GENOMIC REGULATORY ELEMENTS IN TRANSCRIPTION, DEVELOPMENT, AND DISEASE: GENERATING MOUSE MODELS FOR LATERALITY DEFECTS USING CRISPR/CAS9 GENOME ENGINEERING AT THE PITX2 LOCUS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Frances L. Chen

August 2018

GENOMIC REGULATORY ELEMENTS IN TRANSCRIPTION, DEVELOPMENT, AND DISEASE: GENERATING MOUSE MODELS FOR PITX2 RELATED DISEASE USING CRISPR/CAS9 GENOME ENGINEERING

Frances L. Chen, Ph.D.

Cornell University 2018

Doctoral Abstract

The genome contains the code of life: conservation of DNA sequence ensures proper stereotypical patterning and precise formation of the body's tissues replicated in members of the same species, while variation in DNA sequence contributes to unique individual and species' differences in physical form, function, and susceptibility to disease. Genome wide association studies (GWAS) have shown that most sequence variation linked to disease is located not in protein coding genes but in *noncoding* sequence. The noncoding genome is known to play important roles in regulating chromatin structure and gene expression, and understanding its function will translate to improved understanding of genetic disease and health outcomes. In this dissertation, I have used CRISPR/Cas9 genome editing to perturb noncoding regulatory elements at the critical developmental *Pitx2* locus to investigate their function in organogenesis and human disease.

The transcription factor *Pitx2* patterns the left-right (LR) embryonic axis and regulates LR asymmetric organogenesis. Loss of left-specific *Pitx2* expression is

linked to life threatening congenital defects of the heart and intestines. Importantly, mutations affecting noncoding sequence upstream of *Pitx2* while leaving the *Pitx2* gene intact are associated with *Pitx2*-mediated disease, suggesting that cis-regulatory elements may regulate *Pitx2* expression. Here we have investigated differences in chromatin structure and transcription underlying LR asymmetric organogenesis at the *Pitx2* locus. We use transcriptional profiling to identify a novel conserved long noncoding RNA, *Playrr*, that is expressed asymmetric and complementary to *Pitx2* and participates in mutually antagonistic transcriptional regulation with *Pitx2*.

In addition to its essential roles in LR organ development, *Pitx2* and its genomic locus have been linked to atrial fibrillation (AF), the most common arrhythmia worldwide in humans. Despite multiple GWAS identifying noncoding variants at the *Pitx2* locus in association with the most significant genetic contribution to AF risk, the function of these variants and mechanisms by which they may mediate cis-regulation of *Pitx2* and AF pathophysiology remain unknown. Here, using CRISPR/Cas9 genome editing in mice to target the *Playrr* RNA transcript, I demonstrate that *Playrr* mutant mice have evidence of pacemaker dysfunction and are predisposed to AF, mirroring arrhythmia phenotypes found in *Pitx2* loss-of-function mutants and Pitx2 heterozygous mice, suggesting a relationship between *Playrr*, *Pitx2*, and AF.

In conclusion, my work uncovers a role for the lncRNA *Playrr* in arrhythmia and provides multiple novel noncoding mouse models to investigate functions of the cis-regulatory genome and the chromatin level mechanisms in the context of *Pitx2*-related disease.

