypersensitivity (*22*), sequence conservation (*22*) from ENCODE data, and a canonical TBX5 motif, AGGTG (hg19 chr4: 111,712,182-111,712,186) (Fig 2.3D) (*23*).

We performed high resolution chromosome confirmation capture (4C) in murine left atrial tissue to identify genomic regions that interacted with the *Pitx2c* promoter on a genome-wide scale, using the *Pitx2c* promoter as a viewpoint (Fig 3D). The region syntenic to the candidate CRE made contacts with the *Pitx2c* promoter, indicating a *cis*-interaction (Fig 3D). This topology, indicating proximity between the CRE and the *Pitx2c* promoter, supports the region as a candidate *Pitx2* CRE. We then assayed the candidate CRE for TBX5 occupancy by ChIP-qPCR in human left atrial tissue. The CRE was significantly enriched 145-fold by TBX5

ChIP compared with a control locus (Fig 2.3E), indicating TBX5 occupancy at this CRE in human left atria.

The CRE significantly activated transcription in response to TBX5 expression in HEK293T cells and in the HL-1 atrial cardiomyocyte cell line, which expresses endogenous *Tbx5* (Fig 2.3F-G). A mutant CRE with the T-box binding site ablated (AGGTG to TTTTT) failed to activate expression in response to TBX5 in HEK293T cells (p=0.8 versus blank vector control; t-test) or in HL-1 cells (p=0.53 versus blank vector control; t-test) (Fig 2.3, F and G). The CRE harbors a common SNP, rs1906595, residing in the T-box binding site. The minor allele (G) (29%) is in perfect LD with the AF signal tagged by SNP rs2200733 (*24*). The minor allele (G) completes the canonical T-box binding element (AG**G**TG); whereas the major SNP allele (T) disrupts a central nucleotide of the T-box binding motif (AG**T**GT). We found that the major allele completely abolished CRE activity in response to *TBX5* expression in HEK and HL-1 cells and (Fig 2.3F-G). Together these findings identified a TBX5-dependent CRE at *PITX2* and suggested that TBX5 and PITX2 may co-regulate a common atrial gene regulatory network.

Scn5a, Gja1, Dsp, Ryr2, and *Atp2a2* are co-regulated by TBX5 and PITX2

We hypothesized that TBX5 and PITX2 co-regulate a gene regulatory network essential for atrial rhythm. We compared atrial TBX5-dependent transcripts (Fig 3A-C) with PITX2-dependent cardiac transcripts (*11*). *Scn5a*, *Gja1*, *Dsp*, *Ryr2*, and *Atp2a2* were significantly dysregulated genes in both *Tbx5* and *Pitx2* deletion models. These genes are all critical to atrial conduction, have been previously linked to AF (*1, 25, 26*), and were shown to be down-regulated in *Tbx5*-mutant and up-regulated in *Pitx2*-mutant adult mouse hearts (*11*).

To test the hypothesis that TBX5 and PITX2 directly co-regulated the expression of these

genes, we defined regions with overlapping chromosomal occupancy of both PITX2 and TBX5 from published ChIP datasets (*11, 23*). Candidate elements were refined by sequence conservation and co-localization of active enhancer marks p300 and H3k27ac (*23*). The candidate enhancers at all 5 loci demonstrated significant activation in response to TBX5 expression in HEK293T cells: *Gja1 (*mm9 chr10: 88816980-88817471), *Ryr2* (mm9 chr13: 12153077-12153901), *Atp2a2* (mm9 chr5: 122923655-122924226), *Scn5a* (mm9 chr9: 119562160-119563164), and *Dsp* (mm9 chr13: 38244307-38245159) (Fig 2.3H). All of the enhancers also showed significant positive responsiveness in HL-1 atrial cardiomyocytes, which was abrogated by mutation of the T-box binding motifs in each regulatory element. PITX2 co-expression blunted TBX5-dependent activation in all cases (Fig 2.3H). These findings demonstrate that TBX5 and PITX2 directly and antagonistically co-regulate an atrial rhythm gene regulatory network and describe an incoherent feed-forward loop transcriptional architecture, driven by TBX5 and repressed by PITX2.

Reduced *Pitx2* rescued atrial gene expression abnormalities and atrial arrhythmias caused by reduced *Tbx5*

The TBX5/PITX2 incoherent feed-forward loop model predicts that the effects of decreased *Tbx5* dose may be mitigated by decreased *Pitx2* dose. Adult-specific *Tbx5* heterozygous mice ($Tbx5^{fl/+};R26^{CreERt2}$) but not *Pitx2c* heterozygotes ($Pitx2c^{fl/+};R26^{CreERt2}$) showed diminished left atrial expression of *Tbx5* compared with $R26^{CreERt2}$ mice (Fig 2.4A, Table 2.2). $Tbx5^{fl/+};R26^{CreERt2}$ left atria also showed significantly diminished expression of 4 of the 5 shared *Tbx5/Pitx2* targets versus control animals: *Scn5a, Ryr2, Gja1,* and *Atp2a2*. $Pitx2^{fl/+};R26^{CreERt2}$ heterozygotes showed reduced left atrial *Pitx2* and increased *Scn5a,* but no

change of *Tbx5, Ryr2, Gja1, Dsp,* or *Atp2a2* expression versus $R26^{CreERt2}$ mice (Fig 4A, Table 2.2). *Scn5a, Ryr2, Gja1* and *Dsp* expression were all significantly increased in adult-specific compound *Tbx5;Pitx2* heterozygote ($Tbx5^{fl/+};Pitx2^{fl/+};R26^{CreERt2}$) mice compared to *Tbx5* heterozygote mice. In fact, expression of these *Tbx5*/*Pitx2* co-targets was normalized in compound *Tbx5;Pitx2* heterozygotes to their expression in control $R26^{CreERt2}$ mice. Thus, reduced *Tbx5* dose caused decreased expression of *Tbx5/Pitx2* co-targets that was rescued by reduced *Pitx2* dose.

We tested whether atrial rhythm was *Tbx5* dosage sensitive by examining adult-specific *Tbx5* heterozygotes. $Tbx5^{fl/+};R26^{CreERt2}$ mice demonstrated an irregularly irregular rhythm by surface ECG 2 weeks post-TM administration (Fig 2.4B, fig 2.S5A). AF was reproducibly induced in 6 out of 7 $Tbx5^{fl/+};R26^{CreERt2}$ mice using catheter-directed intracardiac pacing with either an S1/S2 coupling interval or burst pacing; whereas 0 out of 10 $R26^{CreERt2}$ mice experienced AF (Fig 2.4E, fig 2.S5B). Atrial myocytes isolated from $Tbx5^{fl/+};R26^{CreERt2}$ heterozygotes demonstrated inappropriate depolarizations and prolonged atrial action potentials at both 50% and 90% repolarization (Fig 4D).

Atrial rhythm was also sensitive to *Pitx2* dosage. Adult-specific *Pitx2* haploinsufficiency ($Pitx2^{fl/+};R26^{CreERt2}$) caused no abnormalities of p-wave duration or PR-interval, the time interval between the ECG P-wave and QRS, by surface ECG (Fig 2.4C). However, $Pitx2^{fl/+};R26^{CreERt2}$ atrial myocytes showed shortened action potential durations compared to $R26^{CreERt2}$ mice as well as increased action potential amplitude (Fig 2.4D). Furthermore, $Pitx2^{fl/+};R26^{CreERt2}$ mice were susceptible to pacing-induced AF (Fig 2.4E, fig 2.S5B).

Remarkably, the atrial rhythm and cellular electrophysiology abnormalities caused by reduced *Tbx5* dose were all rescued by reduced *Pitx2* dose. Compound adult-specific

Tbx5$^{fl/+}$;*Pitx2*$^{fl/+}$;*R26*^{CreERt2} mice showed no atrial rhythm instability or other ECG abnormalities, by surface ECG or Poincaré analysis 2 weeks after tamoxifen treatment (Fig 2.4B, fig 2.S5A). P-wave duration, prolonged in *Tbx5*$^{fl/+}$;*R26*^{CreERt2} mice 1 week after TM administration, was rescued in *Tbx5*$^{fl/+}$;*Pitx2*$^{fl/+}$;*R26*^{CreERt2} mice (Fig 2.4C, fig 2.S5A). No discernible PR interval difference was observed between any of these groups (Fig 2.4C).

Furthermore, decreased *Pitx2* dose rescued the cellular electrophysiology abnormalities observed in *Tbx5*$^{fl/+}$;*R26*^{CreERt2} mice. APD90 prolongation was rescued in *Tbx5*$^{fl/+}$;*Pitx2*$^{fl/+}$;*R26*^{CreERt2} atrial myocytes (Fig 2.4D). Additionally, the ectopic depolarizations observed in *Tbx5*$^{fl/+}$;*R26*^{CreERt2} were also rescued in *Tbx5*$^{fl/+}$;*Pitx2*$^{fl/+}$;*R26*^{CreERt2} mice (Fig 2.4D). Finally, *Tbx5*$^{fl/+}$;*Pitx2*$^{fl/+}$;*R26*^{CreERt2} mice were not susceptible to AF induction by intracardiac pacing (Fig 2.4E, fig 2.S5B). We conclude that the defects in atrial rhythm, AF susceptibility, and cellular electrophysiology caused by reduced*Tbx5* dose were all rescued by reduced *Pitx2* dose.

Discussion

This study links two cardiac transcription factors, TBX5 and PITX2, in a regulatory circuit that controls atrial rhythm. Removal of *Tbx5* profoundly disrupts cardiac channel gene expression and causes primary, spontaneous AF in mice. Despite the high incidence and healthcare burden of AF, there is a limited understanding of the mechanisms underlying its genetic susceptibility. These data provide mechanistic support for a TBX5-driven gene regulatory network in AF risk and link human AF GWAS loci in atrial rhythm control.

We generated a model of primary spontaneous AF in the absence of other cardiac pathology or developmental defects by removing *Tbx5* from the adult mouse. *Tbx5* deletion was sufficient to produce the two components of AF pathophysiology: a "trigger" (ectopic depolarizations) and "substrate" that propagates the trigger (*1*). Atrial myocytes demonstrated triggered activity, including EADs, DADs, and phase 4 depolarizations. Ex vivo optical mapping revealed slowing of atrial conduction velocity, a substrate allowing arrhythmia propagation through reentry. Trigger and substrate after *Tbx5*-removal may explain the resulting primary spontaneous AF, previously elusive in mice.

We then defined a *Tbx5*-dependent network that accounts for both ectopic depolarizations and prolonged action potentials in adult-specific *Tbx5*-mutant mice. Calcium cycling genes can cause ectopic depolarizations (*1*), and we observed direct regulation of *Ryr2* and *Atp2a2* by TBX5. We hypothesize that the primary mechanism for the prolonged calcium transient and prolonged action potential is the decrement in SERCA2 (*Atp2a2*) expression. Decreased SR Ca^{2+} reuptake resulted in prolongation of cytosolic calcium transient, which in turn may contribute to both prolonged action potentials and propensity for after-depolarizations. However, we cannot discount the role of other calcium handling genes, such as *Ryr2* and *Sln*, in the arrhythmogenic

cellular electrophysiologic phenotype. The adult-specific *Tbx5*-mutant atrial cardiomyocyte phenotype was rescued by calcium chelation, demonstrating the importance of altered calcium cycling in this setting.

Decreased expression of potassium channel subunits that modulate $I_{K1,}$ I_{Kr} and I_{to1} currents have been implicated in prolonged action potentials and abnormal triggering propensity (*1*). In this vein, expression of *Kcna7, Kcnd2, Kcnd3, Kcng2, Kcnj11, Kcnj2, Kcnj3, Kcnj5, Kcnj8, Kcnk3,* and *Kcnv2* were all significantly decreased after *Tbx5* deletion. Slowed conduction was observed by ex vivo mapping, likely owing to decreased expression of sodium channel gene *Scn5a* and high conductance gap junction connexins *Gja5* and *Gja1*. Consistent with our previous work (*27*), we did not observe significant expression of *Scn10a* in the adult mouse atria. Thus, *Tbx5* is a driver of rhythm control genes with specific roles, including those that control action potential fidelity and myocardial conduction speed. Decreasing the output of this network causes both arrhythmogenic trigger and substrate, providing a mechanism for the observed arrhythmias (Fig 2.5).

The intersection of *Tbx5*- and *Pitx2*-dependent transcripts identified a subset of essential atrial rhythm genes oppositely regulated by TBX5 and PITX2. We showed here that decreased transcriptional activator *TBX5* or decreased transcriptional repressor *PITX2* cause opposite effects on downstream effector gene regulatory network expression, yet both cause AF susceptibility *(11)*. Thus, AF can occur as a common phenotypic endpoint of opposite perturbations of atrial rhythm gene expression. This observation suggests that substratification of AF patients based on underlying molecular mechanism may afford personalized approaches to this common disorder.

The TBX5-PITX2 loop links critical AF loci and anticipates the identification of genetic variation functionally responsible for AF risk and future genotype-based AF risk stratification. We identified common genetic variation that associates with AF and affects a CRE at *PITX2*. SNP rs1906595 resides within a T-box binding site in the *TBX5*-driven enhancer identified at *PITX2*. AF risk associates with the minor allele and canonical T-box site, and therefore correlates with increased *TBX5*-driven enhancer activity. We speculate that this SNP may increase *PITX2* expression and thereby decrease downstream cardiac channel gene regulatory network expression. This model is consistent with previously published work, in which AF risk SNPs at *PITX2* are associated with increased PITX2 expression (*28*).

We defined the architecture of an atrial gene regulatory network as an incoherent feed-forward loop, driven by TBX5 and repressed by PITX2 (Fig 2.5B). A TBX5-dependent CRE was identified at PITX2, and TBX5- and PITX2-dependent CREs were identified at shared target loci, genes critical to calcium handling (*Ryr2, Atp2a2),* rapid depolarization (*Scn5a),* and intracellular communication (*Gja1, Dsp).* TBX5 and PITX2 acted directly and antagonistically at the terminal atrial channels of the atrial rhythm gene regulatory network. TBX5/PITX2 antagonism was also observed genetically, in which the physiologic effects of reduced *Tbx5* dose, including decreased expression of critical AF genes, atrial rhythm instability, cellular action potential abnormalities, and AF susceptibility were all rescued by reduced *Pitx2* dose.

The robustness of atrial contraction cycles requires the coordinated expression of genes governing action potential fidelity, cell-cell conduction, and free calcium. Variation in these properties is arrythmogenic (*1*). Incoherent feed-forward loops impose molecular buffering on terminal regulatory network transcriptional output in response to upstream variance in gene expression (*29-31*). The *Tbx5/Pitx2* transcriptional architecture is therefore ideal for an atrial

gene regulatory network, which must maintain uniform channel gene expression and cellular electrophysiological behavior for normal atrial rhythm for billions of iterations. The *TBX5*-driven *PITX2*-modulated incoherent feed-forward loop provides a molecular logic for atrial rhythm homeostasis.

The transcriptional architecture we describe is based on *in vitro* studies and *in vivo* genetic experiments in mice. Although the *cis*-regulatory elements we tested were human elements or conserved in humans, we have yet to formally observe the Incoherent Feed Forward Loop transcriptional architecture in human atria. Essential future steps include an evaluation of the transcriptional logic of atrial rhythm in humans, the identification of genetic variation in the essential regulatory elements comprising the network, and testing of the functional impact of variants on atrial gene expression and atrial rhythm control. TBX5 and PITX2 do not act in isolation, but are undoubtedly part of a much larger atrial transcriptional complex. How the complete compendium of transcriptional components interact to effect atrial rhythm homeostasis and how genetic variation in the broader atrial gene regulatory network affects atrial rhythm provide opportunities for future studies.

An essential goal of the functional genetics and genomics approaches applied here is to transition from genetic implication of GWAS to molecular mechanism underlying the genetics of AF risk. This work supports a model in which AF is a common endpoint disease process resulting from opposite perturbations of an atrial rhythm control network. This model suggests that a genotype-based personalized approach to the treatment of AF may ultimately be possible and necessary. We expect that these and analogous efforts will contribute to improved platforms for disease risk determination and therapeutic stratification.

Materials and Methods

Study Design

The objective of our study was to investigate the role of TBX5 in atrial rhythm. We used murine models of conditional *Tbx5* deletion, conditional *Pitx2* deletion, and *in vitro* assays in murine HL-1 cells and human iPS cell-derived cardiomyocytes. Mouse models were chosen for their similarity in cardiac physiology to humans, and for the availability of specific genetic tools. Human iPS cell-derived cardiomyocytes were used as an *in vitro* model of the regulatory loci. Sample sizes were chosen based on power calculations after pilot studies to estimate population mean and standard deviation. All recordings and analyses were conducted in a blinded fashion. Genotype controls were used, and mice randomized into groups when appropriate. Endpoints for studies were selected based on the progression of phenotype. Samples were excluded if replicates were >2 standard deviations away from the population mean. Replication for specific experiments is in the figure legends.

Generation of conditional Tbx5-deleted mice

Tbx5 was deleted from the adult mouse using the published *Tbx5* floxed allele ($Tbx5^{fl}$) (*5*), as described in Supplementary Methods.

Electrocardiogram Recordings

Mice were implanted with subcutaneous telemetry transmitters (ETA-F10; Data Science International) (*15*). Arrhythmia analysis was carried out using Ponemah Physiology Platform software. Signal average and Poincaré plots were generated with a custom Python script.

Catheter-based intracardiac electrophysiology is described in Supplementary Methods.

Myocyte isolation and action potential recordings

To isolate mouse cardiomyocytes, hearts were excised and mounted on a Langendorf apparatus and retrogradely perfused with collagenase type 2 (Worthington Biochemical,) in perfusion buffer. Membrane potential was recorded with a ruptured patch current clamp at 37°C using an Axon Axopatch 200B amplifier. During recording cells were perfused in normal Tyrode buffer. Action potentials were triggered using 0.5 nA x 2 ms current clamp pulses. When noted, 5-mM BAPTA-tetrapotassium was included in the pipette solution to abolish cytosolic calcium transients. In all cases liquid junction potentials were corrected prior to recording action potentials. Additional details are in Supplementary methods.

RNA-Seq and gene expression analysis

Left atrial free wall from $Tbx5^{fl/fl};R26^{CreERt2}$ and $R26^{CreERt2}$ mice was removed and total RNA prepared with a Trizol-based extraction. Sequencing libraries were prepared with riboZero purification (Illumina) and sequenced on an Illumina HiScanSQ Instrument. Samples were sequenced 50 bp single-ended at 10-20 million reads per replicate. Library preparation and sequencing was performed at the University of Chicago Genomics Core. Sequence was aligned to the mouse genome with Tophat2, and differential expression testing performed with the DESeq2 pipeline. Heatmap and Volcano Plot were generated in R with the ggplot2 and gplot packages, respectively. Candidate targets were validated with quantitative real time PCR using exon spanning primers specific to the gene of interest in a distinct cohort of biological replicates.

ChIP-qPCR of human atrial tissue

Left atrial appendage tissue was collected from three patients undergoing robotic valvular surgery and left atrial appendage plication (IRB 12797B). All patients were genotyped as heterozygous at SNP rs1906595. ChIP was performed as previously described using a ChIP grade TBX5 antibody (sc-17866, Santa Cruz Biotechnology) (*32*). To determine fold enrichment, qPCR was performed using input controls compared with DNA bound to immunoprecipitated proteins, using primers specific to the site of interest as well as primers to a site not expected to be enriched. Control primers to the human *GAPDH* locus: 5'–TACTAGCGGTTTTACGGGCG–3' and 5'– TCGAACAGGAGGAGCAGAGAGCGA. Primers tiled across the candidate enhancer were 5'-AGTGGCATCAAGACAGCACA-3' and 5'-CCCCGGATCACCAAATCCAAG-3'; 5'-GTGGGCTGGGTGACTGTATT-3' and 5'-CCCTGCACTCATGCTGGTTA-3'; 5'-GCTGCACAACTTAGCTGCAA-3' and GCCAGAAACAACCTCAAAGCA.

Statistical analysis

For comparison of quantitative metrics of action potentials, calcium transients, gene expression by qRT-PCR, ChIP-qPCR, luciferase response, and ECG parameters, two-tailed t-tests were used to test significance, with a Bonferroni correction for multiple comparisons when appropriate. For count-based analysis of cellular depolarization events and AF inducibility, two-tailed Fisher's Exact Tests were used. RNA-Sequencing differential expression was performed with the DE-Seq2 pipeline, with a false discovery rate <0.05.

List of Supplementary Materials

Materials and Methods

Fig. S2.1. No change in cardiac function after onset of AF in *Tbx5*$^{fl/fl}$*;R26*^{CreERt2} mice

Fig. S2.2. *Tbx5*$^{fl/fl}$*;R26*^{CreERt2} mice develop ventricular arrhythmias following atrial arrhythmias

Fig. S2.3. *Tbx5*$^{fl/fl}$*;R26*^{CreERt2} show paroxysmal atrial fibrillation

Fig. S2.4. Optical action potential from right atrium of *Tbx5*$^{fl/fl}$*;R26*^{CreERt2} mice

Fig. S2.5. Atrial rhythm instability and AF inducibility in *Tbx5*$^{fl/+}$*;R26*^{CreERt2} mice is rescued by *Pitx2* haploinsufficiency

Supplementary Materials and Methods

In vivo electrophysiology

Animals underwent catheter-based intracardiac recordings. Right atrial and ventricular endocardial electrograms were recorded using a 1.1-F octapolar catheter (EPR-800, Millar Instruments) or 1.9-F octapolar catheter with 0.5-mm spacing (Scisense) connected to ADI BioAmp and PowerLab apparatus and recorded using LabChart Software (ADInstruments) using a right internal jugular venous approach. Mice were anesthetized using 2-3% isoflurane in 100% oxygen, titrated to heart rate between 400-450 bpm and secured to a heating pad with body temperature regulated to 37°C using a rectal probe. Subcutaneous needle electrodes placed just under the skin were used to obtain surface electrocardiograms of leads II and III. A vertical skin cut-down at the right jugular vein was performed, and the vein was isolated from connective tissue. A small transverse nick was made in the vein, and the catheter was advanced through the nick and jugular vein to the right ventricle using electrograms as guides. Recordings were made of the atrium and ventricle. AH and HV intervals were measured and all signals were averaged and obtained over at least three minutes per recording. The catheter was connected to a Grass SD9 Stimulator (Astro-Med) controlled by a custom Python-based pacing program on the attached PC. Atrial and ventricular pacing protocols were carried out to obtain atrial, atrio-ventricular and ventricular refractory periods (AERP, AVERP, VERP) as well as sino-atrial recovery and AV Wenckebach times. Effective refractory periods were measured using 8-beat S1 drive trains of 100 ms followed by single extra-stimulus. Atrial and ventricular induction pacing was carried out using burst pacing at 1200 pps for one second and 5-fold extra-stimulation at 50 ms using S1 drive trains of 80 ms. Standard intervals and refractory periods were obtained (*33, 34*). Pacing protocols were carried out prior between 1-2 weeks after

tamoxifen administration. A maximum of two minutes of induction stimulation was applied per mouse. The presence of at least three ventricular cycles of atrial tachycardia or fibrillation at least twice in the same animal was considered to be a positive test in that subject.

Echocardiography

Echocardiographic analysis of heart structure and function was performed at the same timepoints as surface electrocardiography using techniques as described (*15*). Mice were anesthetized with inhaled isoflurane (~2%) delivered via nose cone. The left precordial fur was removed with a topical depilatory agent. Body temperature was maintained with a heated imaging platform. The heart was imaged with a VisualSonics Vevo 770 device (VisualSonics) using a 30 MHz high-frequency transducer. Two-dimensional images were recorded in approximately the parasternal long- and short-axis projections with guided M-mode recordings at the mid-ventricular level in both views. Left ventricular internal dimensions at diastole and systole (LVIDd and LVIDs, respectively) were measured in at least three beats from each projection and averaged, and fractional shortening was calculated, and used to measure LV ejection fraction. When mice were in atrial fibrillation, at least five beats were analyzed and averaged. Echocardiographic measurements were obtained from M-mode images, at the mid-papillary muscle level in the parasternal short-axis view, and also from B-mode images acquired in the parasternal long- and short-axis views. Mitral valve pulsed wave Doppler recordings were used to generate the E/A wave ratio.

Ex vivo optical mapping

Isolated heart preparations were dissected as described (*35*). Briefly, mice were anesthetized with a mixture of ketamine and xylazine with 100 units heparin. After a midsternal incision, the heart was removed and placed on Langendoff retrograde perfusion and superfusion apparatus. The isolated heart was pinned at the apex to bottom of the chamber to prevent stream-induced movement. The right atrial and left atrial appendages were stretched and pinned to flatten them and allow optical measurements from the maximal surface area of the atria. The excitation-contraction uncoupler blebbistatin (10 μM, Tocris Bioscience) was used to eliminate motion artifacts from the optical signals.

Langendorff-perfused hearts were stained with a voltage-sensitive dye RH-237 (5 μl dissolved in 1 mg/ml DMSO) over a period of 5–7 min. Optical potentials were recorded using fluorescent light emitted from the preparation at >650 nm, which was excited by 530–40 nm light source. Emitted light was recorded by a MiCAM Ultima-L CMOS camera (SciMedia) with high spatial (100 x 100 pixels, 230 ± 20 μm/pixel) and temporal (1000–3000 frames/s) resolution. The acquired fluorescent signal was analyzed with custom-developed software (*36*).

Calcium transient measurement

Cytosolic $[Ca2+]_i$ was measured with the high-affinity Ca2+ indicator Fluo-4 (Molecular Probes/Invitrogen,) using a laser scanning confocal microscopy (Zeiss LSM 510,) equipped with a ×63/1.40 NA oil-immersion objective lens. Isolated atrial myocytes were plated on laminin coated coverslips and incubated at room temperature with 10 μM Fluo-4/AM for 15 min in normal Tyrode's solution containing (in mM): 140 NaCl, 4 KCl, 10 glucose, 10 HEPES, and 1 $MgCl_2$, 1 $CaCl_2$, with pH 7.4 using NaOH, followed by a 10 min perfusion wash with Tyrode (at 37˚C) without Fluo-4 before recording. Fluo-4 was excited with the 488-nm line of an argon

laser and fluorescence was measured at >515 nm. Atrial APs were elicited using electrical field stimulation (Grass stimulator; Astro-Med) with submerged platinum electrodes in the perfusion bath. Voltage was set at the AP threshold for ~50% of the plated myocyte population. Ca^{2+} transients were acquired in line-scan mode (3 ms per scan; pixel size 0.12 µm). Ca^{2+} transients are presented as total fluorescence intensity normalized to resting fluorescence (F/F_0) during steady-state conditions before field stimulation.

Preparation of 4C templates

4C templates were prepared as described previously (*37*). In short, 10 adult left atria were isolated in ice cold PBS and pooled for homogenization. Single cell suspensions were obtained by dissociation of tissue with IKA Ultra Turrax T5 FU, followed by dounce homogenization. Chromatin was cross-linked with 2% formaldehyde in PBS with 10% FCS for 10 min at room temperature, nuclei were isolated and cross-linked DNA was digested with a primary restriction enzyme recognizing a 4-bp restriction site (DpnII), followed by proximity ligation. Cross-links were removed and a secondary restriction enzyme digestion (Csp6I), followed again by proximity ligation. For all experiments, 200 ng of the resulting 4C template was used for the subsequent PCR reaction, of which 16 (total: 3.2 µg of 4C template) were pooled and purified for next-generation sequencing. The PCR products were purified using two columns per sample of the High Pure PCR Product Purification Kit (Roche cat. no. 11732676001). The kit separates the PCR products that are larger than 120 bp from the adaptor-containing primers [which are ~75 nucleotides (nt) and ~40 nt in size, respectively].

4C-seq primer design

PCR primers were designed based on the following criteria. The size of the viewpoint fragment was at least 500 bp to allow efficient cross-linking to other DNA fragments. The fragment end (the region between the primary and secondary restriction enzyme) was more than 350 bp to allow efficient circularization during the second ligation step. Primers were designed to be maximally 20 nucleotides in length. The strategy therefore produces sequencing reads (36-mers in this study) composed of the 4C primer sequence (20 nucleotides, specific to a given viewpoint) followed by 16 nucleotides that identify a captured sequence. The reading primer always hybridizes to, and ends at, the 3′ side of the first restriction recognition site. This design ensures analysis of only primary ligation events and provides sufficient sequence information to unambiguously identify most captured sequences. The nonreading primers, with sizes of 18-20 nucleotides, were designed at a distance of ≤100 bp from the secondary restriction site. All primers had a GC-content between 35 and 65% and an optimal basic temperature of 55°C, ranging from 45-65°C. Primers were checked against the mouse genome with MegaBLAST23 (settings -p 88.88 -W 12 -e 1 -F T), which requires primers on the reading side to be matched uniquely in the genome and primers on the nonreading side to have a maximum of three perfectly matching BLAST high-scoring segment pairs (HSP): Reading primer, AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCTGAAG AGGAAGGTGATACAGATC; non-reading primer, CAAGCAGAAGACGGCATACGACTCCCCAGGGATTTCAGT.

4C templates were mixed and sequenced simultaneously in one Illumina HiSeq 2000 lane. The sequence tags generated by the procedure are prefixed by the 4C reading primer that includes the DpnII restriction site sequence. The 4C reading primer sequences were separated from multiplexed 4C-seq libraries and the suffixes were extracted for further processing.

Mapping and filtering of the sequence reads was done as described previously (*38*). The algorithm constructs a background model for remote intra- and interchromosomal contacts to correct for systematic biases that can occur during the 4C-seq experimental protocol. The algorithm is designed to use controls for sequencing errors and non-unique sequences while considering the high coverage (100 to 100,000x) of fragment ends that are proximal to the viewpoint fragment.

iPSC differentiation

Cardiomyocytes were derived from human iPSCs as described previously (*39*).

ATAC Sequencing

ATAC-Seq was performed as described (*21*).

Cloning of candidate regulatory elements

Predictions of regulatory elements suggest that there may be upwards of 100,000 regulatory elements in the human genome, with less than 1000 validated (*34*). Candidate regulatory elements were amplified from HEK293T or C57/B6 mouse genomic DNA. Sequence was verified and then cloned into the pGL4.23 or pGL3 enhancer luciferase response vectors with minimal promoter.

In vitro luciferase response assays

HEK293T cells were co-transfected with the enhancer-luciferase construct (600 ng) and overexpression vector (100 ng) in the pcDNA backbone using FuGENE (Promega) as described

(*32*). Total DNA transfected was normalized with blank vector when necessary. HL-1 cells were transfected with enhancer-luciferase vector alone (600 ng). Cells were cultured for 48 hours after transfection, then lysed and assayed using the Dual-Luciferase Reporter Assay system (Promega).

References

1. J. Heijman, N. Voigt, S. Nattel, D. Dobrev, Cellular and molecular electrophysiology of atrial fibrillation initiation, maintenance, and progression. *Circulation research* **114**, 1483-1499 (2014).
2. K. Nishida, S. Nattel, Atrial fibrillation compendium: historical context and detailed translational perspective on an important clinical problem. *Circulation research* **114**, 1447-1452 (2014).
3. N. R. Tucker, P. T. Ellinor, Emerging directions in the genetics of atrial fibrillation. *Circulation research* **114**, 1469-1482 (2014).
4. B. G. Bruneau, M. Logan, N. Davis, T. Levi, C. J. Tabin, J. G. Seidman, C. E. Seidman, Chamber-specific cardiac expression of Tbx5 and heart defects in Holt-Oram syndrome. *Developmental biology* **211**, 100-108 (1999).
5. B. G. Bruneau, G. Nemer, J. P. Schmitt, F. Charron, L. Robitaille, S. Caron, D. A. Conner, M. Gessler, M. Nemer, C. E. Seidman, J. G. Seidman, A murine model of Holt-Oram syndrome defines roles of the T-box transcription factor Tbx5 in cardiogenesis and disease. *Cell* **106**, 709-721 (2001).
6. A. V. Postma, J. B. van de Meerakker, I. B. Mathijssen, P. Barnett, V. M. Christoffels, A. Ilgun, J. Lam, A. A. Wilde, R. H. Lekanne Deprez, A. F. Moorman, A gain-of-function TBX5 mutation is associated with atypical Holt-Oram syndrome and paroxysmal atrial fibrillation. *Circulation research* **102**, 1433-1442 (2008).
7. P. T. Ellinor, K. L. Lunetta, C. M. Albert, N. L. Glazer, M. D. Ritchie, A. V. Smith, D. E. Arking, M. Muller-Nurasyid, B. P. Krijthe, S. A. Lubitz, J. C. Bis, M. K. Chung, M. Dorr, K. Ozaki, J. D. Roberts, J. G. Smith, A. Pfeufer, M. F. Sinner, K. Lohman, J. Ding,

N. L. Smith, J. D. Smith, M. Rienstra, K. M. Rice, D. R. Van Wagoner, J. W. Magnani, R. Wakili, S. Clauss, J. I. Rotter, G. Steinbeck, L. J. Launer, R. W. Davies, M. Borkovich, T. B. Harris, H. Lin, U. Volker, H. Volzke, D. J. Milan, A. Hofman, E. Boerwinkle, L. Y. Chen, E. Z. Soliman, B. F. Voight, G. Li, A. Chakravarti, M. Kubo, U. B. Tedrow, L. M. Rose, P. M. Ridker, D. Conen, T. Tsunoda, T. Furukawa, N. Sotoodehnia, S. Xu, N. Kamatani, D. Levy, Y. Nakamura, B. Parvez, S. Mahida, K. L. Furie, J. Rosand, R. Muhammad, B. M. Psaty, T. Meitinger, S. Perz, H. E. Wichmann, J. C. Witteman, W. H. Kao, S. Kathiresan, D. M. Roden, A. G. Uitterlinden, F. Rivadeneira, B. McKnight, M. Sjogren, A. B. Newman, Y. Liu, M. H. Gollob, O. Melander, T. Tanaka, B. H. Stricker, S. B. Felix, A. Alonso, D. Darbar, J. Barnard, D. I. Chasman, S. R. Heckbert, E. J. Benjamin, V. Gudnason, S. Kaab, Meta-analysis identifies six new susceptibility loci for atrial fibrillation. *Nature genetics* **44**, 670-675 (2012).

8. H. Holm, D. F. Gudbjartsson, D. O. Arnar, G. Thorleifsson, G. Thorgeirsson, H. Stefansdottir, S. A. Gudjonsson, A. Jonasdottir, E. B. Mathiesen, I. Njolstad, A. Nyrnes, T. Wilsgaard, E. M. Hald, K. Hveem, C. Stoltenberg, M. L. Lochen, A. Kong, U. Thorsteinsdottir, K. Stefansson, Several common variants modulate heart rate, PR interval and QRS duration. *Nature genetics* **42**, 117-122 (2010).

9. N. Tan, M. K. Chung, J. D. Smith, J. Hsu, D. Serre, D. W. Newton, L. Castel, E. Soltesz, G. Pettersson, A. M. Gillinov, D. R. Van Wagoner, J. Barnard, Weighted gene coexpression network analysis of human left atrial tissue identifies gene modules associated with atrial fibrillation. *Circulation. Cardiovascular genetics* **6**, 362-371 (2013).

10. X. Zang, S. Zhang, Y. Xia, S. Li, F. Fu, X. Li, F. Wang, R. Zhang, X. Tian, L. Gao, J.

Zhang, Y. Yang, X. Tu, Q. Wang, SNP rs3825214 in TBX5 is associated with lone atrial fibrillation in Chinese Han population. *PLoS One* **8**, e64966 (2013).

11. Y. Tao, M. Zhang, L. Li, Y. Bai, Y. Zhou, A. M. Moon, H. J. Kaminski, J. F. Martin, Pitx2, an atrial fibrillation predisposition gene, directly regulates ion transport and intercalated disc genes. *Circulation. Cardiovascular genetics* **7**, 23-32 (2014).
12. G. Riley, F. Syeda, P. Kirchhof, L. Fabritz, An introduction to murine models of atrial fibrillation. *Frontiers in physiology* **3**, 296 (2012).
13. U. Alon, Network motifs: theory and experimental approaches. *Nature reviews. Genetics* **8**, 450-461 (2007).
14. A. Ventura, D. G. Kirsch, M. E. McLaughlin, D. A. Tuveson, J. Grimm, L. Lintault, J. Newman, E. E. Reczek, R. Weissleder, T. Jacks, Restoration of p53 function leads to tumour regression in vivo. *Nature* **445**, 661-665 (2007).
15. D. E. Arnolds, F. Liu, J. P. Fahrenbach, G. H. Kim, K. J. Schillinger, S. Smemo, E. M. McNally, M. A. Nobrega, V. V. Patel, I. P. Moskowitz, TBX5 drives Scn5a expression to regulate cardiac conduction system function. *The Journal of clinical investigation* **122**, 2509-2518 (2012).
16. D. Lang, M. Sulkin, Q. Lou, I. R. Efimov, Optical mapping of action potentials and calcium transients in the mouse heart. *Journal of visualized experiments : JoVE,* (2011).
17. J. Andrade, P. Khairy, D. Dobrev, S. Nattel, The clinical profile and pathophysiology of atrial fibrillation: relationships among clinical features, epidemiology, and mechanisms. *Circulation research* **114**, 1453-1468 (2014).
18. A. E. Darby, J. P. Dimarco, Management of atrial fibrillation in patients with structural heart disease. *Circulation* **125**, 945-957 (2012).

19. M. J. Kolek, B. Parvez, R. Muhammad, M. B. Shoemaker, M. A. Blair, T. Stubblefield, G. A. Kucera, J. C. Denny, D. M. Roden, D. Darbar, A common variant on chromosome 4q25 is associated with prolonged PR interval in subjects with and without atrial fibrillation. *Am J Cardiol* **113**, 309-313 (2014).

20. M. S. Olesen, A. G. Holst, J. Jabbari, J. B. Nielsen, I. E. Christophersen, A. Sajadieh, S. Haunso, J. H. Svendsen, Genetic loci on chromosomes 4q25, 7p31, and 12p12 are associated with onset of lone atrial fibrillation before the age of 40 years. *Can J Cardiol* **28**, 191-195 (2012).

21. J. D. Buenrostro, B. Wu, H. Y. Chang, W. J. Greenleaf, ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]* **109**, 21 29 21-29 (2015).

22. E. P. Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).

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

24. D. F. Gudbjartsson, D. O. Arnar, A. Helgadottir, S. Gretarsdottir, H. Holm, A. Sigurdsson, A. Jonasdottir, A. Baker, G. Thorleifsson, K. Kristjansson, A. Palsson, T. Blondal, P. Sulem, V. M. Backman, G. A. Hardarson, E. Palsdottir, A. Helgason, R. Sigurjonsdottir, J. T. Sverrisson, K. Kostulas, M. C. Ng, L. Baum, W. Y. So, K. S. Wong, J. C. Chan, K. L. Furie, S. M. Greenberg, M. Sale, P. Kelly, C. A. MacRae, E. E. Smith, J. Rosand, J. Hillert, R. C. Ma, P. T. Ellinor, G. Thorgeirsson, J. R. Gulcher, A. Kong, U. Thorsteinsdottir, K. Stefansson, Variants conferring risk of atrial fibrillation on

chromosome 4q25. *Nature* **448**, 353-357 (2007).

25. I. E. Christophersen, P. T. Ellinor, Genetics of atrial fibrillation: from families to genomes. *Journal of human genetics*, (2015).

26. S. Kostin, G. Klein, Z. Szalay, S. Hein, E. P. Bauer, J. Schaper, Structural correlate of atrial fibrillation in human patients. *Cardiovascular research* **54**, 361-379 (2002).

27. M. van den Boogaard, S. Smemo, O. Burnicka-Turek, D. E. Arnolds, H. J. van de Werken, P. Klous, D. McKean, J. D. Muehlschlegel, J. Moosmann, O. Toka, X. H. Yang, T. T. Koopmann, M. E. Adriaens, C. R. Bezzina, W. de Laat, C. Seidman, J. G. Seidman, V. M. Christoffels, M. A. Nobrega, P. Barnett, I. P. Moskowitz, A common genetic variant within SCN10A modulates cardiac SCN5A expression. *The Journal of clinical investigation* **124**, 1844-1852 (2014).

28. R. I. Martin, M. S. Babaei, M. K. Choy, W. A. Owens, T. J. Chico, D. Keenan, N. Yonan, M. S. Koref, B. D. Keavney, Genetic variants associated with risk of atrial fibrillation regulate expression of PITX2, CAV1, MYOZ1, C9orf3 and FANCC. *Journal of molecular and cellular cardiology* **85**, 207-214 (2015).

29. L. Goentoro, O. Shoval, M. W. Kirschner, U. Alon, The incoherent feedforward loop can provide fold-change detection in gene regulation. *Molecular cell* **36**, 894-899 (2009).

30. W. Ma, A. Trusina, H. El-Samad, W. A. Lim, C. Tang, Defining network topologies that can achieve biochemical adaptation. *Cell* **138**, 760-773 (2009).

31. S. Mangan, U. Alon, Structure and function of the feed-forward loop network motif. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 11980-11985 (2003).

32. A. D. Hoffmann, X. H. Yang, O. Burnicka-Turek, J. D. Bosman, X. Ren, J. D. Steimle, S.

A. Vokes, A. P. McMahon, V. V. Kalinichenko, I. P. Moskowitz, Foxf genes integrate tbx5 and hedgehog pathways in the second heart field for cardiac septation. *PLoS Genet* **10**, e1004604 (2014).

33. J. Gehrmann, C. I. Berul, Cardiac electrophysiology in genetically engineered mice. *Journal of cardiovascular electrophysiology* **11**, 354-368 (2000).

34. V. V. Patel, M. Arad, I. P. Moskowitz, C. T. Maguire, D. Branco, J. G. Seidman, C. E. Seidman, C. I. Berul, Electrophysiologic characterization and postnatal development of ventricular pre-excitation in a mouse model of cardiac hypertrophy and Wolff-Parkinson-White syndrome. *Journal of the American College of Cardiology* **42**, 942-951 (2003).

35. A. V. Glukhov, T. P. Flagg, V. V. Fedorov, I. R. Efimov, C. G. Nichols, Differential K(ATP) channel pharmacology in intact mouse heart. *Journal of molecular and cellular cardiology* **48**, 152-160 (2010).

36. J. I. Laughner, F. S. Ng, M. S. Sulkin, R. M. Arthur, I. R. Efimov, Processing and analysis of cardiac optical mapping data obtained with potentiometric dyes. *American journal of physiology. Heart and circulatory physiology* **303**, H753-765 (2012).

37. M. Simonis, J. Kooren, W. de Laat, An evaluation of 3C-based methods to capture DNA interactions. *Nature methods* **4**, 895-901 (2007).

38. H. J. van de Werken, G. Landan, S. J. Holwerda, M. Hoichman, P. Klous, R. Chachik, E. Splinter, C. Valdes-Quezada, Y. Oz, B. A. Bouwman, M. J. Verstegen, E. de Wit, A. Tanay, W. de Laat, Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nature methods* **9**, 969-972 (2012).

39. J. T. Hinson, A. Chopra, N. Nafissi, W. J. Polacheck, C. C. Benson, S. Swist, J. Gorham, L. Yang, S. Schafer, C. C. Sheng, A. Haghighi, J. Homsy, N. Hubner, G. Church, S. A.

Cook, W. A. Linke, C. S. Chen, J. G. Seidman, C. E. Seidman, HEART DISEASE. Titin mutations in iPS cells define sarcomere insufficiency as a cause of dilated cardiomyopathy. *Science* **349**, 982-986 (2015).

Acknowledgments: We thank H. Balkhy for assistance in obtaining tissue samples; C.-W. Wu for assistance in cellular electrophysiology studies; This research was supported in part by NIH through resources provided by the Computation Institute and the Biological Sciences Division of the University of Chicago and Argonne National Laboratory, under grant 1S10OD018495-01. We specifically acknowledge the assistance of Lorenzo Pesce. **Funding:** The study was funded by grants from the NIH (R01 HL114010 to IPM; R01HL118761, R01DE023177, and U54 HD083092 to J.F.M; R01 HL114395 and R01 HL126802 to I.R.E.; and HL128075 to E.M.M.), The Leducq Foundation (FP058566-01-PR, to IPM, JM and VC), and the American Heart Association (Established Investigator Award 13EIA14690081 to I.P.M.). This research was supported in part by NIH through resources provided by the Computation Institute and the Biological Sciences Division of the University of Chicago and Argonne National Laboratory, under grant 1S10OD018495-01. **Author contributions:** R.D.N. was involved in design execution, and analysis of all experiments; wrote the manuscript; performed statistical analysis. M.T.B. performed and analyzed whole animal electrophysiology and echocardiography; wrote the manuscript. B.B. performed and analyzed ex vivo optical mapping experiments; wrote the manuscript. S.R.M. performed and analyzed single cell electrophysiology experiments; wrote the manuscript. X.Y. analyzed genomics experiments; performed statistical analysis. M.v.B. performed and analyzed chromosome conformation capture; wrote the manuscript. J.B. performed and analyzed gene expression, chromatin immunoprecipitation, and cellular experiments. M.G. performed and analyzed whole animal electrophysiology, and chromatin immunoprecipitation experiments. Y.Q. performed and analyzed ex vivo optical mapping experiments. T.W. performed and analyzed iPS cardiomyocyte experiments; wrote the manuscript. M.Z. performed and analyzed enhancer identification, in vitro cellular experiments.