BIOGRAPHICAL SKETCH

Frances Chen was born in March 1990 to Lucia and Joseph Chen in San Jose, CA where she spent her childhood in the Bay Area, practicing in her shower how to withstand cold water and reading about dolphin species in order to become a marine biologist. After moving to Colorado in 1999, Frances gave up on her dreams to study ocean mammals and started raising her first guide dog puppy for the blind, which sparked her interest in behavior and inspired her to pursue a career in veterinary medicine. Six guide dog puppies later, Frances graduated from Niwot High School and moved to Ithaca, NY to begin what would eventually be a twelve year tenure at Cornell University, where she began her undergraduate and research training. In 2010 she initiated an independent undergraduate research project under the supervision of Dr. Donna Cassidy-Hanley performing deletion mapping in the binucleate genome of *T. thermophila* as an undergraduate in the lab of Dr. Ted Clark, eventually completing her senior honors thesis with this work. After accepting her admission into the College of Veterinary Medicine, Frances spent the summer of 2012 working in the lab of Dr. Natasza Kurpios, where she became enraptured by the comparative species and genomics approaches used to study organ morphogenesis. This initial experience with the embryo prompted her to apply for and enter the Combined DVM/PhD program as a DVM/PhD candidate in the Kurpios lab in 2013 in order to pursue a clinician-scientist career in translational biomedical research. In 2015 thanks to Dr. Kurpios' mentorship, Frances was awarded an NIH F30 fellowship. In Jan 2018 Frances returned to the DVM curriculum. After obtaining her PhD, Frances will graduate with her DVM in 2020, with plans to build upon her goals for clinician-scientist training by pursuing postdoctoral positions in behavioral/biomedical genetics that leverages comparative genomics and functional genetics approaches and acquiring quantitative and computational skillsets.

ACKNOWLEDGMENTS

First and foremost, I thank Dr. Natasza Kurpios for taking on the ambitious task of advising an equally ambitious (and often foolishly so) combined degree DVM/PhD student; if not for her initial investment in my potential and overwhelmingly enthusiastic support I would not have fallen in love with studying the genome via the Dao of the embryo and made the decision to pursue the path of a clinician scientist. I will never stop admiring her boundless enthusiasm and ability in fervently spreading her passion for science, and her support in my training has always opened doors for me. For that, I am forever grateful.

I would like to acknowledge the members of my Graduate Committee, Drs. John Schimenti, Paul Soloway, and Lisa Fortier for their kindness, patience and support of my graduate training. I thank them and their former/current lab members, including but not limited to: Dr. Ewelina Bolcun-Filas, Rob Monroe, Chris Abratte, Nithya Kartha; David Taylor and Roman Spektor; and Jennifer Cassano for scientific discussions, personable professional guidance, and invaluable advice. I would especially like to thank Dr. John Schimenti again, as well as Drs. James Martin for their co-sponsorship and co-advising on my F30 fellowship. I would like to thank Dr. Drew Noden for providing his embryological expertise and providing invaluable teaching and feedback on my experiments. Lastly, I would like to thank the faculty mentors who have always opened their doors and ears to me for feedback and advising: Tracy Stokol, John Parker, David Lin, Bruce Kornreich, Sabine Mann, Charles Danko, and Gunther Hollopeter. Thank you so much for making time for me.

I would like to acknowledge all past and present members of the Kurpios lab for playing an integral role in shaping my PhD experience and for pushing me to places of better understanding and awareness. To Dr. Eva Oxford, for her invaluable mentorship. Thank you Aravind, for your scientifically philosophical camaraderie and to both you and Aparna for your compassion. Thank you to my lab 妹妹 Shing Hu, for her objectivity and her friendship. To former veterinary leadership students: Hanna Telama, Lucy Watson, Alice Watson, Michelle Teunissen, and Becky Dubowitz; and to my past and current undergraduate mentees: Jacqueline Pino, Brandon Mui, Sophie Kupiec-Weglinski, Christina Cong, and Sienna Perry- thank you for your unique contributions and for teaching me as much as I've hopefully taught you.

I would also like to thank Dr. Helene Marquis for her unwavering commitment to the Cornell Combined Degree Program as well as her wholehearted support of each of the individual DVM/PhD students, myself included. Because of her willingness to listen as well as to lead, I was able to appreciate and make the most of the professional opportunities granted to me. I would like to thank all of my fellow Cornell Combined Degree colleagues and friends past and present for being an endless source of inspiration and support- especially to Erin Chu, Jess Brown, Michelle White, Erica Lachenauer, Eileen Troconis Gonzales, and Kristina Ceres for being your impressive, individual selves and for creating a community I am so proud to be a part of.