J.F.M. analyzed enhancer identification and in vitro cellular experiments and wrote the manuscript. C.E.S. and J.S. analyzed iPS cardiomyocyte experiments and wrote the manuscript. V.C. analyzed chromosome conformation capture experiments and wrote the manuscript. I.R.E. analyzed ex vivo optical mapping experiments and wrote the manuscript. E.M.M analyzed whole animal physiology, in vitro assays, and wrote the manuscript. C.R.W. performed and analyzed cellular electrophysiology experiments and wrote the manuscript. I.P.M. was involved in designing, performing, and analyzing all experiments; wrote the manuscript; performed statistical analysis. All authors approved the final version. **Competing Interests:** None. **Data and Materials Availability:** The RNA and ATAC sequencing data from this study have been deposited in the Gene Expression Omnibus database.

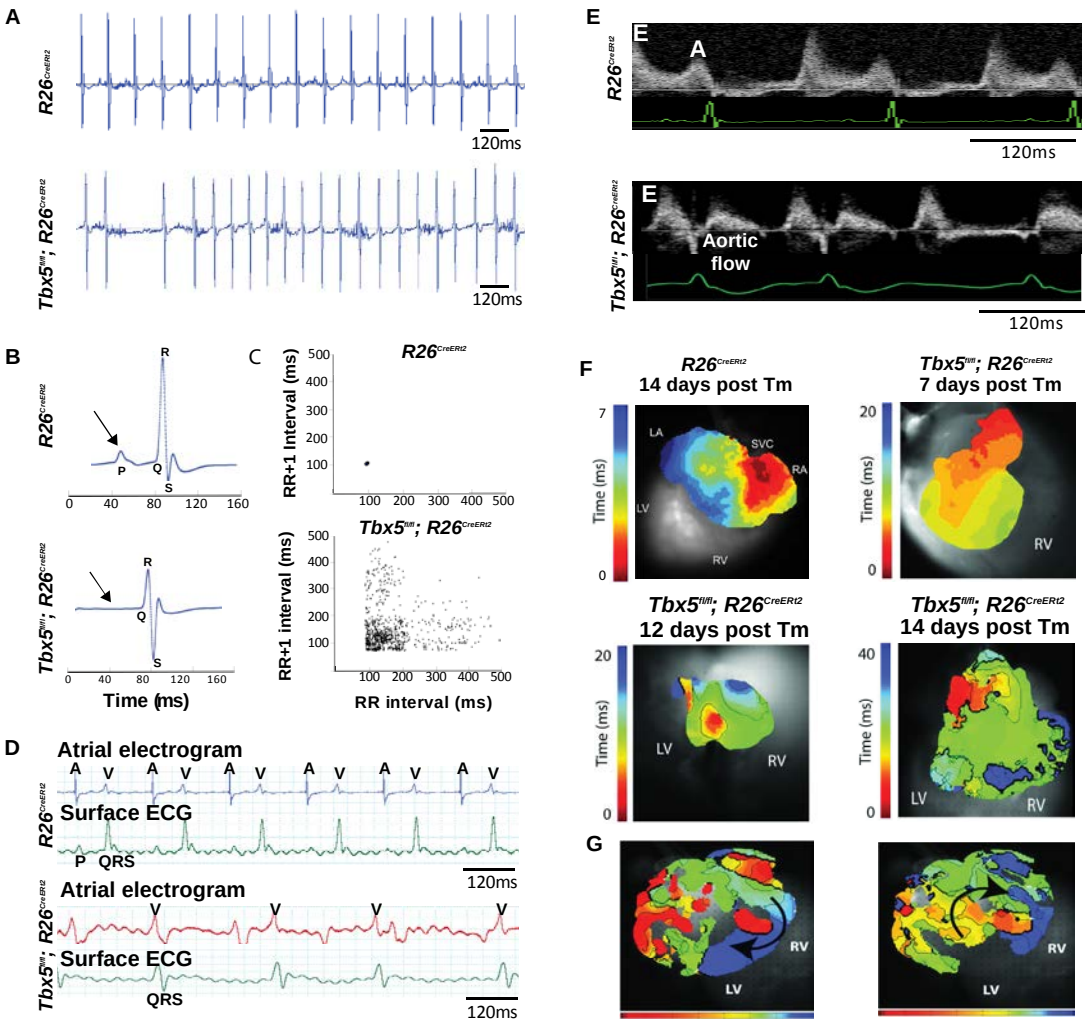

Figure 2.1. Removal of *Tbx5* from the adult mouse results in spontaneous, sustained atrial fibrillation. (A) Representative ambulatory telemetry electrocardiogram (ECG) of $R26^{CreERt2}$ control mice and $Tbx5^{fl/fl};R26^{CreERt2}$ 10 days after receiving tamoxifen. ECGs are representative of n=24 $Tbx5^{fl/fl}$; $R26^{CreERt2}$ and n=10 $R26^{CreERt2}$ mice. (**B** and **C)** Signal-averaged ECG waveforms of ~1000 consecutive beats (B) and Poincaré plots of R-R interval against the subsequent R-R+1 interval (C) in $R26^{CreERt2}$ (n=10) and $Tbx5^{fl/fl};R26^{CreERt2}$ (n=24) mice. P-wave is present in control but absent in $Tbx5^{fl/fl};R26^{CreERt2}$ mice (arrows). Number of mice with absent p-wave and irregularly irregular heart rhythm described in A and B, in $Tbx5^{fl/fl};R26^{CreERt2}$ vs

(Fig 2.1, Continued)

R26^{CreERt2} *P*=7.6E-9 by two-tailed Fisher's exact test. (**D)** Intracardiac atrial electrogram recordings and surface ECG in *R26*^{CreERt2} and *Tbx5*$^{fl/fl}$*;R26*^{CreERt2} mice. A, atrial electrical signal; V, far field ventricular electrical signal. Data are representative of n=6 *Tbx5*$^{fl/fl}$*;R26*^{CreERt2} and n=3 *R26*^{CreERt2} mice. Number of mice with irregular atrial electrogram in *Tbx5*$^{fl/fl}$*;R26*^{CreERt2} vs *R26*^{CreERt2}p=0.012 by two-tailed Fisher's exact test. (**E)** Pulse wave Doppler across mitral valve alongside surface ECG. "A" wave is present in *R26*^{CreERt2} and absent in *Tbx5*$^{fl/fl}$*;R26*^{CreERt2} recordings, indicative of a lack of coordinated atrial contraction. Data are representative of n=6 *Tbx5*$^{fl/fl}$*;R26*^{CreERt2} and n=3 *R26*^{CreERt2} mice. Number of mice with absent "A" wave in *Tbx5*$^{fl/fl}$*;R26*^{CreERt2} vs *R26*^{CreERt2}p=0.012 by two-tailed Fisher's exact test. **(F)** Representative atrial voltage activation maps from *R26*^{CreERt2} and *Tbx5*$^{fl/fl}$*;R26*^{CreERt2} mice 7, 12, and 14 days after completion of tamoxifen treatment. Atrial activation maps demonstrate that conduction waves traverse the atria in ~12 ms after 7 days, ~20 ms after 12 days, and ~40 ms after 14 days post TM. n=2 for each of the 3 groups. Trans-atrial conduction speed in *Tbx5*$^{fl/fl}$*;R26*^{CreERt2} vs *R26*^{CreERt2} p=0.02 by ANOVA. (**G)** Atrial macro reentrant pathways observed in *Tbx5*$^{fl/fl}$*;R26*^{CreERt2} mice 14 days after TM treatment. Reentrant pathway travels right to left within the posterior atrial wall and left to right through the anterior atrial wall. Images are representative of n=2 mice at 14 days.

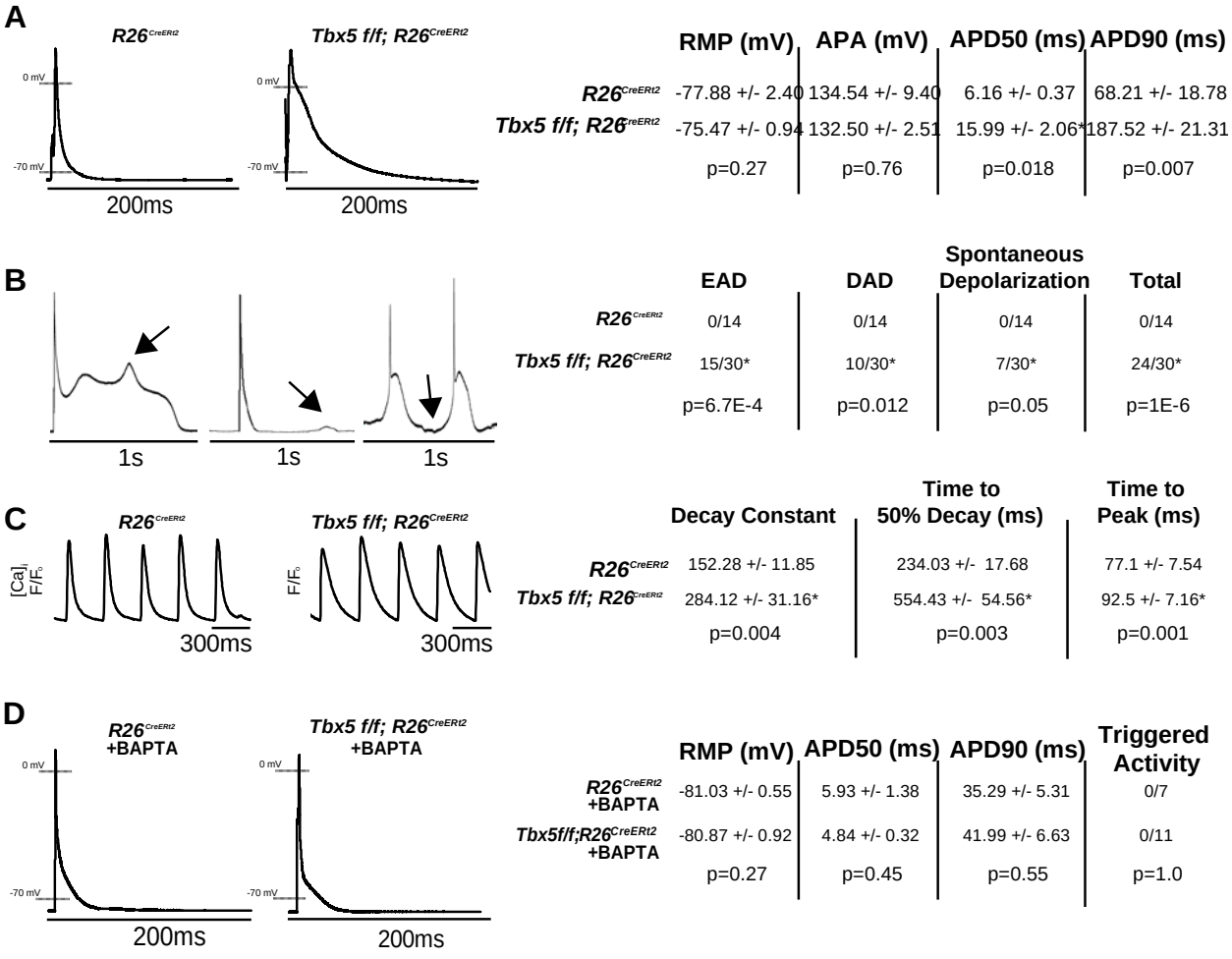

Figure 2.2. Action potential abnormalities in *Tbx5$^{fl/fl}$;R26^{CreERt2}* atrial cardiomyocytes is mediated by disrupted calcium handling. (**A)** Representative action potentials (AP) from atrial cardiomyocytes isolated from *R26^{CreERt2}* and *Tbx5$^{fl/fl}$;R26^{CreERt2}* mice and the corresponding properties of the APs. Resting membrane potential (RMP), AP amplitude (APA), and AP duration at 50% (APD50) and at 90% (APD90) repolarization. Data are means +/- SEM (*Tbx5$^{fl/fl}$;R26^{CreERt2}*, n=30 cardiomyocytes; *R26^{CreERt2}*, n=14 cardiomyocytes; n>5 animals per group). P values determined by two-tailed t-test. (**B)** Representative abnormal depolarization events—early after depolarization (EAD), delayed after depolarization (DAD), and spontaneous phase 4 depolarization—observed in atrial cardiomyocytes. Total number of abnormal

(Fig 2.2, Continued)

spontaneous events were recorded in $R26^{CreERt2}$ (n=14) and $Tbx5^{fl/fl};R26^{CreERt2}$ (n=30) atrial cardiomyocytes from n>5 animals per group. Data are means +/- SEM. P values determined by one-tailed Fisher's exact test. (**C)** Representative tracings of calcium release in isolated cardiomyocytes from $R26^{CreERt2}$ and $Tbx5^{fl/fl};R26^{CreERt2}$ mice. Calcium was imaged with Fluo-4 dye and myocytes were paced at 1 Hz. Properties of calcium transient spikes, including decay constant (tau), time to 50% decay, and time to peak, were recorded across $R26^{CreERt2}$ (n=11) and $Tbx5^{fl/fl};R26^{CreERt2}$ (n=54) cardiomyocytes from n>5 animals in each group. Data are means +/- SEM. P values determined by two-tailed t-test. (**D)** Representative APs of atrial cardiomyocytes isolated from $R26^{CreERt2}$ and $Tbx5^{fl/fl};R26^{CreERt2}$ adult mutant mice treated with calcium chelating agent BAPTA (5 mM). AP properties of these cardiomyocytes were determined, including resting membrane potential (RMP), APD50, and APD90. Data are means +/- SD from $Tbx5^{fl/fl};R26^{CreERt2}$ (n=11) and $R26^{CreERt2}$ (n=7) cardiomyocytes across n>3 animals per group. *P* values determined by either two-tailed t-test for continuous measurements or two tailed Fisher's exact test for count-based measurements (EADs, DADs, Spontaneous Depolarizations, and total abnormal depolarization events).

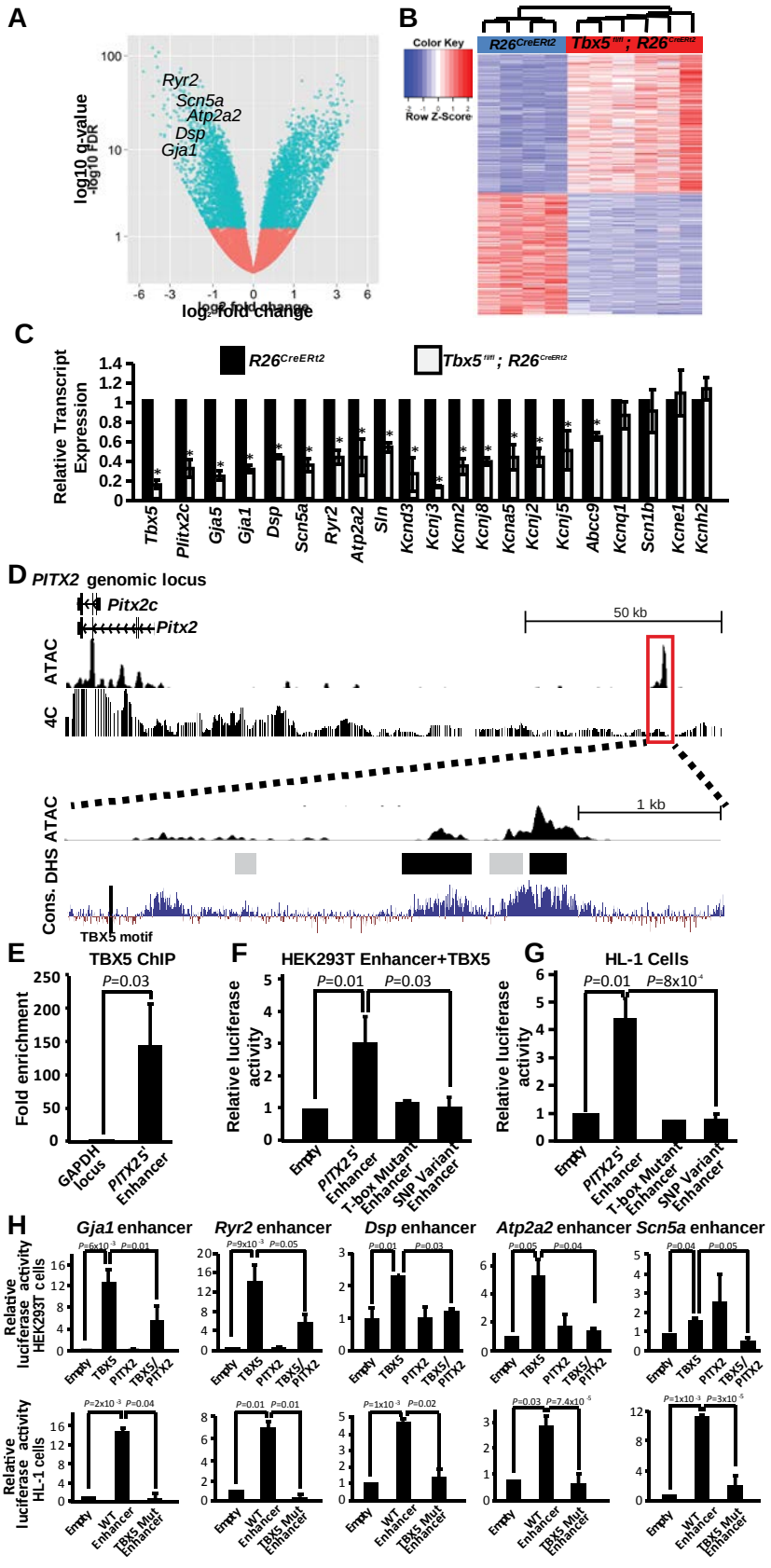

Figure 2.3. TBX5-PITX2 gene regulatory network for atrial rhythm control. (**A**) Volcano

(Fig 2.3, Continued)

plot of relative transcript expression from the left atria of *Tbx5*$^{fl/fl}$*;R26*^{CreERt2} vs *R26*^{CreERt2} mice. All significantly misregulated genes (q<0.05) are labeled blue, all nonsignificant transcripts in red. TBX5 and PITX2 shared targets labeled. (**B**) Heatmap of all significantly misregulated (q<0.05) transcripts in left atria of *Tbx5*$^{fl/fl}$*;R26*^{CreERt2} (n=6) and *R26*^{CreERt2} (n=4) mice. Cladogram shows clustering of biological replicates. (**C**) Relative gene expression by qPCR of known AF ion channels from left atria. Data are means normalized to *GAPDH* and relative to *R26*^{CreERt2} expression. Data are means +/- SEM (n=4 *R26*^{CreERt2} and n=7 *Tbx5*$^{fl/fl}$*;R26*^{CreERt2}). Experiments were performed in technical triplicate. *p<0.05 versus *R26*^{CreERt2} controls by two-tailed t-test. See Table 1 for complete list of genes analyzed. (**D**) *PITX2* genomic locus (hg19) aligned with ATAC-Seq of iPSC-derived cardiomyocytes and 4C with the *PITX2c* promoter as viewpoint. The region cloned for enhancer activity is boxed in red. Below, a smaller scale view of assayed enhancer showing ATAC-Seq, ENCODE DNase hypersensitivity (DHS) (darker to lighter gray showing relative strength of DNAse signal, darker being stronger signal), and Vertebrate Conservation (Cons.). The TBX5-binding motif, which should occur by chance every ~1000 nucleotides, is labeled. (**E**) Fold enrichment after TBX5 ChIP by qPCR from a control locus and *PITX2* enhancer in the human left atrial appendage. Data are means +/-SEM (*n* = 3). (**F** and **G**) *In vitro* luciferase response assay of the candidate PITX2 regulatory element in HEK293T cells co-transfected with TBX5 expression vector or HL-1 atrial cardiomyocytes. Wild type enhancer, full mutant variant enhancer lacking the TBX5 binding-motif mutant enhancer, and SNP rs1906595 variant enhancer are shown. Data are means +/- SEM normalized to empty pGl4.23 vector with TBX5 expression (n=4 WT enhancer, n=4 Empty, n=3 T-box mutant, n=4 SNP variant). (**H)** *In vitro* luciferase response assay of candidate TBX5/PITX2 co-regulated elements

(Fig 2.3, Continued)

at *Gja1*, *Ryr2*, *Dsp*, *Atp2a2*, and *Scn5a* in HEK cells co-transfected with TBX5 or TBX5 and PITX2 or HL-1 cardiomyocytes and corresponding T-box mutant enhancers. Data are means +/- SEM, normalized to blank vector with corresponding overexpression (n=5 for *Atp2a2* enhancer+TBX5 and *Atp2a2* enhancer+TBX5/PITX2; n=3 for all other groups). Experiments in (E to H) were performed in technical triplicate. P values in (E to H) were determined by two-tailed t-test.