Finally, I would like to thank the friends/family who have been and done everything: my parents Lucia and Joseph Chen; Alyssa Wetterau, Rachel Evanowski, and Jean and Cal Snow; Alina Goh, Sharon Liu, Cindy Xue, and Tony Won; Sophie Liu; Susan Lee and Moseh Cho; and Timothy Lin. Without you, I would not be here.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Preface

In my dissertation work I set out to leverage CRISPR/Cas9 genome engineering technology to take a loss-of-function, reverse genetics approach to 1) determine the function of *noncoding* elements at the *Pitx2* locus and 2) create novel noncoding mutant mouse models to understand the genomic basis of left-right (LR) organ pathophysiology related to Pitx2, the critical transcription factor driving the evolutionarily conserved LR asymmetry and protecting from Atrial Fibrillation (AF), the most common human arrhythmia. My major findings are organized into two main chapters that follow the Introduction to my dissertation (Chapter 1). Chapter 2 represents a co-authored published manuscript in which we discovered differences in chromatin architecture of the *Pitx2* locus mirroring asymmetric transcription in the context of LR gut morphogenesis and uncovered a long noncoding RNA (lncRNA) we named *Playrr* (*Pitx2* locus-asymmetric regulated RNA. Chapter 3 represents an independent first-author *manuscript in preparation* where I investigate the role of *Playrr* in regulating cardiac Pitx2 expression and demonstrate a biological role for this lncRNA in susceptibility to cardiac arrhythmias. Finally, in Chapter 4, I conclude by highlighting outstanding questions and future studies necessary to further understand Playrr function and describe additional CRISPR/Cas9 edited alleles I have generated at the *Pitx2* locus as important in vivo models for investigating the genomic mechanisms for human arrhythmias.

1.2 Background

1.2.1 The genetic basis of disease: GWAS maps majority of disease association to noncoding variation in genome

The historical central dogma of molecular biology (Crick, 1970) outlines the sequential flow of genetic information to biological function as DNA to RNA to protein, and this paradigm has dominated the way in which the field of genetics has experimentally approached the relationship between DNA sequence/genes to phenotypes and function. However, thanks to the advent of massively parallel sequencing technologies and the collaborative efforts of large scale sequencing consortia, we have exponentially magnified our view into the genome and gained access to the billions of base pairs of linear DNA sequence that constitute the instructions for life. This has revolutionized biomedical research and allowed for genome wide, systems level approaches to answering biological questions at the molecular level. Strikingly, one of the most surprising findings that emerged from this base pair resolution of the genome was that less than 2% of this sequence constitutes protein coding genes and that the majority of genome consists of non-protein-coding DNA sequence elements and a multitude of transcribed RNA species whose functions are largely unknown. Furthermore, through genome wide association studies (GWAS), which have mapped more than 1,000 loci to physiological traits relevant to health and disease (Marigorta et al, 2018), the majority of these associated variants in the genome occur in these noncoding regions of the genome. Therefore, in order to gain

understanding of the genetic basis of disease that will translate to preventative and therapeutic interventions for human health, it is critical to functionally annotate and demonstrate the function of these noncoding regulatory sequences in vivo.

The overarching goal of my dissertation is to perform functional studies of specific noncoding regulatory elements at the *Pitx2* locus in the contexts of LR intestinal looping morphogenesis (Chapter 2, Welsh et al., 2015) and heart development and arrhythmias (Chapter 3, Chen et al., manuscript in preparation).

1.2.2 The relationship between genome structure and gene expression: hierarchical chromatin organization, cis-regulatory elements, and transcriptional regulation