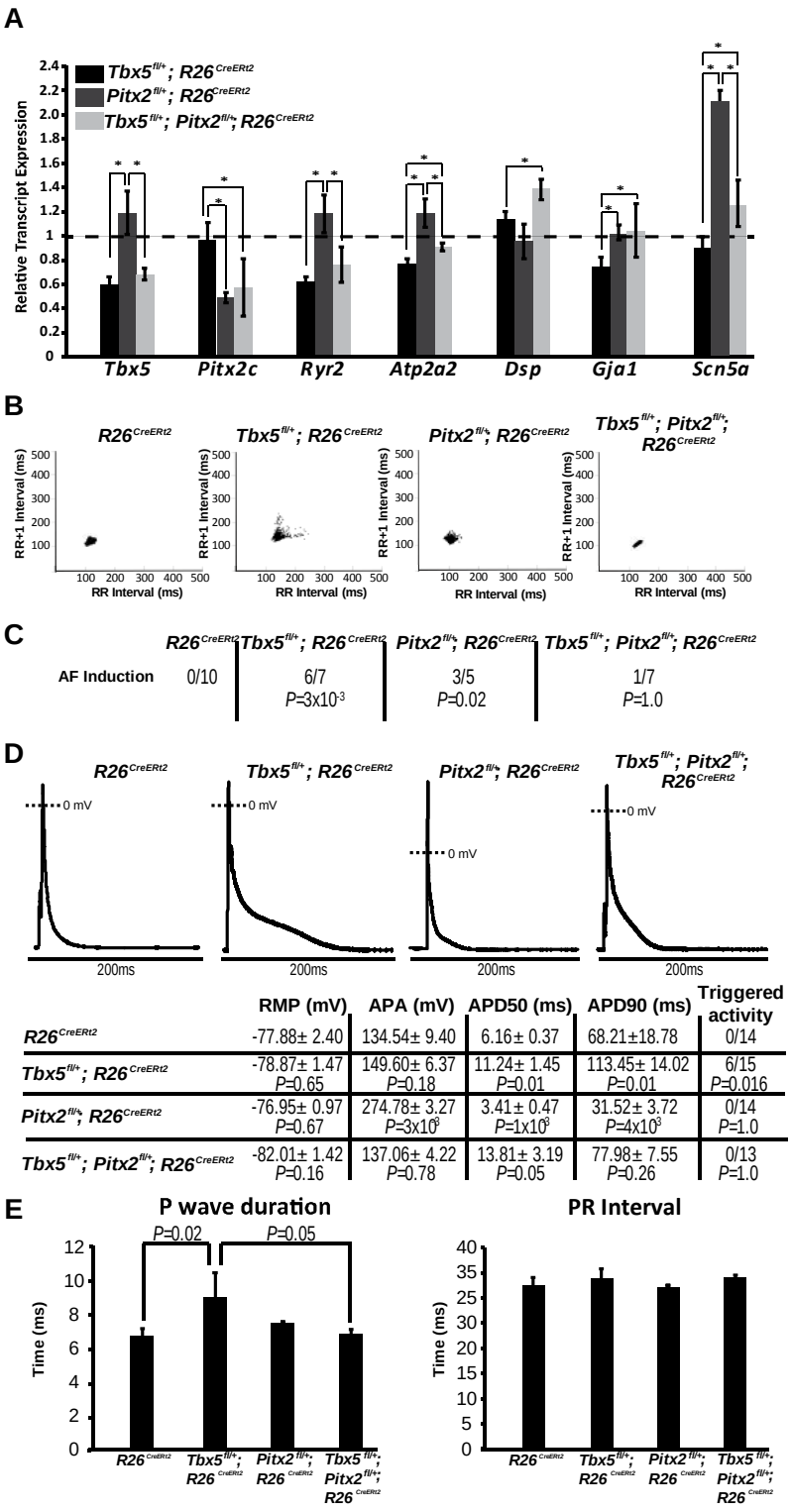

Figure 2.4. ***Pitx2*** **haploinsufficiency rescues** ***Tbx5*** **haploinsufficiency in mice. (A)** Relative

(Fig 2.4, Continued)

transcript expression by qPCR in the left atria from *Pitx2* and *Tbx5* heterozygotes and *Tbx5;Pitx2* compound heterozygotes. Data are means +/- SEM normalized to $R26^{CreERt2}$ mice (set as 1) (n=11 $R26^{CreERt2}$; n=15 $Tbx5^{fl/+};R26^{CreERt2}$; n=4 $Pitx2^{fl/+};R26^{CreERt2}$; n=8 $Tbx5^{fl/+};Pitx2^{fl/+};R26^{CreERt2}$). *P<0.05, two-tailed t-test. Experiments were performed in technical triplicate. **(B)** Representative Poincaré plot of R-R interval vs. the subsequent R-R+1 interval. (n=7 $R26^{CreERt2}$, n=6 $Tbx5^{fl/+};R26^{CreERt2}$, n=9 $Pitx2^{fl/+};R26^{CreERt2}$, n=7 $Tbx5^{fl/+};Pitx2^{fl/+};R26^{CreERt2}$). (**C)** P-wave duration and PR interval calculated from ambulatory telemetry ECG recordings from mice in (B). P-values were determined by two-tailed t-test. Data are means +/- SEM. (**D)** Representative action potential (AP) recordings from atrial myocytes isolated from $R26^{CreERt2}$, $Tbx5^{fl/+};R26^{CreERt2}$, $Pitx2^{fl/+};R26^{CreERt2}$ and $Tbx5^{fl/+};Pitx2^{fl/+};R26^{CreERt2}$ mice. Resting membrane potential (RMP), AP amplitude (APA), and AP duration at 50% (APD50) and 90% (APD90) repolarization were determined from n>3 animals per group ($R26^{CreERt2}$ n=14 cells, $Tbx5^{fl/+};R26^{CreERt2}$ n=15, $Pitx2^{fl/+};R26^{CreERt2}$ n=14; $Tbx5^{fl/+};Pitx2^{fl/+};R26^{CreERt2}$ n=13). P-values were determined versus $R26^{CreERT2}$ controls. Inappropriate depolarization P values were measured by two-tailed Fisher's exact test; APD90 and APD50 were determined by two-tailed t-test. (**E)** Pacing induction by intra-atrial pacing of $R26^{CreERt2}$ (n=10), $Tbx5^{fl/+};R26^{CreERt2}$ (n=7), $Pitx2^{fl/+};R26^{CreERt2}$ (n=5), and $Tbx5^{fl/+};Pitx2^{fl/+};R26^{CreERt2}$ (n=7) mice. P-values determined by two-tailed Fisher's Exact Test.

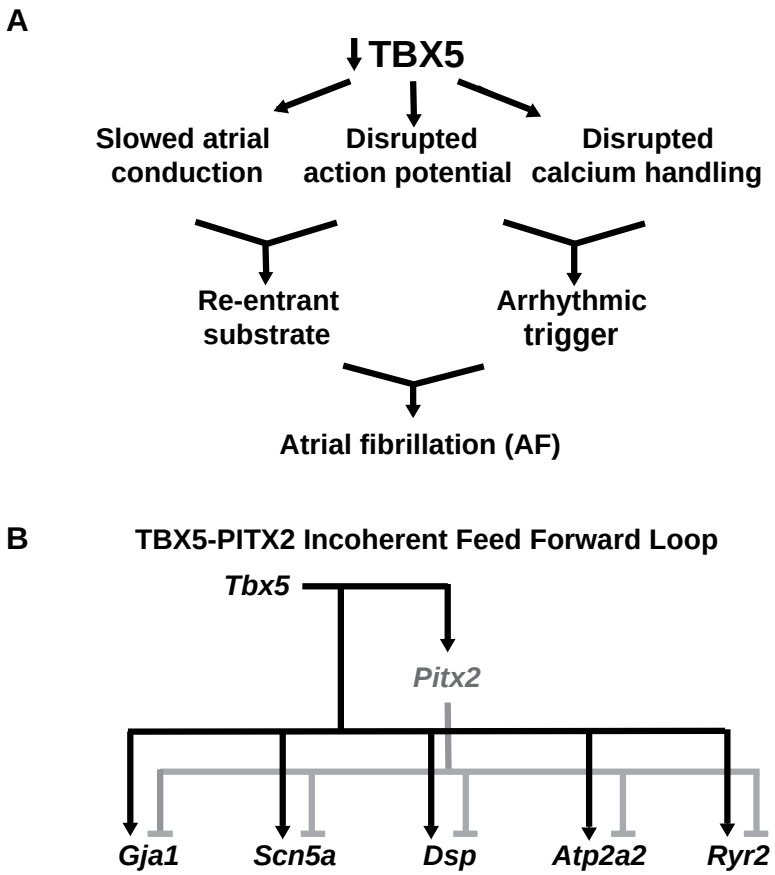

Fig 2.5. A TBX5-PITX2 regulatory loop regulates atrial rhythm. (A) Loss of TBX5 in the adult atrium leads to slowed atrial conduction, prolonged action potential, and disrupted calcium handling, leading to atrial fibrillation trigger and substrate. (**B)** A TBX5-PITX2 incoherent feed-forward loop regulates atrial conduction genes *Scn5a, Gja1, Ryr2, Atp2a2*, and *Dsp*. TBX5 drives PITX2 expression. TBX5 and PITX2 positively and negatively regulate downstream targets. Mysregulation of this loop disrupts atrial conduction.

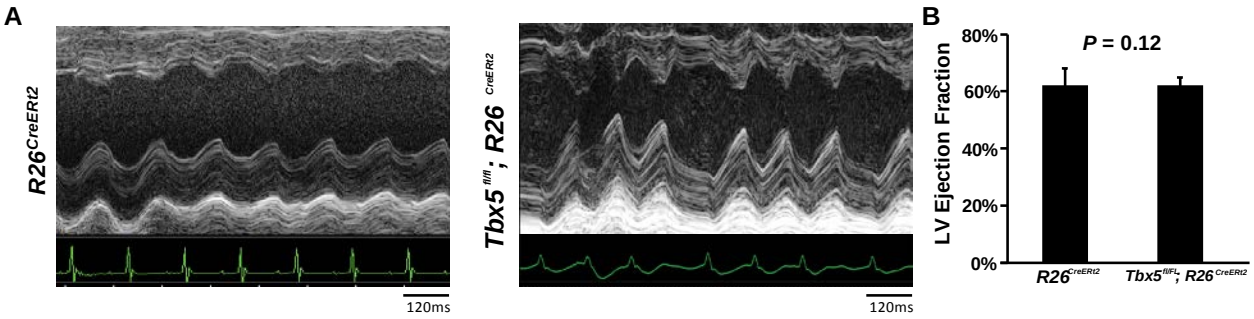

Fig. S2.1. No change in cardiac function after onset of AF in *Tbx5$^{fl/fl}$;R26^{CreERt2}* mice. (**A**) M-mode echocardiography from *R26^{CreERt2}* and *Tbx5$^{fl/fl}$;R26^{CreERt2}* mice shown above surface ECGs. (**B**) Left ventricular ejection fraction (LVEF) calculated from the M-mode ECGs in (A). Data are averages ± SEM (n >5). P value determined by t-test.

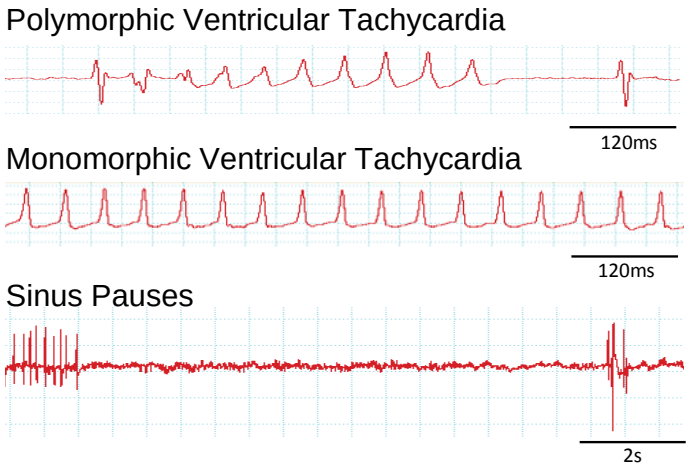

Fig. S2.2. *Tbx5*$^{fl/fl}$*;R26*^{CreERt2} mice develop ventricular arrhythmias following atrial arrhythmias. Sustained ventricular tachycardia, polymorphic ventricular tachycardia resembling Torsades des Pointes, and prolonged sinus pauses (lasting >10 s) served by anesthetized surface ECG mimicking lead II in *Tbx5*$^{fl/fl}$*;R26*^{CreERt2} mice. All ventricular arrhythmias began at 4 weeks post-tamoxifen, after the presentation of AF in each individual case.

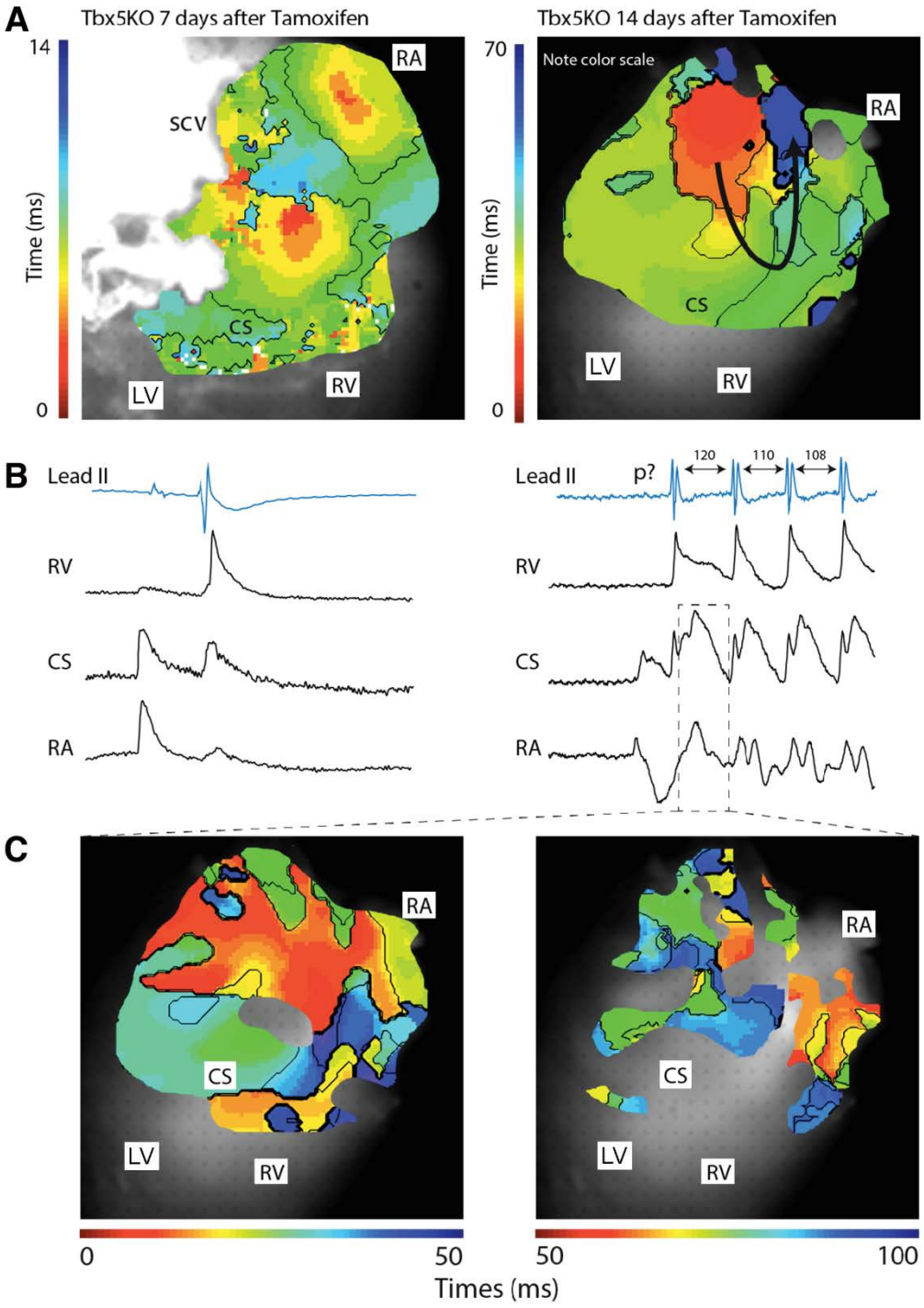

Fig. S2.3. ***Tbx5*$^{fl/fl}$*;R26*^{CreERt2} **show paroxysmal atrial fibrillation.** (**A**) Reconstructed atrial activation patterns at 7 and 14 days after tamoxifen injection in *Tbx5* mutant mice. **(B)** ECGs

(Fig S2.3, Continued)

tracing with simultaneously recorded optical action potentials used for the activation map of the respective activation patterns above in (A**)**. (**C**) Activation maps show the reentrant activation pattern in the period marked above in (B). CS, coronary sinus; LV, left ventricle; RA, right atrium; RV, right ventricle; SCV, superior caval vein.

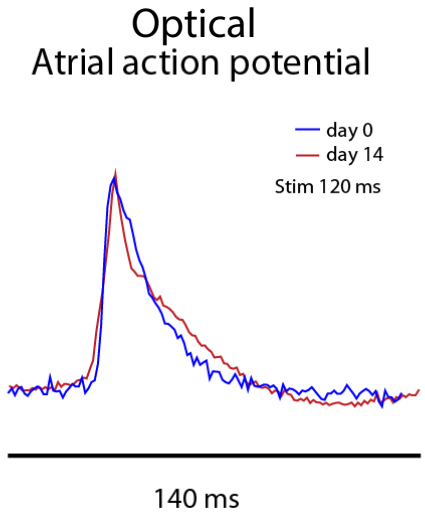

Fig. S2.4. Optical action potential from right atrium of *Tbx5*$^{fl/fl}$;*R26*^{CreERt2} *mice.* Optically recorded atrial action potentials at days 0 and 14 after tamoxifen injection. Data are representative of n = 2 mice from each group.

Fig. S5. Atrial rhythm instability and AF inducibility in *Tbx5*$^{fl/+}$;*R26*^{CreERt2} mice is rescued by *Pitx2* haploinsufficiency. A) and **B)** Representative surface ECG (A) and representative pacing induction of atrial fibrillation (B) from *R26*^{CreERt2} controls, *Tbx5* and *Pitx2* heterozygotes, and *Tbx5;Pitx2* compound heterozygote mice.

Gene	$Tbx5^{fl/fl};R26^{CreERt2}$ vs $R26^{CreERt2}$ fold change by RNASeq	q-value RNASeq	$Tbx5^{f/f};R26^{CreERt2}$ vs $R26^{CreERt2}$ fold change by qRT-PCR	p-value qRT-PCR
Pitx2c	0.53	0.07	0.22±0.09	0.03
Gja5	0.05	$3x10^{-4}$	0.23±0.05	$2.49x10^{-6}$
Gja1	0.24	$3x10^{-4}$	0.30±0.04	$1.06x10^{-5}$
Dsp	0.36	$3x10^{-4}$	0.44±0.05	$1.99x10^{-8}$
Scn5a	0.30	$3x10^{-4}$	0.34±0.07	1.06E-5
Ryr2	0.22	$3x10^{-4}$	0.42±0.07	$2.14x10^{-6}$
Atp2a2	0.42	$3x10^{-4}$	0.42±0.17	0.002
Sln	0.41	$3x10^{-4}$	0.52±0.05	$1.2x10^{-4}$
Kcnd3	0.67	$7x10^{-3}$	0.25±0.17	0.002
Kcnj3	0.14	$3x10^{-4}$	0.12±0.01	$9.56x10^{-12}$
Kcnn2	0.77	0.10	0.33±0.08	$4.83x10^{-5}$
Kcnj8	0.67	$7x10^{-3}$	0.38±0.04	$3x10^{-6}$
Kcna5	0.78	0.04	0.42±0.13	$4x10^{-4}$
Kcnj2	0.60	$3x10^{-4}$	0.42±0.09	$1.4x10^{-4}$
Kcnj5	0.21	$3x10^{-4}$	0.49±0.2	$4.7x10^{-5}$
Abcc9	0.58	$3x10^{-4}$	0.63±0.04	$1.7x10^{-5}$
Cacna1c	0.95	0.88		
Camk2a/2b/2d/2g	0.24/0.83/0.59/1.25	0.11/1/0.2/1		
Slc8a1/a2/a3	0.81/3.08/0.61	0.18/1/1		
Calm2/3	1.27/1.44	$0.03/4x10^{-4}$		

Table 2.1. Gene expression changes in $Tbx5^{fl/fl};R26^{CreERt2}$ mouse left atria. Gene expression of ion channels implicated in AF and of *Tbx5/Pitx2* co-targets in left atrial tissue from $Tbx5^{f/f};R26^{CreERt2}$ adult mutant mice normalized to $R26^{CreERt2}$ by RNASeq or qRT-PCR, in independent cohorts. Data are means +/- SEM (n=4 $R26^{CreERt2}$ and n=6 $Tbx5^{f/f};R26^{CreERt2}$ for RNASeq; n=4 $R26^{CreERt2}$ and n=7 $Tbx5^{fl/fl};R26^{CreERt2}$ for qRT-PCR). Nonsignificant changes in the RNASeq were not validated by realtime, with the exception of *Pitx2c*, which was just above significance by RNASeq.

	$Tbx5^{f/+};R26^{CreERt2}$ fold change	$Pitx2^{f/+};R26^{CreERt2}$ fold change	$Tbx5^{f/+};Pitx2^{fl/+};R26^{CreERt2}$ fold change
Tbx5	0.55±0.06 *	1.11±0.17 #	0.64±0.05 *
Pitx2c	0.91±0.13	0.45±0.10 * #	0.54±0.22 *
Ryr2	0.59±0.04 *	1.12±0.15 #	0.72±0.14 *
Atp2a2	0.73±0.05 *	1.10±0.11 #	0.86±0.08 #
Dsp	1.07±0.07	0.88±0.14	1.31±0.21
Gja1	0.70±0.04 *	0.95±0.06 #	0.98±0.03 #
Scn5a	0.85±0.08 *	1.95±0.10 * #	1.18±0.23 #

Table 2.2. Gene expression changes in $Tbx5^{fl/+};R26^{CreERt2}$, $Pitx2^{fl/+};R26^{CreERt2}$, and $Tbx5^{fl/+};Pitx2^{fl/+};R26^{CreERt2}$ mouse left atria. Gene expression TBX5/PITX2 shared targets in left atrial tissue normalized to $R26^{CreERt2}$ by qRT-PCR, in independent cohorts. Data are means +/- SEM (n=4 $R26^{CreERt2}$ and n=6 $Tbx5^{f/f};R26^{CreERt2}$ for RNASeq; n=4 $R26^{CreERt2}$ and n=7 $Tbx5^{fl/fl};R26^{CreERt2}$ for qRT-PCR). * p<0.05 vs $R26^{CreERt2}$, # p<0.05 vs $Tbx5^{fl/+};R26^{CreERt2}$. n=11 $R26^{CreERt2}$; n=15 $Tbx5^{fl/+};R26^{CreERt2}$; n=4 $Pitx2^{fl/+};R26^{CreERt2}$; n=8 $Tbx5^{fl/+};Pitx2^{fl/+};R26^{CreERt2}$. For $Tbx5^{fl/+};R26^{CreERt2}$ vs $R26^{CreERt2}$ *Tbx5* p=1.5E-5; *Pitx2c* p=0.86 *Scn5a* p=0.028 *Ryr2* p=3.98E-5 *Gja1* p=6.98E-3 and *Atp2a2* p=5.5E-3. For $Pitx2^{fl/+};R26^{CreERt2}$ vs $R26^{CreERt2}$ *Pitx2c* p=0.05 *Scn5a* p=0.03 *Tbx5 p*=0.47 *Ryr2* p=0.44 *Gja1* p=0.52 *Atp2a2* p=0.60. For $Tbx5^{fl/+};Pitx2^{fl/+};R26^{CreERt2}$ vs $Tbx5^{fl/+};R26^{CreERt2}$ *Tbx5*: p=0.56; *Scn5a*: p=7.5E-3; *Ryr2*: p=0.01; *Gja1*: p=0.02; *Atp2a2*: p=0.19; *Dsp*: p=0.05. $Tbx5^{fl/+};Pitx2^{fl/+};R26^{CreERt2}$ vs $R26^{CreERt2}$ *Scn5a*: p=0.58; *Ryr2*: p=0.07; *Gja1*: p=0.99; *Atp2a2*: p=0.19; *Dsp*: p=0.31. By two tailed t-test. Experiments performed in technical triplicate.

CHAPTER III: DIFFERENTIAL ENHANCER TRANSCRIPTION DEFINES A TRANSCRIPTION FACTOR-DEPENDENT GENE REGULATORY NETWORK

Abstract

Transcription factors (TFs) determine context-dependent gene expression by binding and modulating enhancers. Defining functional TF-dependent enhancers from the thousands of TF binding locations genome-wide remains a fundamental challenge for understanding TF-dependent gene regulatory networks. Based on evidence that non-coding RNA (ncRNA) is transcribed from enhancers, we hypothesized that TF-dependent ncRNA transcriptional profiling would identify functional TF-dependent enhancers. We defined TF-dependent ncRNAs for TBX5, a critical cardiac TF, by deep sequencing ncRNAs from wildtype versus *Tbx5* mutant mouse atrium. Genome-wide, *Tbx5*-dependent ncRNAs were enriched for chromatin accessibility and tissue-specific marks of active enhancers. *Tbx5*-dependent ncRNA-defined enhancers were enriched for TBX5 binding and demonstrated robust *Tbx5*-dependent activity *in-vitro*. The direction and magnitude of *Tbx5*-dependent enhancer transcription correlated positively with that of *Tbx5*-dependent target gene expression, providing a quantitative metric for *Tbx5*-dependent enhancer function. *Tbx5*-dependent atrial ncRNAs identified enhancers required for expression of a calcium-handling network previously associated with *Tbx5* function in the atrium, elucidating a physiologically relevant gene regulatory network. Application of TF-dependent enhancer transcription may allow broad elucidation of TF-dependent gene regulatory networks.

Introduction

Identification of tissue specific enhancer elements using histone modifications, chromatin status [1-3] and TF occupancy has been successful in many contexts [4]. However, only a minuscule percentage of the thousands of the candidate TF-dependent enhancers identified by these approaches have been functionally validated [5] and these approaches lack specificity and quantitative resolution of enhancer strength [6]. For example, we attempted to identify TF-dependent enhancers for TBX5, a canonical cardiac transcription factor, applying common tissue-specific markers for enhancers to cardiomyocytes [7,8]. Overlap of genomic enhancer marks in the adult mouse heart, including Histone 3 Lysine 27 Acetylation (H3K27Ac) [9], DNase I hypersensitivity [10], and ATAC-seq in HL-1 atrial cardiomyocytes revealed high concordance with TBX5 occupancy (**Fig 3.1A**). The overlap between TF localization and canonical enhancer markers identified 7,500 common regions suggesting a lack of specificity for functional TF-dependent enhancers, consistent with published observations that TF occupancy by ChIP is non-specific [11,12]. These observations indicated the need for alternate strategies for the identification of functional TF-dependent enhancers.

Recent evidence suggests that enhancer transcription is a widespread property of functional enhancers (*8-16*). Roles for enhancer-associated noncoding RNAs (ncRNAs) have been reported in transcriptional regulation, development, and disease (*8-16*). We hypothesized that TF-dependent ncRNA enhancer transcripts may define a TF-dependent gene regulatory network. We have previously published a TBX5-driven gene regulatory network that modulates calcium handling to maintain normal cardiac rhythm in the atrium [7].

Results

To identify TF-dependent noncoding RNA (ncRNAs) in this context, we sequenced non-polyadenylated RNA from the atrium of control (*R26*^{CreERt2}; n=3) and *Tbx5* mutant (*Tbx5*$^{fl/fl}$*;R26*^{CreERt2}; n=4) adult mice, as enhancer associated RNAs are canonically under poly-adenylated [13]. Approximately 40.5k noncoding transcripts were identified by *de novo* transcript assembly. Of these, 3610 noncoding transcripts were *Tbx5*-dependent intergenic ncRNAs (FDR<0.05, FC>1.5, at least 1kbp away from known coding gene TSSs, ENCODE mm10) [14]. *Tbx5*-dependent ncRNA expression was robust across biological replicates, distinguishing *Tbx5*$^{fl/fl}$*;R26*^{CreERt2} samples from *R26*^{CreERt2} controls (**Fig 3.1B**). 3,067 (85%) *Tbx5*-dependent ncRNA transcripts were located within 2Mbp up-stream or down-stream of 2,383 *Tbx5*-dependent protein-coding genes defined by RNA-seq of polyA transcripts from the same samples, within known distance parameters of enhancer-gene target pairs (**Fig 3.1C)** [7].

We analyzed the genomic features of *Tbx5*-dependent ncRNAs (**Fig 3.S1A**). The vast majority of the *Tbx5*-dependent ncRNAs were *de novo* transcripts (>85%), with a small number previously annotated as known lncRNAs (5%), antisense transcripts (2%), or other predicted noncoding transcripts, consistent with the observation that ncRNAs are highly tissue-specific and that global ncRNA discovery in the atrium of the heart has not been previously reported [15]. *Tbx5*-dependent ncRNAs showed a bimodal length-distribution, with relative enrichments at ~500bp and at ~2000bp (**Fig 3.S1B**), similar to GENCODE annotated lncRNAs (**Fig 3.S1B1**). *Tbx5*-dependent ncRNAs showed no

preference for uni- vs. bi-directional transcription (**Fig 3.S1C**). Sequence conservation indicated average Phastcon30 scores between 0 and 0.2 for the majority of *Tbx5*-dependent ncRNAs, with a distribution similar to known lncRNAs and lower than promoters (**Fig 3.S1D**).

We hypothesized that *Tbx5*-dependent ncRNAs marked *Tbx5*-dependent enhancers. *Tbx5*-dependent ncRNAs were generated from locations enriched for enhancer marks genome-wide (**Figs 1E, S2**). Specifically, we observed enrichment between the 3067 *Tbx5*-dependent ncRNAs and DNAse hypersensitivity in the adult mouse heart, open chromatin (ATAC-sequencing) in HL1 cardiomyocytes, and H3K27Ac in the adult mouse heart [9,16,17]. Furthermore, *Tbx5*-dependent ncRNAs defined locations enriched genome-wide for TBX5 occupancy in HL-1 cells (**Fig 3.2A**). We further hypothesized that *Tbx5*-dependent ncRNAs marked *Tbx5*-dependent enhancers at regions of open chromatin, a global feature of enhancers [18]. We overlapped differential ncRNA expression with cardiac DNAse Hypersensitivity [19] and defined chromatin accessible *Tbx5*-dependent noncoding transcripts as those within a 1kbp window of from DNAse peak-centers. 1703 candidate regulatory elements were identified by this intersection, from a set of 28,393 TBX5-bound DNAse regions (**Fig 3.2B**).

We functionally interrogated candidate regulatory elements identified by *Tbx5*-dependent ncRNA transcription for enhancer activity. 11 of 12 candidate enhancers demonstrated robust *cis*-activity in HL-1 cardiomyocytes, indicative of cardiac enhancer function (P-values from 3.3e-6 to 0.03, **Fig 3.2C**). Overall, these enhancers demonstrated unusually strong *cis*-activation, with 9 activating more than 10-fold and 3 activating more than 50-fold. These enhancers were strongly TBX5-dependent activity, as the activity of

10 of the 11 enhancers was significantly diminished by mutation of their canonical T-box sites.

We asked if differential ncRNA expression could discern functional TF-dependent enhancers from candidates marked by TF-ChIP. TBX5 ChIP marked five candidate enhancers at the *Ryr2* gene, encoding the cardiac ryanodine receptor, a TBX5-dependent gene critical for cardiac function [7] (**Fig 3.2D, E**). The single candidate enhancer marked by *Tbx5*-dependent noncoding transcription showed strong cardiac activity by luciferase assay (**Fig 3.2E**, p=0.05). In contrast, the other candidates showed weak or no cardiac activity (**Fig 3.2E**, p=0.06, 0.39, 0.02, 0.002 for Enhancers 1-3, 5). This observation suggested that TF-dependent ncRNAs may improve the predictive performance of TF occupancy-based enhancer identification. We tested the predictive power of TF-dependent ncRNA expression to accurately identify tissue-specific enhancers genome-wide, comparing the precision and specificity of *Tbx5*-dependent ncRNA-defined enhancers against 240 cardiac and 1162 non-cardiac enhancers tested by mouse transient transgenic reporter assays *in-vivo*, from the VISTA database [5]. *Tbx5*-dependent transcription increased the predictive power of DNAse hypersensitivity (DHS), H3K27Acetylation, and open chromatin by ATAC-seq (>2-fold for precision, with similar specificity, **Fig 3.2F**), suggesting that *Tbx5*-dependent ncRNA expression is an improved metric for functional enhancer identification.

To further define atrial-specific TBX5-dependent enhancers, we overlaped *Tbx5*-dependent ncRNAs with open chromatin in the mouse atrium, defined by ATAC-Seq. From 16,000 open regions, 145 local *Tbx5*-dependent ncRNAs marked 152 candidate *Tbx5*-dependent enhancers. We compared TF motif enrichment between enhancers

defined by *Tbx5*-dependent ncRNA versus all atrial open chromatin regions. Atrial open chromatin showed enrichment for motifs of general TF families, such as MYC and E2F that provide non-tissue-specific transcriptional functions (**Fig 3.2G**). By contrast, *Tbx5*-activated enhancers showed significant enrichment for the T-box motif itself (enrichment score=6.7, P =0.03). TBX5-occupancy was concordant with open chromatin in HL-1s (**Fig 3.1A, 2H**). TBX5 occupancy by ChIP-seq in atrial HL-1 cells was highly enriched at open atrial locations with *Tbx5*-dependent ncRNAs (72 of the 152 locations, OR=18, P<2e-16; **Fig 3.2H**). The candidate enhancers marked by differential transcription but not TBX5 binding represent either indirect TBX5-dependent elements or false negatives by ChIP. Although a candidate enhancer located at the *Sln* locus did not demonstrate TBX5 localization by ChIP-seq in HL-1 cells [16], we analyzed the region and observed TBX5-occupancy by ChIP-qPCR (p=0.03 vs control *Gapdh* locus) (**Fig 3.S5**). The element demonstrated robust TBX5-dependent activity in HL-1 cardiomyocytes *in-vitro* by luciferase assay (P=2.3E-4 WT vs T-box Mutant). These observations provide evidence that *Tbx5*-dependent enhancer transcription identified locations of TBX5 binding, *a-priori*.

We sought to define the gene regulatory networks controlled by *Tbx5*-dependent CRE transcription and examined enhancer-target gene interactions. We postulated that there may be a quantitative relationship between the direction and magnitude of TF-dependent differential enhancer transcription and the direction and magnitude of differential transcription of the enhancer's target gene(s). The direction of altered transcription of *Tbx5*-dependent ncRNAs and local (within 2Mbp) polyA transcripts was often coherent, consistent with formation of regulatory units. Specifically, 86% of down-

regulated were local to down-regulated *Tbx5*-dependent coding-genes (1577/1841; p<0.001) and 66% of up-regulated *Tbx5*-dependent ncRNAs were local to up-regulated *Tbx5*-dependent coding-genes (1490/1769 *p*=0.003) from *Tbx5* mutant atria (**Figs 1D, S3A**, **S3B**). The significant directional coherence between *Tbx5*-dependent ncRNAs and local *Tbx5*-dependent coding transcripts was consistent with a directed relationship between TF-dependent transcription and TF-dependent enhancer function. Targeted chromosome conformation capture analysis (4C-Seq) [20] in the atrial HL-1 cardiomyocyte cell line demonstrated significant interactions between enhancer defined by *Tbx5*-dependent transcription and surrounding *Tbx5*-dependent genes located within each topological associated domain in each case (at genetic locations *Ryr2*, *Atp2a2*, *Ubc*, and *Wrn*; **Fig 3.S4**). Application of the Pi-score computational method [21] suggested the potential for local action of enhancers defined by *Tbx5*-dependent transcription: 81 out of 93 enhancers with *Tbx5*-activated transcription showed quantitative directional coherence with the expression of the nearest *Tbx5*-dependent gene (**Fig 3.3A**), consistent with *cis*-regulation of the nearest *Tbx5*-dependent target gene by enhancers. We analyzed the fold-changes of *Tbx5*-dependent enhancer transcription with that of adjacent *Tbx5*-dependent genes and observed a significant correlation between their direction and magnitude (coefficient=0.58, P<2e-16 by Pearson Correlation test; **Fig 3.3B**). This positive correlation suggested that TF-dependent enhancer transcription represents a quantitative metric of TF-dependent enhancer activity.

Enhancer-transcribed ncRNAs may function to stabilize enhancer-promoter contacts by association with chromatin [22]. We therefore determined the sub-nuclear localization of six *Tbx5*-dependent enhancer ncRNAs by nuclear fractionation in HL-1

cardiomyocytes into chromatin vs soluble nuclear fractions [3]. We found that each *Tbx5*-dependent ncRNA was significantly enriched in the chromatin vs the soluble nuclear fraction, to a degree similar to that for *Xist*, a canonical chromatin-binding ncRNA (qPCR P<6e-3, n=3) (**Fig 3.3C**). By contrast, control RNA from the *Gapdh* coding region showed no chromatin enrichment. Thus, *Tbx5*-dependent CRE ncRNAs are tightly localized to chromatin suggesting a possible direct role in target gene regulation [23,24].

To reveal functional gene sets within the identified *cis*-regulatory network we performed Gene Ontology (GO) enrichment analysis [25] on the putative targets of enhancers with *Tbx5*-activated transcription (**Fig 3.3D**). We found enrichment for two statistically significant terms, "Calcium ion-related biological process" and "Calcium ion transmembrane transport," remarkably consistent with previous work demonstrating that *Tbx5* controls atrial rhythm by driving calcium handling physiology and gene expression [7]. We directly assessed the function of the *Tbx5*-dependently transcribed enhancer at the genes encoding the cardiac Ryanodine Receptor (*Ryr2)* and the cardiac sarcoplasmic reticulum (SR) Calcium ATPase *Atp2a2* (**Figs 1D, E**), responsible for calcium-induced calcium release and reuptake respectively during the cardiac action potential [26].

We examined the functional requirement of the *Tbx5*-dependent enhancer-associated ncRNA at the *Atp2a2* or *Ryr2* loci. To test the function the ncRNA for gene expression and calcium cycling, we compared antisense mediated knockdown of the enhancer ncRNA versus the mRNA in HL-1 cardiomyocytes [24,27]. Control knockdown of the mRNA caused a 50% and 40% decrement in *Ryr2* and *Atp2a2* gene expression

respectively. Enhancer ncRNA expression was unchanged (P=5.3E-5, 5.1E-4 for mRNA; P=0.43, and 0.82 for enhancer ncRNAs, respectively; **Fig 3.3E,F**). In contrast, knockdown of the *Ryr2* and *Atp2a2* enhancer ncRNAs (chr13:12226210-12226916, and chr5:122969667-122970801, mm9) caused a decrease in enhancer ncRNA expression and, importantly, a corresponding decrease in *Ryr2* or *Atp2a2* mRNA expression (P=5.64E-6, 0.04, 1.62E-6, 4.1E-4 respectively; **Fig 3.3E,F**). We assessed the effect of enhancer-RNA knockdown on RNA Polymerase 2 (Pol2) occupancy at the cognate promoter. Knockdown of the *Ryr2* enhancer ncRNA caused reduced Pol2 occupancy at the *Ryr2* promoter (Pol2 ChIP enriched vs ctrl locus 8.7±0.79-fold vs 1.6±0.20-fold following knockdown, P=0.02, **Fig 3.3G**). This observation suggested that the *Ryr2* enhancer ncRNA is required for recruitment or stability of Pol2 binding at the *Ryr2* promoter.

Ryr2 enhancer ncRNA-knockdown caused altered transmembrane calcium kinetics. In control HL-1 cells, β-adrenergic stimulus (1 nM isoproterenol, ISO) caused a dramatic increase in the rise rate of cytosolic calcium transients (rise rate change from 33.3±7.4 to 84.8±9.9, P<0.05). Knockdown of the *Ryr2* enhancer transcript blocked the effect of ISO (rise rate change 49.3±11.3 to 41.0±6.1; P=NS) (**Fig 3.3E, F**). Similarly, *Ryr2*-mRNA-knockdown also blunted the effect of β-adrenergic stimulus (rise rate change from 34. 8±5.9 to 33.3±7.4 P=NS) (**Fig 3.3E,F**). We conclude that the TF-dependent ncRNA at *Ryr2* is necessary for *Ryr2* gene expression and normal *Ryr2*-dependent calcium release, a cellular phenotype dependent on the *Tbx5*-driven gene regulatory network [7].

Discussion

Defining TF-dependent *cis*-regulatory networks in health and disease remains a fundamental challenge. By analyzing TF-dependent ncRNAs, we defined TF-dependent enhancers in a sensitive, quantitative manner. TF-dependent enhancer transcription identified enhancers that were highly enriched for the T-box binding element and for TBX5-localization by ChIP-sequencing. This approach provides an *a-priori* method for identifying highly active TF-bound enhancers without ChIP, therefore bypassing the common limitations of ChIP-based approaches [28]. TF-dependent enhancer-associated transcripts were highly enriched in the chromatin fraction of the nucleus and essential for target gene expression. These enhancers defined a *cis*-regulatory network for a specific physiologic pathway previously associated with TF function. Dependence of TF-dependent target gene transcription on TF-dependent enhancer transcription may be a fundamental aspect of TF-driven *cis*-regulatory networks.

Methods

ATAC-seq library preparation and sequencing: We cultured HL-1 cardiomyocytes according to published protocols [32]. ATAC (assay for transposase-accessible chromatin) protocol was performed as previously described[33]. Briefly, nuclei were isolated according to published protocols, transponsed, and libraries were amplified and normalized with the Illumina Nextera DNA Library prep kit (FC-121-1031) according to the manufacturer's protocols and Greenleaf et al[33]. Libraries were quantitated by the Agilent Bio-analyzer, pooled in equimolar amounts, and sequenced with 50 bp single-end reads on the HiSeq2500 following the manufacturer's protocols.

Data analysis: The 50 bp reads were examined by fastqc. About 30-70 million reads were sequenced for each sample with high quality reads (quality score >30). All reads for each sample were aligned to the UCSC mouse genome mm9 with Bowtie (v1.0.0) using default settings except (-v 1 -q -m 1). Duplicates , low quality (Q<30), and chrM reads were removed, index were added by samtools (v 0.1.18). Between 12 million and 26 million high-quality reads per sample that mapped to genomic DNA were retained. Peaks were called for each sample using MACS2 [34] with parameters "--nomodel --bw 200 --bdg". Overlap peaks identified from three replicates were generated for the downstream analysis. Peaks mapping to ENCODE blacklist regions were excluded from the identification. Peak annotation was performed using the Bioconductor package seq2pathway [35]. All sequencing tracks were viewed using the Bioconductor package Gviz [36].

Differentially expressed coding genes identified from RNA-seq analysis

Mouse model: Mice were raised according to university protocols (protocol # 71737). A tamoxifen inducible Cre recombinase driven by the ubiquitous R26 promoter was used to excise exon 3 of TBX5 as previously described [37]. Excision of *Tbx5* was induced by IP tamoxifen for 3 days, 0.167mg tamoxifen/day. Mice were sacrificed at 1 week post tamoxifen and right and left atrial appendage were dissected.

RNA extraction: Left atrium of these 6-8 week old mice were mechanically homogenized in TRIzol Reagent (Invitrogen), chloroform was added, aqueous phase was isolated, ethanol was added, and then RNA was isolated using RNeasy Mini Column (Qiagen).

RNA-Seq library preparation: Libraries were prepared from this RNA starting with 1g per sample and using the mRNA-seq Sample Prep Kit (Illumina) as per recommended instructions. After Ribozero purification and removing only ribosomal RNA, barcoded libraries were prepared according to Illumina's instructions (2013) accompanying the TruSeq RNA Sample prep kit v2 (Part# RS-122-2001). Libraries were quantitated using the Agilent Bio-analyzer (model 2100) and pooled in equimolar amounts. The pooled libraries were sequenced with stranded 50bp single-end reads on the HiSeq2500 in Rapid Run Mode following the manufacturer's protocols (2013).

Data analysis: The left 3 bp sequencing reads were trimmed due to their unexpected lowest quality score among all 50bp reads. About 62-72 million reads were sequenced for each sample with high quality reads (quality score >30). Fastq files were aligned to build GRCm38.p3 of the mouse genome provided by the GENCODE Consortium (mm10).

Transcript alignment was performed using TopHat (version 2.0.10, using the parameters --segment-length 24 --segment-mismatches 2 --no-coverage-search). Reads mapping to mitochondrial genome were excluded. Between 43 million and 48 million successfully mapped reads were analyzed for differential expression using the R package DESeq [38]. Significance was considered when FDR<0.05, fold change >1.5. To ensure cardiac expression we additionally require a normalized mean value in the wide type group >1.

Differentially expressed non-coding RNAs identified from RNA-seq analysis

RNA-Seq library preparation: Total RNA was extracted by TRIzol Reagent (Invitrogen), followed by ribosomal and polyA depletion. After RiboZero purification and oligo-dT depletion, RNA Barcoded Libraries were prepared according to Illumina's instructions (2013) accompanying the TruSeQ RNA Sample prep kit v2 (Part# RS-122-2001). Libraries were quantitated using the Agilent Bio-analyzer (model 2100) and pooled in equimolar amounts. The pooled libraries were sequenced with 50bp stranded single-end reads on the HiSEQ4000 in Rapid Run Mode following the manufacturer's protocols (2013).

Data analysis: The left 3 bp fastq sequencing reads were trimmed due to their unexpected lowest quality score among all 50bp reads. About 92-135 million reads were sequenced for each sample with high quality reads (quality score >30). Fastq files were aligned to build GRCm38.p3 of the mouse genome provided by the GENCODE Consortium (mm10). Transcript alignment was performed using TopHat (version 2.0.10) as previously described [39] and between 89 million and 129 million reads were

successfully mapped. De novo assembly was performed by Cufflinks (version 2.2.1), as it can recover transcripts that are transcribed from segments of the genome that are missing from the current genome assembly. Featurecount was performed using Rsubread. Analysis of differential expression was performed using a count-based differential expression analysis strategy for RNA sequencing and normalized using the R package DESeq [38]. False discovery rate (FDR) was calculated after removing the coding gene counts. Significance was considered when FDR<0.05, fold change >1.5.