1.2.2.1 The linear genome is hierarchically organized into a 3D chromatin structure: spatial control of gene expression

As with all biological phenomena, function cannot be considered without understanding the constraints and utility provided by structure. In order to investigate the mechanisms by which DNA regulatory sequences control transcription, it is necessary to understand their relationship in the context of chromatin structure. The linear DNA double helix is organized via wrapping around an octameric core of histone proteins into the fundamental building block of chromatin, the nucleosome (Kornberg, 1974; Arents et al., 1991; Luger et al., 1997). Compaction of nucleosomes serves to fold and condense chromatin necessary for chromosome packaging and transcriptional repression while nucleosome remodeling allows for opening and transcriptional activation. Additionally, starting from early electron microscopy

studies in the nucleus (Heitz, 1928), it has been known that chromatin can be roughly organized into two structural and correlated functional types: 1) heterochromatin, tightly coiled chromatin generally associated with transcriptional repression and close association with the nuclear lamina and 2) euchromatin, loosely coiled chromatin accessible to transcriptional machinery and associated with transcriptional activation (Gasser, 2002; Pollard, 2002; Lanctot, 2007). However, these initial observations were limited in their resolution to probe the precise mechanisms relating DNA sequence, 3D chromatin structure, and transcriptional regulation. Nevertheless, recent technologies, including DNA fluorescent in situ hybridization (DNA-FISH) and Chromosome Conformation Capture (3C) based methods, have allowed for new and important insights as to how the hierarchical structure of chromatin in 3D space may allow for gene regulation via multiple levels of chromatin structure (Bridger and Volpi, 2010; Dekker et al., 2002). Chapter 2 represents our published work that explores chromatin level gene regulation at the *Pitx2* locus through these higher resolution methods, including DNA-FISH and incorporation of Hi-C and ChIA-PET data sets, to be described below.

1.2.2.2 Long Range chromatin looping constrained by chromatin architecture: Topologically Associated Domains (TADs) and sub-TADs.

At the level of linear DNA sequence, proper gene expression requires that gene promoters and their regulatory sequences along (in 'cis') and between chromosomes (in 'trans') communicate with each other across scales of kilobases to hundreds of kilobases via long range chromatin looping (Carter et al., 2002; Tolhuis et al., 2002).

Methods such as high resolution microscopy with fluorescent DNA labeling (DNA FISH) (Bridger et al., 2010) and imaging of chromatin and chromatin conformation capture (3C) technologies (Dekker et al., 2002) have not only provided snapshots of these specific long range physical interactions across many genetic loci but have also painted a more global picture of the genome as a dynamic regulatory landscape where the linear DNA sequence is hierarchically organized from dynamically interacting chromatin loops to larger chromatin domains (hundreds of kilobases to megabase scale). By chemically crosslinking and then sequencing chromatin, 3C and related assays generate a statistical average/likelihood of dynamic DNA contacts across a locus (3C, 4C, 5C) or across the genome (Hi-C; ChIA-PET) (Dekker et al., 2002; Rusk, 2009). These data allow for the representation of the genome as a global interactive topographical map subdivided by experimentally defined units of preferentially self-interacting regions of DNA called topologically associated domains (TADs) on the hundreds of kilobases to megabase scale (Dixon et al., 2012; Valton and Decker, 2016). TADs can be further subdivided into so-called sub-TADs (on the hundreds of kilobases to megabase scale). TADs have been shown to be invariant across cell types and seem to represent constitutive features of chromatin organization (Dixon et al., 2015; Dixon et al., 2012; Nora et al., 2012). In contrast, sub-TADs appear to have cell type specific organization and are correlated with cell type specific regulatory events (Dixon et al., 2016; Dowen et al., 2014; Rao et al., 2014). This compartmentalization of the 3D genome represents a useful framework to investigate the relationships between chromatin structure, transcriptional domains and the cis and trans factors that interact with this hierarchical topology.

TADs can be correlated with transcriptional machinery occupancy and gene expression to infer relationships between chromatin interaction patterns and gene regulation (Dekker et al., 2015; Ulianov et al., 2016). Indeed, genes within the same TAD have been shown to show correlated patterns of expression across differentiation (Nora et al., 2012). Therefore, understanding TAD structure represents one platform to probe mechanisms of gene regulation within chromatin structure.

The boundaries of TADs are demarcated and formed by historically so-called 'insulators,' regions that limit enhancer activity and that in most studied cases have shown to be mediated by cohesion/CTCF complexes (Wendt et al, 2008; Dixon et al., 2012; Busslinger et al., 2017). This insulation of enhancer activity and contact with promoters represents a key mechanism by which the structure of the genome regulates gene expression. Not only have TAD boundaries been shown to confine transcriptional activity (Wendt et al, 2008) but in a few specific cases have even been shown to prevent pathological activation/dysregulation of gene expression (Lupiáñez et al, 2015; Valton and Dekker, 2016). In fact, disrupting TAD structure has been shown to dysregulate gene expression and recapitulate disease phenotypes (Lupiáñez et al, 2015; Franke et al, 2016). For instance, genome engineering at the *Epha4* locus producing three types of large structural chromosomal rearrangements mirroring known mutations found in humans with congenital limb abnormalities recapitulated similar limb phenotypes in mouse (Lupiáñez et al, 2015). Significantly, it was demonstrated that these pathological chromosomal rearrangements resulted in misexpression and ectopic contact of noncoding enhancers and gene promoters only if the structural variant disrupted a CTCF-associated TAD boundary (Lupiáñez et al,

2015). In another locus specific study, structural variation in copy number at the Sox9 locus had varying consequences on gene misexpression and phenotype depending on if duplications of noncoding DNA occurred within a TAD (intra-TAD), across TAD boundaries, and/or formed a new TAD (Franke et al, 2016).

Investigating chromatin structure from the TAD perspective of chromatin architecture thereby bears relevancy for examining the mechanisms by which noncoding regulatory elements function in transcriptional regulation and understanding of human disease. In Chapter 2, we demonstrate that chromatin architecture at the *Pitx2* locus is dependent on CTCF and relate long range interactions across the *Pitx2* locus with TAD structure.

1.2.3 Emergence of a field: functional roles for long noncoding RNAs

Following the striking finding that less than two percent of the genome produced protein coding transcripts yet three quarters of the human genome is transcribed (Carnici et al., 2005; Djebali et al., 2012), attention has now turned to characterizing the nature and function of the thousands of noncoding RNA transcripts uncovered by transcriptomic analyses of mouse and human cells (Carnici et al., 2005; Guttman et al., 2009; Djebali et al., 2012). One novel class of these RNA species uncovered by transcriptomic approaches are long noncoding RNAs, originally classified arbitrarily as any transcript greater than 200 nucleotides to distinguish them from previously characterized shorter noncoding RNAs.

Over the past two decades, research on noncoding RNA species such as microRNAs and piRNAs have produced a paradigm shift in molecular biology by

demonstrating the versatility of RNA molecules over proteins as widespread and critical gene regulators that can provide precise and powerful modulatory control (Varani, 2015). Consequently, there was much excitement surrounding the discovery of this class of RNAs and the potential that long noncoding RNAs could offer greater understanding of gene regulation. However, our basic understanding of this novel class of RNA species remains poor; to this day only a handful have been functionally annotated and a valid counterargument to the promise of functional potential of these RNA species is that most of them represent 'transcriptional noise' (Struhl, 2007). Therefore, significant goals of the field are to determine which lncRNAs play significant biological roles, assign functions to these uncharacterized transcripts in cellular and organismal physiology, and uncover the non-RNA and RNA-mediated mechanisms by which lncRNA loci act. In Chapter 3, I detail my work uncovering an in vivo biological function for the *Pitx2* locus derived lncRNA, *Playrr*, in the context of cardiac arrhythmia, representing a significant contribution to this nascent field.

1.2.3.1 Classification of lncRNAs

LncRNAs are defined as autonomously transcribed noncoding RNA transcripts greater than 200 nucleotides in length (Morris and Mattick, 2014; Ranschoff et al., 2018). Further characterization has revealed that lncRNAs constitute a broad class of heterogeneous transcripts that can be classified according to evolutionary conservation based on sequence, structure, or synteny (position relative to other genes), as well as the DNA sequence elements and genomic locations they are transcribed from (enhancers, antisense promoters, intronic, intergenic, etc) (Kim et al., 2015; Wu et al., 2017). There have been several classification schemes proposed as to how these

features may correlate with or provide potential clues for guiding experimental approaches to assess biological function (Mercer et al., 2009; Wu et al., 2017; Kopp et al., 2018).