Associating ncRNAs and regulatory elements

We observed around half of the identified TF-dependent ncRNAs were contiguous to rather than directly overlapped with open chromatin regions called. To capture these ncRNAs, we expanded the definition of 'co-localization' of ncRNAs and enhancers that reside within 1kbp between their centers of regions. Compared to edge-to-edge distance control, this strategy counts the cases of two adjacent short regions but excluded two adjacent long regions. Since longer regions have more chance to directly overlap with each other, this strategy adjusted the bias introduced by width of peaks.

Simulation study on ncRNAs local to coding-genes that were TF-dependent

The null hypothesis is that TF-dependent ncRNAs and genes are randomly localized across the whole genome. To test this hypothesis, we ran n =1000 simulations. In each simulation run i, we randomly sampled x ncRNAs from the 40.5K noncoding transcript

background and y genes from the mouse genome, termed set {X} and set {Y}, respectively. We then counted the number of ncRNAs in the set {X} within the 2Mb window of any genes with in the set of {Y}, and recorded it as a_i. A is the observed count of TF-dependent ncRNAs within the 2Mbp window of TF-dependent genes. An empirical p-value was estimated for TF-activated and TF-repressed ncRNAs, respectively (**Equ 1**), let the expression marked by |.| be the total number of counts meeting a given criteria.

$$\boldsymbol{P\ value} = \frac{\sum_{i=0}^{n} |a_i \geq \boldsymbol{A}|}{\boldsymbol{n}} \qquad \textbf{(Equ 1)}$$

Combined Pi-score to quantify significance of differential expression

To prioritize differential expressed genes together with their cis-regulatory elements, we designed a computational method called a 'combined Pi-score'. Previously, the Pi-score has been proposed to improve the prioritization of differentially expressed genes by combining expression fold change and statistical significance (ie, the p-value) [26]. We modified the original Pi-score to fit the current data distribution (**Equ 2**).

$$\boldsymbol{\pi_i = \varphi_i \cdot (-log_{10} P_i)} \qquad \textbf{(Equ 2)}$$

Where side effect $\boldsymbol{\varphi_i}$ was either the fold change when fold change is finite, or it was an assigned value ±X representing the edging values of all finite fold changes when fold change is infinite. Here, we averaged the Pi-score of a ncRNA *i* and the Pi-score of its gene target to calculate the combined Pi-score (**Equ 3**). In this study, different strategies to select the target genes of a ncRNA were compared.

$$c\pi^i = [\pi^i_{ncRNA} + \max(\pi^i_{target})]/2 \qquad \textbf{(Equ 3)}$$

Briefly, We calculated the Pi-score for both the noncoding transcriptome and coding transcriptome and compared the combined Pi-scores of two models of ncRNA targeting. The “nearest mapping” model assumes that ncRNAs identify enhancers that regulate the nearest target gene, the “broader mapping” model assumed that all potential TBX5-dependent target genes within a window of 2 Mb could be the targets of an enhancer.

Motif enrichment Analysis

We tested the motif enrichment for the identified enhancers as well the whole set of chromatin accessible regions in HL1 cells. 637 pre-compiled threshold-free position weight matrices (PWM) for mouse from the Bioconductor package MotifDb were downloaded to test. Motif enrichment was calculated using the threshold-free affinity scoring algorithm with a lognormal distribution background [40]. Significance was considered when $P<0.05$ and enrichment score >1.

Functional enrichment of gene targets of enhancers

A background corrected Fisher’s exact test was performed using the Bioconductor package Seq2pathway[35]. Functional enrichment test was examined against 253 Gene Ontology Biological Process (BP) terms that have directly evidence and annotated to less than 1000 gene members. Significance was considered when FDR<0.05 and the score of odds ratio >1.

Landscape for mouse heart enhancer marks

The genome-wide landscape of H3K27Ac in whole mouse adult heart was called by macs2 and shared by the authors (GSE52123) [19]. There were two biological replicates and 44,707 regions commonly marked by H3K27Ac. The footprinting for DNase I hypersensitive sites in whole mouse adult heart was downloaded from ENCODE/University of Washington ([20]). Transcription factor and P300 landscapes in HL1 cardiomyocytes were downloaded from GEO (GSE21529) [18].

Retrieving phastCons sequence conservation scores

The phastCons30 sequence conservation scores were retrieved by querying against the UCSC Table Browser using the Bioconductor package rtracklayer [41]. An average score was calculated for each selected regions, removing the NA values. For certain large classes of genomic regions, eg, lincRNAs, open chromatin regions, we randomly sample 2000 regions first, and then retrieve the conservation scores for each sampled region to represent the feature of the whole class.

Meta-gene analysis

To compare the accessibility of chromatin at selected regions for groups of features (**Fig 3.1A**), meta-gene analysis was performed using the Bioconductor package metagene (v 2.2.1). Multiple combinations of group of features were compared in a single analysis, and BAM files instead of the peaks produced by peak callers were used to increase the sensitivity. Briefly four steps were performed for meta-gene analysis. 1) The read coverages of all selected regions were extracted from BAM files and normalized to read per million aligned (RPM). 2) Binned the position of each selected region with a size parameter 50bp, and calculate the averaged RPM within each bin. 3) Estimate the

confidence interval of each binned position using bootstrap. 4) And finally, plot the mean value as a line and the estimation as a ribbon.

The same analysis was also done to compare the TF-dependent expression at selected regions for groups of features (**Fig 3.3A**).

4C experiments

4C-seq was performed as described previously [25]. In short, HL-1 cardiomyocytes were collected and cross-linked with 1% formaldehyde in PBS with 10% FCS for 10 min at room temperature, nuclei were isolated and cross-linked DNA was digested with a primary restriction enzyme recognizing a 4-bp restriction site (DpnII), followed by proximity ligation. Cross-links were removed and a secondary restriction enzyme digestion (Csp6I), followed again by proximity ligation. For all experiments, 200 ng of the resulting 4C template was used for the subsequent PCR reaction, of which 8 (total: 1.6 µg of 4C template) were pooled and purified for next-generation sequencing.

Sequencing reads were aligned using bowtie2 (v2.2.5 with parameter -q set to 1) and processed as described previously [25]. In short, reads are mapped to a restricted mouse reference genome (mm9), consisting of sequences directly flanking the 4C primary restriction enzyme site (DpnII), called 4C frag-ends. Non-unique frag-ends are discarded in subsequent analysis. After mapping, the highest covered frag-end was removed from the dataset in the normalization process and data is read-depth normalized to 1 million aligned intrachromosomal reads. 4C-Seq coverage profiles are calculated as "running means", i.e. coverage averages of 21 consecutive 4C frag-ends.

Nuclear Fractionation for Assesing Chromatin Enrichment

Fraction of HL-1 nuclei was performed as previously described{Werner:2015bb, Bhatt:2012cc, Wuarin:1994ts}. In short, ~15 million cells from 3 different populations were lysed with 0.1% Triton X-100 and nuclei were pelleted 1,300 x g for 12 minutes, 4°. Nuclei were re-suspended in 250 µl NRB (20 mM HEPES pH 7.5, 50% Glycerol, 75 mM NaCl, 1 mM DTT, 1 x protease inhibitor cocktail), and then an equal volume of NUN buffer (20 mM HEPES, 300 mM NaCl, 1 M Urea, 1% NP-40 Substitute, 10 mM MgCl2, 1 mM DTT) was added and cells were incubated for 5 minutes on ice. After centrifugation (1,200 x g, 5 min, 4°C) RNA from the supernatant (soluble nuclear extract) and chromatin (pellet) were extracted using Trizole, purified on Zymo RNA Clean and Concentrator columns, and then converted to cDNA using random hexamer primers and High Performance MMLV Reverse Transcriptase from Epicentre. qPCR was then performed using SYBR green master mix (Life Technologies), with 18S rRNA (F primer: CGCAGCTAGGAATAATGGAATAGG, R primer: GCCTCAGTTCCGAAAACCAA) as an internal standard for both fractions.

Relative Luciferase activity

Luciferase response assays were performed as previously described[16]. Candidate regulatory elements were amplified from C57/B6 mouse genomic DNA or synthesized via Gibson Assembly. Sequence was verified and then cloned into the pGL4.23 enhancer luciferase response vector with minimal promoter. HL-1 Cardiomyocytes were co-transfected with luciferase response vector and a pRL control using Lipofectamine 3000,

cultured for 48 hours after transfection, then lysed and assayed using the Dual-Luciferase Reporter Assay system (Promega).

Antisense mediated knockdown

Custom antisense oligonucleotides (ASOs) were designed and synthesized by IDT. Pooled ASOs were transfected with Lipofectamine 3000 into HL-1 cardiomyocytes at a concentration of 10nM. Cells were sorted into transfected and non-transfected populations by Flow Cytometry based on co-transfection with a vector encoding mCherry. ncRNA expression, *Ryr2* coding transcript, and calcium kinetics were measured. ASOs against the *Ryr2* lncRNA were: 5'-CCCTTCATATCACGTTGGAA-3', 5'-ATCACCTGCCCTGGTTCTTT-3', 5'-TATGATCACCTGCCCTGGTT-3', 5'-CTTCATGCTTATCTGACAAG-3'. Antisense oligos against the *Ryr2* mRNA were 5'-CGATGTCCTTGGCAGGCTCA-3' 5'-CATCGTCCATGTGGCCTTCG-3' , 5'-ATTCATCCATGTGCCCATGC-3', 5'- GTCTGTCCCACTGGCCTTTG -3', and 5'-ATGTCTGGCGTGTCACCATC -3'.

Calcium transient measurement

Cytosolic $[Ca^{2+}]_i$ was measured with the high-affinity Ca*2+* indicator Fluo-4 (Molecular Probes/Invitrogen,) using a laser scanning confocal microscopy (Zeiss LSM 510,) equipped with a ×63/1.40 NA oil-immersion objective lens. HL-1 cells were sorted and plated on coated coverslips and incubated at room temperature with 10 μM Fluo-4/AM for 15 min in normal Tyrode's solution containing (in mM): 140 NaCl, 4 KCl, 10

glucose, 10 HEPES, and 1 $MgCl_2$, 1 $CaCl_2$, with pH 7.4 using NaOH, followed by a 10 min perfusion wash with Tyrode (at 37˚C) without Fluo-4 before recording. Fluo-4 was excited with the 488-nm line of an argon laser and fluorescence was measured at >515 nm. Spontaneous Ca^{2+} transients were acquired in line-scan mode (3 ms per scan; pixel size 0.12 μm). Ca^{2+} transients are presented as total fluorescence intensity normalized to resting fluorescence (F/F_0) during steady-state conditions before field stimulation. Cells were recorded at baseline, then isopropteranol was perfused to a final concentration of 1nM and calcium transients were recorded again.

References

1. Bonn, S. *et al.* Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nat Genet* **44**, 148-56 (2012).
2. Creyghton, M.P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* **107**, 21931-6 (2010).
3. Werner, M.S. & Ruthenburg, A.J. Nuclear Fractionation Reveals Thousands of Chromatin-Tethered Noncoding RNAs Adjacent to Active Genes. *Cell Rep* **12**, 1089-98 (2015).
4. Visel, A. *et al.* ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854-8 (2009).
5. Visel, A., Minovitsky, S., Dubchak, I. & Pennacchio, L.A. VISTA Enhancer Browser--a database of tissue-specific human enhancers. *Nucleic Acids Res* **35**, D88-92 (2007).
6. Zentner, G.E. & Scacheri, P.C. The chromatin fingerprint of gene enhancer elements. *J Biol Chem* **287**, 30888-96 (2012).
7. Nadadur, R.D. *et al.* Pitx2 modulates a Tbx5-dependent gene regulatory network to maintain atrial rhythm. *Sci Transl Med* **8**, 354ra115 (2016).
8. Wang, D. *et al.* Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* **474**, 390-4 (2011).
9. He, A. *et al.* Dynamic GATA4 enhancers shape the chromatin landscape central to heart development and disease. *Nat Commun* **5**, 4907 (2014).

10. Stergachis, A.B. *et al.* Conservation of trans-acting circuitry during mammalian regulatory evolution. *Nature* **515**, 365-70 (2014).

11. Cusanovich, D.A., Pavlovic, B., Pritchard, J.K. & Gilad, Y. The functional consequences of variation in transcription factor binding. *PLoS Genet* **10**, e1004226 (2014).

12. Danko, C.G. *et al.* Identification of active transcriptional regulatory elements from GRO-seq data. *Nat Methods* **12**, 433-8 (2015).

13. Kim, T.K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182-7 (2010).

14. Rosenbloom, K.R. *et al.* The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res* **43**, D670-81 (2015).

15. Quinn, J.J. & Chang, H.Y. Unique features of long non-coding RNA biogenesis and function. *Nat Rev Genet* **17**, 47-62 (2016).

16. He, A., Kong, S.W., Ma, Q. & Pu, W.T. Co-occupancy by multiple cardiac transcription factors identifies transcriptional enhancers active in heart. *Proc Natl Acad Sci U S A* **108**, 5632-7 (2011).

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

18. Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455-61 (2014).

19. Mouse, E.C. *et al.* An encyclopedia of mouse DNA elements (Mouse ENCODE). *Genome Biol* **13**, 418 (2012).

20. van de Werken, H.J. *et al.* Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nat Methods* **9**, 969-72 (2012).

21. Xiao, Y. *et al.* A novel significance score for gene selection and ranking. *Bioinformatics* **30**, 801-7 (2014).

22. Li, W., Notani, D. & Rosenfeld, M.G. Enhancers as non-coding RNA transcription units: recent insights and future perspectives. *Nat Rev Genet* **17**, 207-23 (2016).

23. Lai, F. *et al.* Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* **494**, 497-501 (2013).

24. Li, W. *et al.* Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* **498**, 516-20 (2013).

25. Wang, B., Cunningham, J.M. & Yang, X.H. Seq2pathway: an R/Bioconductor package for pathway analysis of next-generation sequencing data. *Bioinformatics* **31**, 3043-5 (2015).

26. Sondergaard, M.T. *et al.* The Arrhythmogenic Calmodulin p.Phe142Leu Mutation Impairs C-domain Ca2+-binding but not Calmodulin-dependent Inhibition of the Cardiac Ryanodine Receptor. *J Biol Chem* (2016).

27. Ploner, A., Ploner, C., Lukasser, M., Niederegger, H. & Huttenhofer, A. Methodological obstacles in knocking down small noncoding RNAs. *RNA* **15**, 1797-804 (2009).

28. Landt, S.G. *et al.* ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res* **22**, 1813-31 (2012).

Acknowledgments: This research was supported in part by NIH through resources provided by the Computation Institute and the Biological Sciences Division of theUniversity of Chicago and Argonne National Laboratory, under grant 1S10OD01849501.We specifically acknowledge the assistance of Lorenzo Pesce.

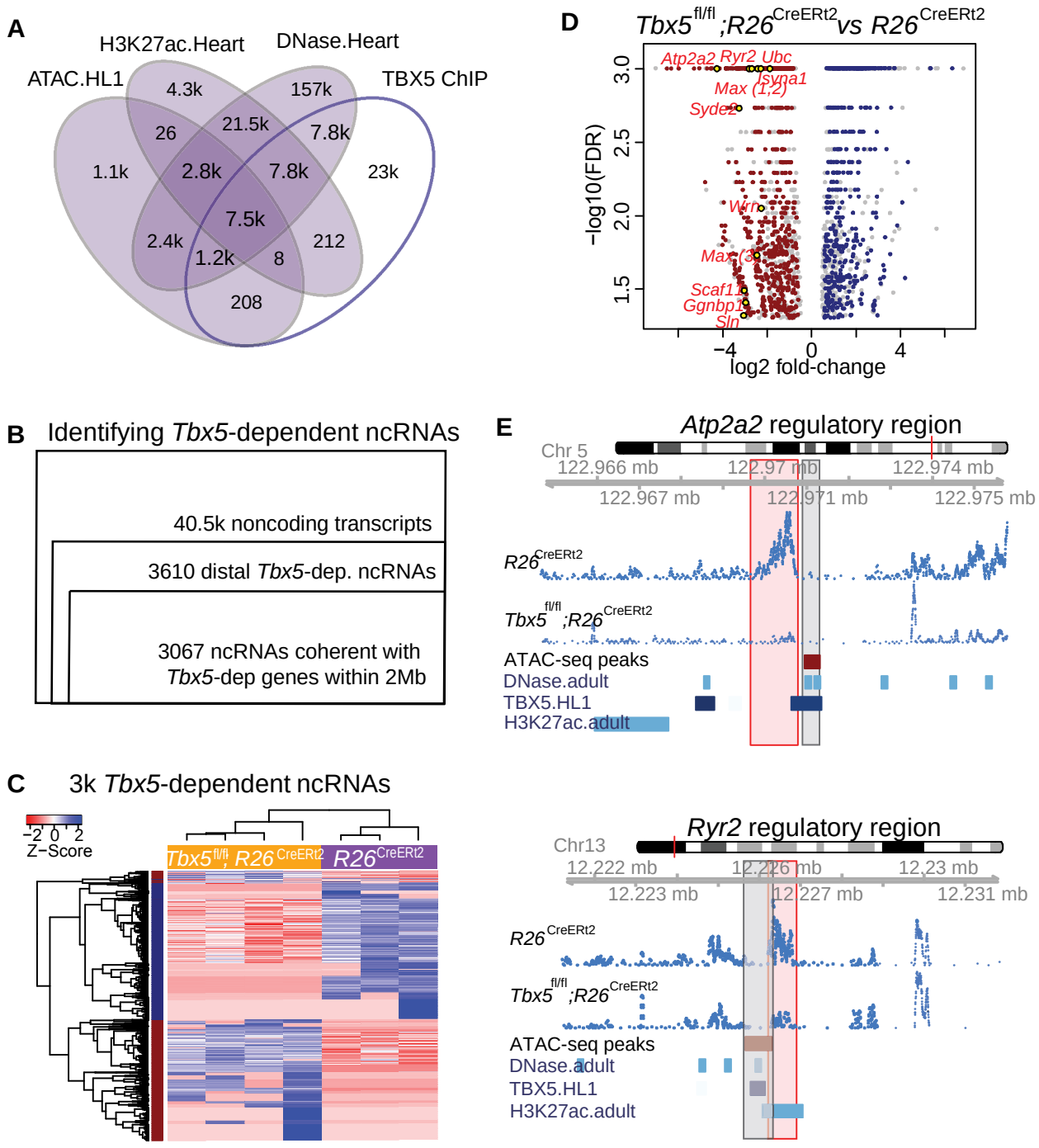

Figure 3.1. **TF-dependent noncoding transcription defines regulatory elements**.

A) Venn-diagram of peak call overlaps for HL-1 ATAC-Sequencing (ATAC.HL1), Histone 3 Lysine 27 Acetylation in mouse heart (H3K27ac.Heart, GSE52123), DNase

(Fig 3.1, Continued)

Hypersensitivity in mouse heart (DNase.heart, ENCODE) and TBX5 ChIP-seq (TBX5.HL1, GSE21529). **B**) Workflow for identifying TF-dependent ncRNAs. Total noncoding transcripts from mouse left atrium, narrowed to *Tbx5*-dependent, distal intergenic ncRNAs defined as >1kbp away from known transcriptional start sites of known coding genes (GENCODE mm10), and narrowed again to coherent changes with nearby *Tbx5*-dependent genes within a 2Mbp window.

C) Heatmap of identified *Tbx5*-dependent ncRNAs in $Tbx5^{fl/fl}; R26^{CreERt2}$ (yellow, left) and corresponding $R26^{CreERt2}$ control (purple, right) in left atrial tissues. Hierarchical cluster analysis based on Euclidean distance of normalized sequencing counts. 1577 *Tbx5*-activated ncRNAs were downregulated after *Tbx5* deletion and 1490 *Tbx5*-repressed ncRNAs were upregulated after *Tbx5* deletion.

D) Volcano-plot of significantly misregulated TF-dependent ncRNAs, select identifications were labeled by nearest TBX5 dependent gene. Plot of $\log_2$ fold-change of ncRNAs in $Tbx5^{fl/fl}; R26^{CreERt2}$ compared to $R26^{CreERt2}$ vs $-\log_{10}$ False Discovery Rate (FDR) for the same comparison (FDR<0.05, |FC|>2). The ncRNAs within 2Mb of coherently mis-expressed TBX5-dependent genes are red or blue for activated and repressed, respectively. Gray dots represent those ncRNAs without coherently mis-expressed coding genes in the 2Mb window.

E) Example genomic views of two of the most significantly TF-dependent ncRNAs, adjacent to the *Atp2a2* (top) and *Ryr2* (bottom) genes, respectively. Top track is chromosomal location, followed by the ncRNA read density from $R26^{CreErt2}$control and $Tbx5^{fl/fl}; R26^{CreERt2}$. Below is ATAC-Seq peak call in HL-1 cells, cardiac DNase

hypersensitivity (ENCODE), TBX5 ChIP-seq (GSE21529) and cardiac H3k27 Acetylation (GSE52123). The identified differential ncRNA is marked in the red box, putative regulatory element as defined by above enhancer marks in gray box.

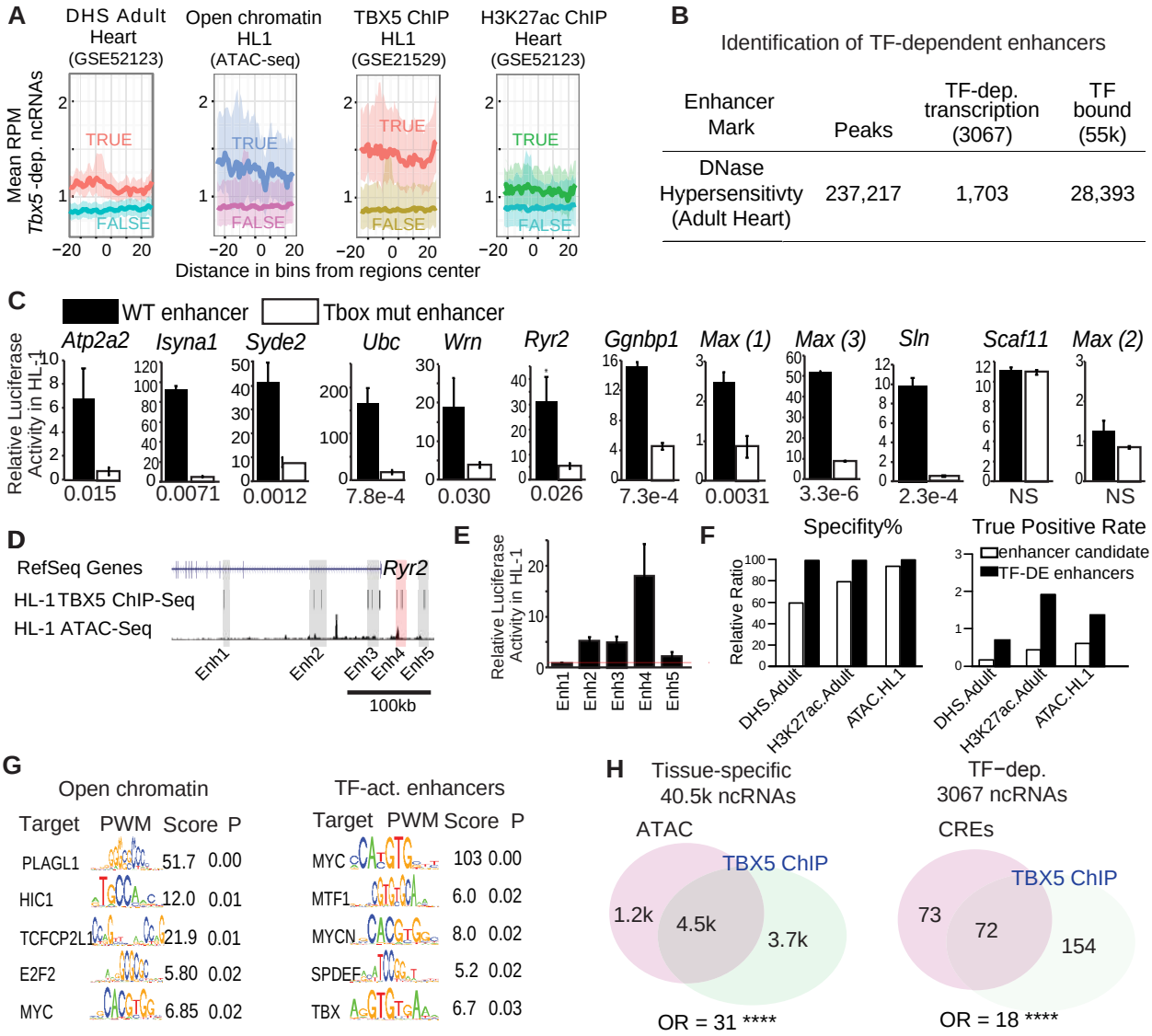

Figure 3.2. **TF-dependent ncRNAs predict functional enhancers.**

A) Meta-gene plot showing Reads per Million (RPM) of *Tbx5*-dependent ncRNAs, being divided by enhancer makers such as DNase Hypersensitivity, ATAC Sequencing in HL-1 cells, TBX5 ChIP-Seq, and H3K27Ac, respectively. Each subpanel is centered around all *Tbx5*-dependent ncRNAs, the averaged RPM of the ncRNAs hold an enhancer mark (top line, TRUE) vs those ncRNAs that do not (bottom line, FALSE).

Fig 3.2 (continued)

B) Identifying TF-dependent ncRNAs associated with cis-regulatory elements (CREs), by overlapping with accessible chromatin defined by DNase Hypersensitivity. Overlap of ncRNAs (left) and TF ChIP (right), respectively.

C) Relative luciferase activity of select enhancers identified in HL-1 cardiomyocytes, labeled by nearest TBX5-dependent gene. Activity of wild type enhancers (black) and T-box mutant enhancers (white) are shown. P-value of wildtype vs mutant enhancers are given below each graph. n=3 each, activity normalized to co-transfected Renilla vector, and to vector with scrambled insert.

D) Genomic view of *Ryr2* locus showing four candidate TBX5-dependent regulatory elements identified by TBX5 ChIP-seq (gray) and by differential ncRNA expression (Enh 4, red).

E) Relative luciferase activity in HL-1 cardiomyocytes of candidate *Ryr2* Enhancers, normalized to co-transfected Renilla vector, and to vector with scrambled insert (right, n>3 replicates).

F) Predictive precision and specificity of enhancer marks and *Tbx5*-dependent ncRNA defined enhancers were calculated and compared, against 240 cardiac and 1162 non-cardiac enhancers reported in the VISTA database. The enhancers were defined by DNase hypersensitivity (DHS), H3K27Acetylation, and open chromatin by ATAC-seq.

G) Motif analysis showing enriched motifs in HL-1 open chromatin (ATAC, left) and ncRNA-defined TBX5-dependent enhancers (right). Motif patterns viewed as position weight matrix (PWM) with enrichment statistics. Enrichment scores >1.5 and P-value<0.05 were considered significant and the top five most significant motifs reported.

(Fig 3.2, Continued)

H) Venn diagram of ATAC-seq and TBX5 ChIP-seq in background of ncRNAs (left) and TF-dependent elements overlapped with open chromatin in background of TF-dep ncRNAs (right). Odds Ratio (OR) for each background. **** indicates P<1e-10.

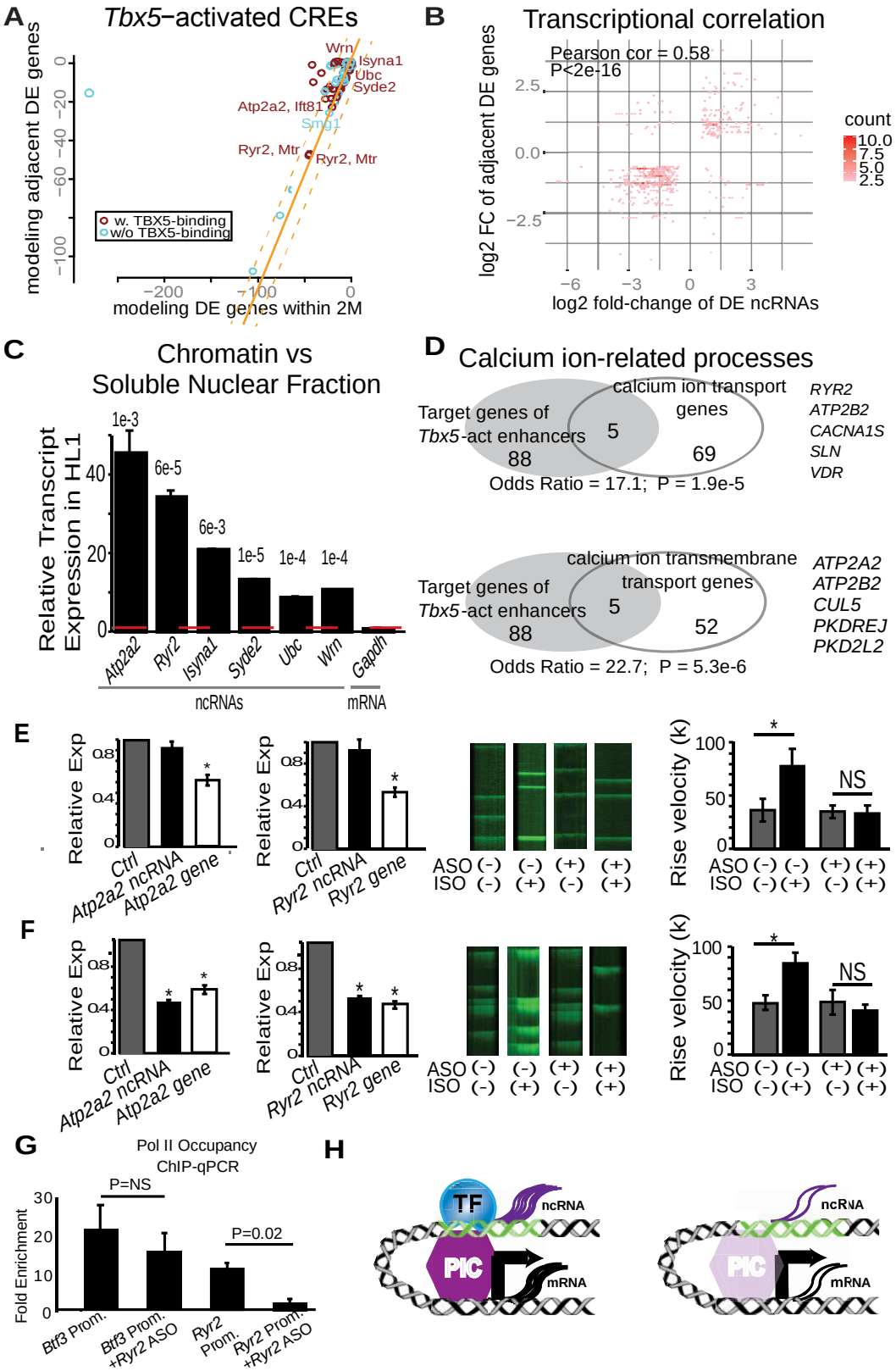

Figure 3.3. Enhancer Transcription mediates target gene expression

(Fig 3.3, Continued)

A) Sparse scatterplot for combined Pi-scores calculated from *Tbx5*-activated enhancers and *Tbx5*-activated genes. X-axis is the pi-score using the "nearest" *Tbx5*-dependent gene where Y-axis is the pi-score using all potential TF-dependent gene targets within a 2Mbp window. Orange line indicates y=x. The dashed lines marge the standard deviations. Example enhancers were labeled by nearest *Tbx5*-dependent genes. Red indicates enhancers with TBX5 occupancy, and blue indicates enhancers without TBX5 occupancy by ChIP-seq.

B) Scatterplot in hexagonal binning for the 3067 identified TF-dependent ncRNAs. X-axis is the differential expression fold change on log2 scale of these ncRNAs, and Y-axis is the differential expression fold change on log2 scale of any nearest gene targets. The fold change with the maximum absolute value is used when both neighbor genes are TF-dependent.

C) Relative transcript expression of chromatin enriched vs soluble nuclear fraction in HL1 cardiomyocyte cells of ncRNAs, and *Gapdh* coding mRNA transcript as a control. Expression was normalized to 18S Ribosomal RNA. * indicates $P<0.05$

D) Gene Ontology functional enrichment for gene targets within a 2Mb window of the TBX5-activated enhancers. Odds Ratio (OR) and P-value of overlap with GO terms "Calcium Ion Transport Genes" and "Calcium Ion Transmembrane Transport Genes"

E-F) Relative transcript expression of *Atp2a2* (left) and *Ryr2* (right) lncRNA, mRNA and *Gapdh* after knockdown of mRNA (E) and lncRNA (F). Representative calcium transient line scans from control (left) and ASO knockdown (right) HL-1 cardiomyocytes in presence and absence of isopropteranol (+ISO, -ISO resp) after knockdown of *Ryr2*

(Fig 3.3, Continued)

mRNA (top) or lncRNA (bottom) with antisense oligonucleotides (ASO). Quantification of rise velocity under control and isoproterenol.

G) Fold-enrichment of RNA Polymerase 2 (Pol2) occupancy by ChIP at housekeeping gene *Btf3* promoter (prom.) and *Ryr2* promoter from control (left) and after antisense oligonucleotide knockdown of *Ryr2* enhancer ncRNA (ASO, right). Normalized to control locus near *Gapdh* gene.

H) TF-dependent regulatory element transcription. mRNA and enhancer transcription occur in the presence of Transcription Factor (TF) through the Transcription Preinitiation Complex (PIC) (left). Loss of the TF results in decreased PIC complex leading to decreased enhancer and mRNA transcription (right).

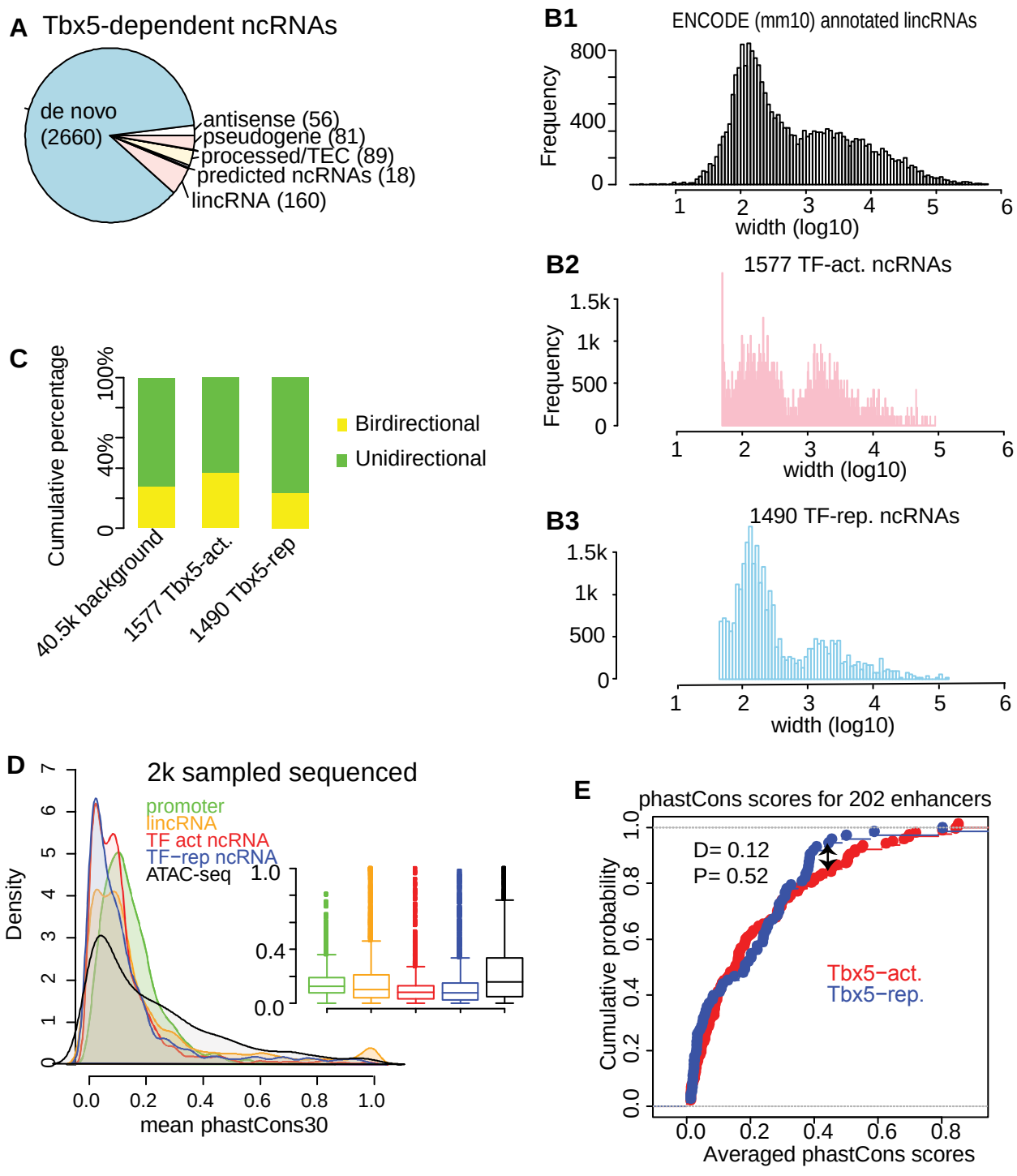

Figure S3.1. Genomic features of the identified TF-dependent ncRNAs recapitulate known features of annotated lincRNAs.

(Fig S3.1, Continued)

A) Pie chart showing the composition for the identified TF-dependent ncRNAs using *de novo* assembly (Cufflinks). The 3067 ncRNAs were categorized into 6 groups. GENCODE (mm10) annotated ncRNAs with relatively few instances were grouped together as 'predicted ncRNAs'. *De novo* indicates a previously unannotated noncoding transcript.

B) Histogram distribution of the ncRNA width on log10 scale. The distribution of GENCODE annotated lincRNAs, the identified TF-activated ncRNAs, and the TF-repressed ncRNAs are shown.

C) Stacked bar-graphs showing percentage of unidirectional (green) and bi-directional (yellow) noncoding transcripts for the total noncoding transcript calls, the *Tbx5*-activated, and the *Tbx5*-repressed ncRNAs identified.

D) Histograms and boxplots of average phastCons30 sequence conservation scores for promoter regions (green), GENCODE annotated lincRNAs (yellow), TBX5-activated and TBX5-repressed ncRNAS (red, blue resp.), and ATAC-Seq (black).

E) Two-sample Kolmogorov–Smirnov statistics. The cumulative probabilities exhibit no difference in conservation between *Tbx5*-activated and *Tbx5*-repressed cis-regulatory elements (red and blue lines, respectively).

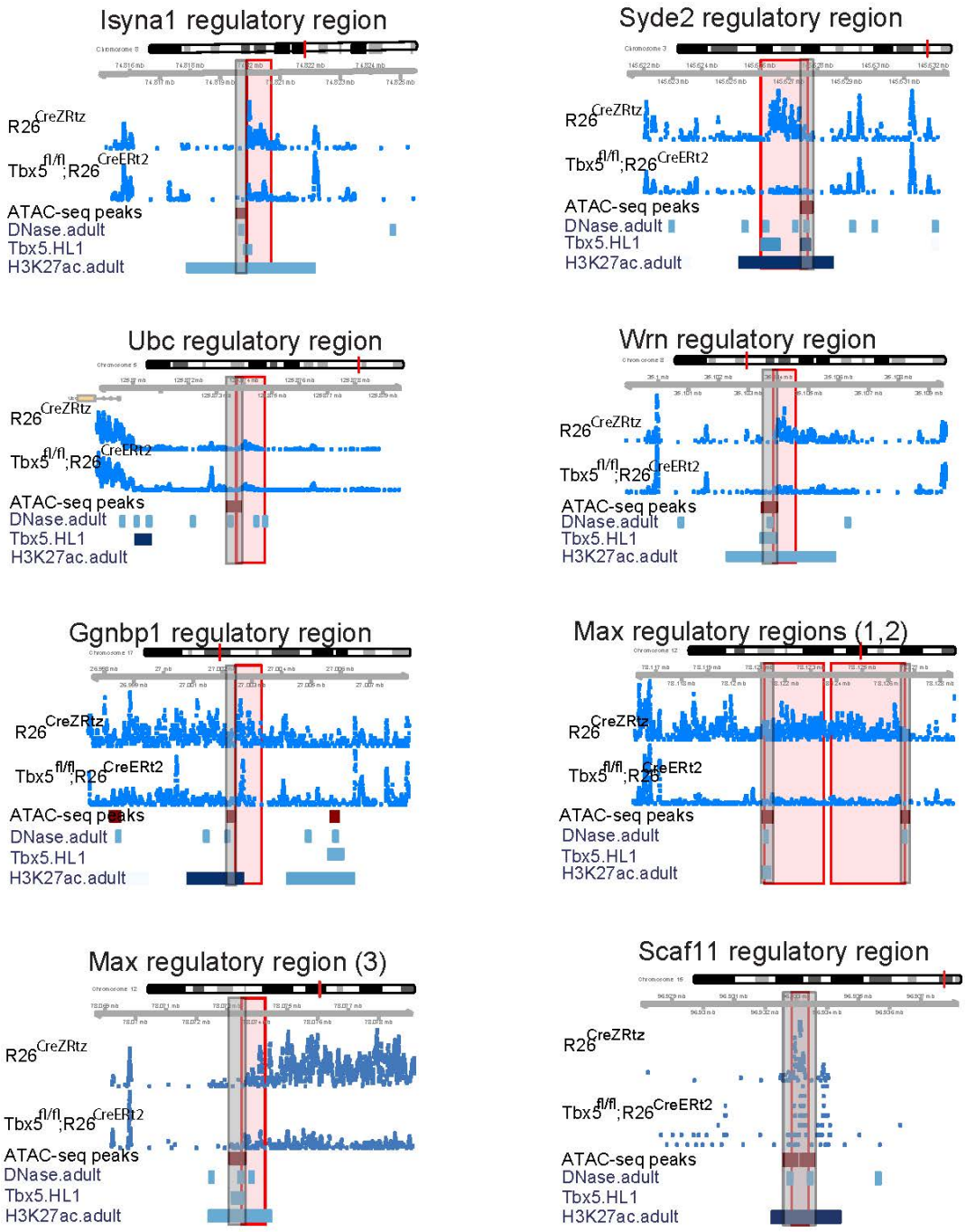

Figure S3.2. Genomic view of nine TF-dependent ncRNAs (mm9). Labeled by nearest TBX5-dependent coding gene. In each subpanel, the ncRNAs are shown and marked by a pink box. The genomic regions marked by general enhancer makers are shown below, which overlap with or are adjacent to the identified ncRNA, boxed in gray.

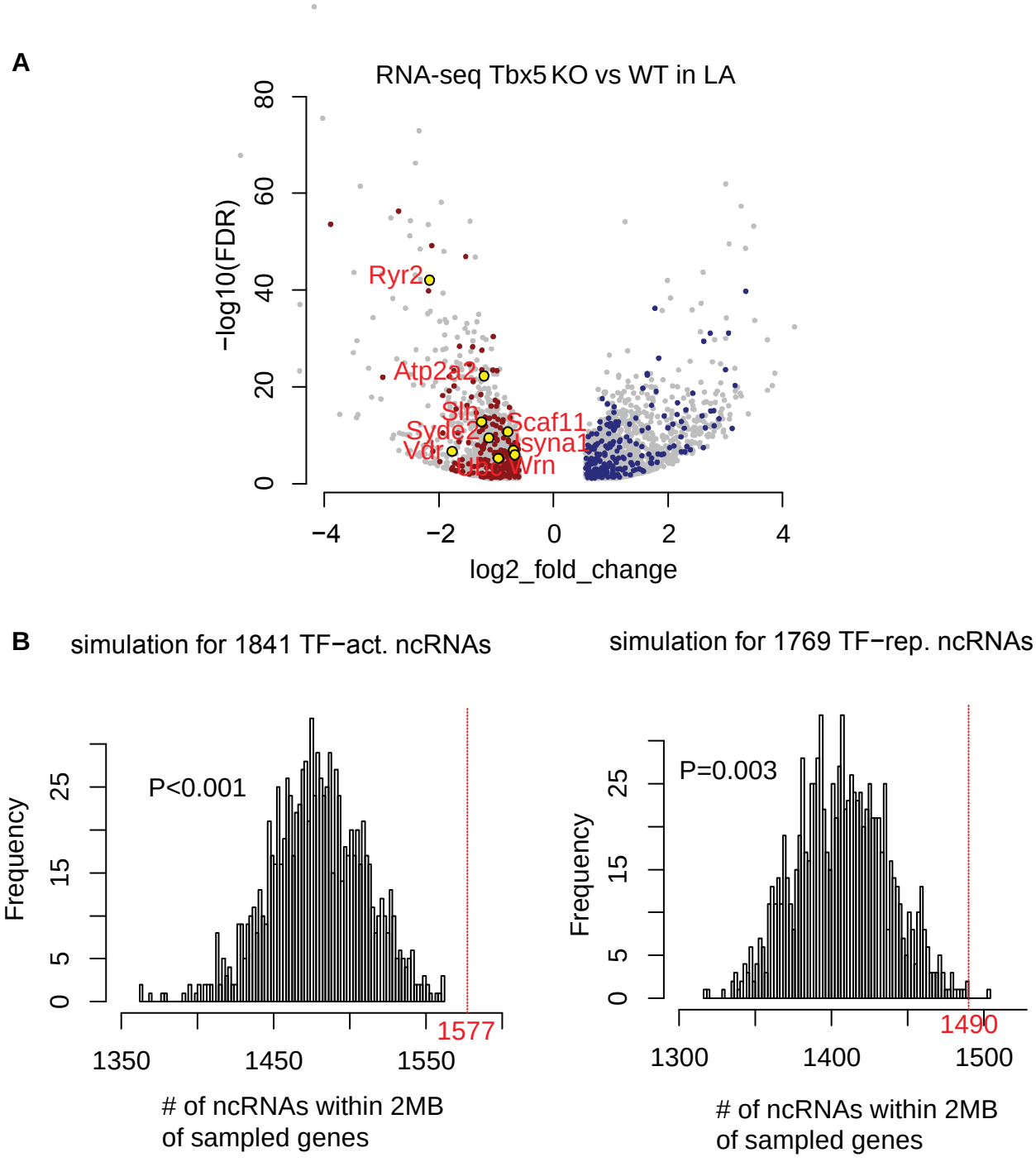

Figure S3.3. Identifying TF-dependent ncRNA targets from TF-dependent expressed genes. A) TF-dependent genes are coherent with TF-dependent ncRNAs.

(Fig S3.3, Continued)

Volcano-plot of TF-dependent genes (FDR<0.05, |FC|>1.5). Genes residing within 2Mb of coherently TBX5-dependent ncRNAs are colored (red for *Tbx5*-activated, blue for *Tbx5*-repressed).

B) Simulation study shows the empirical significance of *Tbx5*-dependent ncRNAs local to *Tbx5*-dependent coding-genes within a 2Mb window. Histogram showing the distribution of randomly sampled ncRNAs within this window to randomly sampled genes. The red dashed line showing the observed number in a *Tbx5*-activated or *Tbx5*-repressed, respectively. An empirical P-value was calculated.

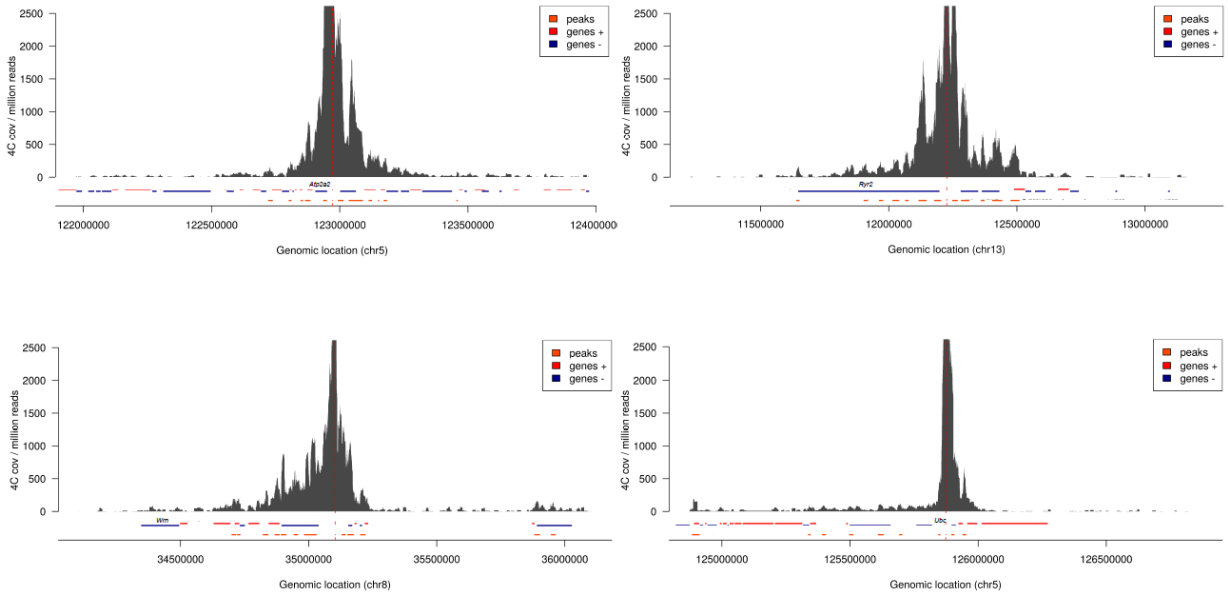

Fig S3.4. Chromosomal interactions of identified TF-dependent ncRNAs.

Genomic loci showing circularized chromatin conformation capture (4C) from viewpoint of 4 identified regulatory elements (near *Atp2a2*, *Wrn*, *Ryr2*, and *Ubc*). Total reads (top), and significant contacts (bottom, orange) plotted. Annotated genes shown in blue and red (sense and antisense).

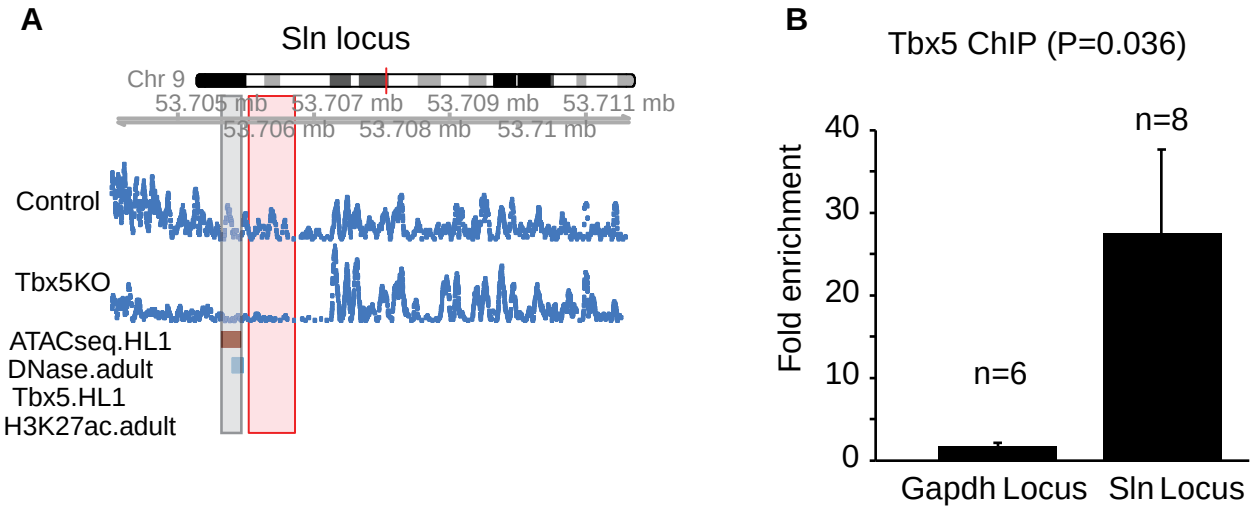

Figure S3.5. TBX5-dependent enhancer at *Sln*, not identified by TF ChIP-seq

A) Genomic view of a *Tbx5*-activated ncRNA region adjacent to gene *Sln*. This locus shows significant *Tbx5*-dependence on transcription expression in mouse atrium, open chromatin detected by ATAC-seq in HL1 cells, DHS signature in adult mice, but not TBX5 occupancy, nor H3K27Ac in HL1 cells (GSE21529).

B) ChIP-qPCR of TBX5 comparing this *Sln* locus with the *Gapdh* locus. The *Sln* locus shows enrichment for TBX5 occupancy (n=8, P=0.036). Two-side t-test assuming unequal variances was performed.

CHAPTER IV: CONCLUSIONS

Regulatory Architecture in Atrial Rhythm

The first chapter of this dissertation described a multi-tiered gene regulatory network for atrial rhythm. The network is driven by the T-box transcription factor TBX5. TBX5 directly drives expression of the transcriptional repressor *PITX2*, the most implicated locus in AF. TBX5 and PITX2 co-regulate a network of essential atrial rhythm-control channels. This network comprises almost half of all AF susceptibility loci identified by Genome Wide Association Studies (GWAS) [1-10].

Interestingly, *Tbx5* and *Pitx2* antagonistically co-regulate a shared gene regulatory network. Reduction of *Tbx5* or *Pitx2* causes opposite effects on the atrial rhythm network and AF susceptibility. Thus AF susceptibility can be achieved from opposite perturbations to atrial channel gene expression. The paradigm of dosage-sensitivity in cardiac rhythm control has long been implicit in the literature of the field. Both gain and loss of function mutations in ion channels and transcriptional regulators have been observed to cause atrial fibrillation [10-20]. The observed dosage-sensitivity shows that atrial fibrillation is a common endpoint of perturbation to atrial conduction. This defines a previously unappreciated molecular complexity to AF causation and the need for a personalized approach to AF diagnosis and therapy – opposite perturbations to the network may require different clinical interventions.

We observed the antagonistic relationship between TBX5 and PITX2 at the cellular level: loss of *Tbx5* causes prolongation of the atrial action potential while loss of *Pitx2* causes shortening of the atrial action potential. More sophisticated cellular voltage and current clamp studies have been performed in the context of *Pitx2* deficiency, and suggest mechanisms for the efficacies of different antiarrhythmic agents when these perturbations are present[21]. However,

more in depth analysis of the causes of the prolonged action potential in *Tbx5*-deficiency remain to be performed. We observed that intracellular calcium-chelation rescued the prolongation of the atrial action potential and propensity for abnormal triggered activity seen in *Tbx5*-deficiency. Further cellular electrophysiology measuring sarcoplasmic reticulum calcium load, calcium reuptake kinetics, and calcium extrusion in the context of *Tbx5*-deficiency may suggest specific pharmacologic therapies or predict efficacy of antiarrhythmic drugs, similar to the studies performed with *Pitx2*.

The transcriptional architecture we uncovered for atrial rhythm control is an incoherent-feed forward loop. Incoherent feed-forward loops impose molecular buffering on terminal regulatory network transcriptional output in response to upstream variance in gene expression [22-24]. This architecture provides buffering to cardiac ion channel gene expression, establishing a molecular mechanism for cardiac rhythm robustness. This work offers a gene regulatory paradigm essential for homeostasis of organ physiology.

There are a number of known transcriptional regulators linked to AF and related arrhythmias, notably *TBX5, PITX2, PRRX1, ZHFX3, NKX2.5 GATA4, GATA6*[19]. Understanding the interactions between these transcriptional regulators is critical to expanding our knowledge of the transcriptional networks underlying AF. The interactions between *Tbx5*, *Nkx2.5* and *Gata4/6*, and their transcriptional repressor *Zfpm2* (FOG2) have been studied in a number of contexts, notably during cardiac development[25,26-28,29, 30]. However, the interactions of these factors in the adult heart, and specifically their role in atrial rhythm, have yet to be defined. The study of *Tbx5* and *Pitx2* provide a framework to pursue future studies of transcription factor interactions in the context of dosage sensitivity in atrial rhythm.

An additional mechanism suggested by the described dosage sensitive model involves feedback regulation through microRNAs. MicroRNAs, small regulatory RNAs that post-transcriptionally regulate mRNA transcription, represent canonical dosage-sensitive feedback mechanisms in regulatory networks. MicroRNAs can function as buffers against fluctuation in gene expression by sensitive, post-transcriptional fine tuning of gene expression [31]. Indeed, a number of microRNA regulatory loops have been described in atrial rhythm [32,33]. Understanding the interplay of microRNAs, transcriptional regulators, and downstream effectors in the adult mouse heart will further refine our understanding of the mechanisms that maintain stable gene expression and normal cardiac rhythm.

Defining Transcription Factor Dependent *Cis*-Regulatory Elements

The second chapter of this thesis focuses on the challenge of defining transcription factor-dependent regulatory elements. We aim to define the essential *cis*-regulatory elements downstream of TBX5 responsible for maintaining atrial rhythm.

Defining the functional cis-regulatory elements within a gene regulatory network remains a fundamental challenge. Current whole-genome approaches to enhancer identification can be challenging, and the majority of genomic regions marked by currently accepted methods are not all functionally relevant. The accepted approach, Chromatin Immunoprecipitation (ChIP), describes genome-wide locations of TF binding; however, it does not distinguish functional TF-enhancer localization from more common non-functional localization[34].

Enhancer transcription has recently been identified as essential for enhancer function [35]. We demonstrated that transcription factor (TF)-dependent enhancer transcription identifies essential enhancer-transcribed non-coding RNAs (ncRNAs) and functionally robust TF-

dependent enhancers. We applied a novel approach, genome-wide TF-dependent ncRNA profiling, to define TF-dependent enhancers. *Tbx5*-dependent ncRNA-seq defined *Tbx5*-dependent enhancers and associated non-coding transcripts. *Tbx5*-dependent enhancer transcription was a more specific mark for functional enhancers than accepted enhancer marks including chromatin accessibility, histone modifications, and TBX5 binding.

Interestingly, *Tbx5*-dependent ncRNA-defined enhancers were highly enriched for TBX5 binding *a-priori*. This approach therefore bypassed the need for ChIP-seq to define TF-dependent enhancers. Furthermore, this approach provided a quantitative metric for TF-dependent enhancer function, as the direction and magnitude of *Tbx5*-dependent enhancer transcription correlated strongly with that of *Tbx5*-dependent target gene expression. This suggests that enhancer transcription represents an extremely sensitive readout of enhancer function. A powerful application of this approach would be the ability to detect minor perturbations in enhancer function, such as those caused by single nucleotide polymorphisms (SNPs) that subtly affect transcription factor binding or enhancer function.

Lastly, differential ncRNA-seq identified a physiologically relevant gene regulatory network, described in Chapter I - expression of a calcium-handling network essential for normal atrial rhythm. Interestingly, *Tbx5*-dependent ncRNAs were required for *Tbx5*-target gene expression and calcium-handling physiology, suggesting a functional role for the ncRNAs in gene regulation. Defining enhancer-associated ncRNA mechanisms in recruitment of transcriptional regulators, formation of enhancer-promoter loops, and chromatin reorganization is necessary to defining their functional role. Separating the role of the noncoding transcript from the associated enhancer using genome editing technologies such as CRISPR, as previously described[36], may also shed further insight on the role of these ncRNAs.

Differential ncRNA-seq not only identifies robust *Tbx5*-dependent elements, but also suggests the identification of indirect downstream elements. As we describe, loss of TBX5, canonically a transcriptional activator, was associated with loss of transcription of many candidate regulatory elements, suggesting activity of this TF at these elements. However, we also observed increased transcription from a number of loci, suggesting activation of normally silent regulatory elements. These elements likely represent indirect targets that activate only in the context of TBX5-deficiency. Activation of these context-specific enhancers suggests the exciting possibility of identifying enhancers that do not appear active in healthy or wild-type tissue. This observation has enormous implications for identifying disease-associated enhancers that would be impossible to identify using the multitude of datasets for defining the regulatory landscape of normal tissue[37].

TF-dependent ncRNA-seq is a novel approach for identification of TF-dependent functional enhancers and associated functional ncRNAs, allowing direct construction of a TF-dependent gene regulatory network. TF-dependent ncRNA-seq is conceptually novel and methodologically straightforward. This approach overcomes many of the limitations of current approaches to define TF-dependent enhancers. We apply this approach to identify functionally relevant TF-dependent ncRNAs and TF-dependent enhancers genome-wide that define a physiologically relevant TF-dependent gene regulatory network. This approach may become part of the standard molecular methodologies for the identification of TF-dependent enhancers. We propose testing this approach to define regulatory networks across biological contexts.

Conclusions

This dissertation defines a transcriptional architecture for cardiac rhythm control, and proposes a novel approach for characterizing context-specific *cis*-regulatory landscapes. Defining the regulatory networks underlying cardiac rhythm is essential to guide improved diagnostic, therapeutic, and preventive strategies to reduce the profound healthcare burden associated with cardiac arrhythmias.

References

1. Benjamin, E.J., *et al.* Variants in ZFHX3 are associated with atrial fibrillation in individuals of European ancestry. *Nat Genet* **41**, 879-881 (2009).
2. Christophersen, I.E. & Ellinor, P.T. Genetics of atrial fibrillation: from families to genomes. *J Hum Genet* (2015).
3. Ellinor, P.T., *et al.* Meta-analysis identifies six new susceptibility loci for atrial fibrillation. *Nat Genet* **44**, 670-675 (2012).
4. Ellinor, P.T., *et al.* Common variants in KCNN3 are associated with lone atrial fibrillation. *Nat Genet* **42**, 240-244 (2010).
5. Gudbjartsson, D.F., *et al.* Variants conferring risk of atrial fibrillation on chromosome 4q25. *Nature* **448**, 353-357 (2007).
6. Lubitz, S.A., *et al.* Novel genetic markers associate with atrial fibrillation risk in Europeans and Japanese. *Journal of the American College of Cardiology* **63**, 1200-1210 (2014).
7. Mahida, S., Lubitz, S.A., Rienstra, M., Milan, D.J. & Ellinor, P.T. Monogenic atrial fibrillation as pathophysiological paradigms. *Cardiovasc Res* **89**, 692-700 (2011).
8. Pfeufer, A., *et al.* Genome-wide association study of PR interval. *Nat Genet* **42**, 153-159 (2010).
9. Sinner, M.F., *et al.* Integrating genetic, transcriptional, and functional analyses to identify 5 novel genes for atrial fibrillation. *Circulation* **130**, 1225-1235 (2014).
10. Tucker, N.R. & Ellinor, P.T. Emerging directions in the genetics of atrial fibrillation. *Circ Res* **114**, 1469-1482 (2014).

11. Bers, D.M., Pogwizd, S.M. & Schlotthauer, K. Upregulated Na/Ca exchange is involved in both contractile dysfunction and arrhythmogenesis in heart failure. *Basic Res Cardiol* **97 Suppl 1**, I36-42 (2002).

12. Bruneau, B.G., *et al.* Chamber-specific cardiac expression of Tbx5 and heart defects in Holt-Oram syndrome. *Developmental biology* **211**, 100-108 (1999).

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

14. Christophersen, I.E., *et al.* Rare variants in GJA5 are associated with early-onset lone atrial fibrillation. *Can J Cardiol* **29**, 111-116 (2013).

15. McDermott, D.A., Hatcher, C.J. & Basson, C.T. Atrial Fibrillation and Other Clinical Manifestations of Altered TBX5 Dosage in Typical Holt-Oram Syndrome. *Circ Res* **103**, e96 (2008).

16. Olesen, M.S., Nielsen, M.W., Haunso, S. & Svendsen, J.H. Atrial fibrillation: the role of common and rare genetic variants. *Eur J Hum Genet* **22**, 297-306 (2014).

17. Patel, C., Silcock, L., McMullan, D., Brueton, L. & Cox, H. TBX5 intragenic duplication: a family with an atypical Holt-Oram syndrome phenotype. *Eur J Hum Genet* **20**, 863-869 (2012).

18. Postma, A.V., *et al.* A gain-of-function TBX5 mutation is associated with atypical Holt-Oram syndrome and paroxysmal atrial fibrillation. *Circ Res* **102**, 1433-1442 (2008).

19. Zhou, M., Liao, Y. & Tu, X. The role of transcription factors in atrial fibrillation. *J Thorac Dis* **7**, 152-158 (2015).

20. Zhou, Y.M., Zheng, P.X., Yang, Y.Q., Ge, Z.M. & Kang, W.Q. A novel PITX2c lossoffunction mutation underlies lone atrial fibrillation. *Int J Mol Med* **32**, 827-834 (2013).

21. Syeda, F., *et al.* PITX2 Modulates Atrial Membrane Potential and the Antiarrhythmic Effects of Sodium-Channel Blockers. *Journal of the American College of Cardiology* **68**, 1881-1894 (2016).

22. Goentoro, L., Shoval, O., Kirschner, M.W. & Alon, U. The incoherent feedforward loop can provide fold-change detection in gene regulation. *Mol Cell* **36**, 894-899 (2009).

23. Ma, W., Trusina, A., El-Samad, H., Lim, W.A. & Tang, C. Defining network topologies that can achieve biochemical adaptation. *Cell* **138**, 760-773 (2009).

24. Mangan, S. & Alon, U. Structure and function of the feed-forward loop network motif. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 11980-11985 (2003).

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

26. Nadeau, M., *et al.* An endocardial pathway involving Tbx5, Gata4, and Nos3 required for atrial septum formation. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 19356-19361 (2010).

27. Ang, Y.S., *et al.* Disease Model of GATA4 Mutation Reveals Transcription Factor Cooperativity in Human Cardiogenesis. *Cell* **167**, 1734-1749 e1722 (2016).

28. Garg, V., *et al.* GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. *Nature* **424**, 443-447 (2003).

29. Maitra, M., *et al.* Interaction of Gata4 and Gata6 with Tbx5 is critical for normal cardiac development. *Developmental biology* **326**, 368-377 (2009).

30. He, A., Kong, S.W., Ma, Q. & Pu, W.T. Co-occupancy by multiple cardiac transcription factors identifies transcriptional enhancers active in heart. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 5632-5637 (2011).

31. Alon, U. Network motifs: theory and experimental approaches. *Nat Rev Genet* **8**, 450-461 (2007).

32. Luo, X., Yang, B. & Nattel, S. MicroRNAs and atrial fibrillation: mechanisms and translational potential. *Nature reviews. Cardiology* **12**, 80-90 (2015).

33. Wang, J., *et al.* Pitx2-microRNA pathway that delimits sinoatrial node development and inhibits predisposition to atrial fibrillation. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 9181-9186 (2014).

34. Cusanovich, D.A., Pavlovic, B., Pritchard, J.K. & Gilad, Y. The functional consequences of variation in transcription factor binding. *PLoS Genet* **10**, e1004226 (2014).

35. Engreitz, J.M., *et al.* Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* **539**, 452-455 (2016).

36. Werner, M.S., *et al.* Chromatin-enriched lncRNAs can act as cell-type specific activators of proximal gene transcription. *Nature structural & molecular biology* **24**, 596-603 (2017).

37. Mouse, E.C., *et al.* An encyclopedia of mouse DNA elements (Mouse ENCODE). *Genome Biol* **13**, 418 (2012).