1.2.3.2 Known biological functions of lncRNAs

A few case examples have provided functional paradigms for this diverse class of noncoding RNAs. Although it is difficult to infer generalizable properties from locus specific studies to the remainder of the thousands of uncharacterized lncRNAs, these model loci have nonetheless validated important biological roles for lncRNAs and provided valuable insights for investigating RNA-mediated mechanisms of gene regulation.

The most classic and well studied example is that of *Xist* (X inactive specific transcript), the long noncoding RNA that is the master epigenetic regulator of X chromosome inactivation (XCI), the phenomenon by which eutherian mammals achieve X chromosome dosage compensation between *XX* (females) and *XY* (males) (Penny et al., 1996). The mechanisms by which *Xist* achieves epigenetic silencing of the entire X chromosome in females include: a) *Xist* binding to discrete X chromosome locations and spreading across the entire chromosome, b) relationship with chromatin architecture and topology, c) *Xist* interaction with and recruit of Polycomb proteins, an extensively studied chromatin remodeler responsible for epigenetic silencing, and d) interaction with nuclear scaffolding (Cerase et al., 2015). Importantly, the example of *Xist* provides key experimental paradigms on which to study the epigenetic function of long noncoding RNA transcripts at multiple regulatory levels via genetic, biochemical, and structural approaches. Shortly

following the discovery of *Xist,* additional functional studies at model loci have linked lncRNAs to a similarly wide range of gene regulatory processes involving allele specific gene silencing (*H19* at the imprinted *Igf2* locus) (Kallen et al., 2013; Gabory et al., 2010) recruitment of Polycomb proteins, (*H19, Kcnq1ot1*) (Zhao et al., 2010) formation of repressive chromatin loops (*COLDAIR* and *COLDWRAP* at Flowering Locus C (FLC) (Heo et al., 2011; Kim et al., 2017), and direct transcriptional activation and repression and scaffolding with chromatin modification complexes (*HOTAIR* at the *HoxC* gene cluster) (Rinn et al., 2007; Tsai et al., 2010).

In addition to loss and gain-of-function genetic studies *in vitro* demonstrating functional roles for lncRNAs at the cellular level, several groups have now performed genetic analyses in mouse models to demonstrate that some lncRNAs play critical and indispensable roles in lineage specific differentiation and embryonic development. Notable examples include *Braveheart,* a lncRNA transcript whose depletion in ES cells resulted in failure to activate cardiac lineage gene expression (Klattenhoff et al., 2013) and *Fendrr*, a lateral plate mesoderm expressed lncRNA essential for cardiac and body wall development (Grote et al., 2013).

1.2.3.3 Overcoming challenges in lncRNA functional studies

Despite this growing amount of evidence supporting functional roles for a few key lncRNAs, many other lncRNA transcripts have been shown to be dispensable when perturbed both in vitro and in vivo, or with only minor effects on transcription of genes in cis (Paralkar et al., 2016; Amândio, et al., 2016). Additionally, several properties of lncRNAs in general have made them difficult to experimentally approach compared to protein coding transcripts, as much of the toolkit used to investigate their

nature and function are only recently developed technologies and/or have been adapted from studying protein coding genes (Leone and Santoro, 2016). For instance, detection in tissue is more challenging as lncRNAs are generally found to be highly tissue specific, lowly expressed compared to mRNAs, and predominantly confined to the nucleus (Derrien et al., 2012; Calibi et al., 2011). Similarly, traditional reverse genetic approaches used for protein coding genes applied to lncRNA loci often do not dissect out functional RNAs from their underlying DNA cis-regulatory elements or even the act of transcription itself (Goff and Rinn, 2015). Indeed, a key challenge to overcome when designing functional studies to ascertain lncRNA biological roles is how to best unlink these RNA transcripts from their DNA sequence elements. LncRNAs are often closely associated with and often even arise from DNA cis-regulatory elements such as enhancers (Kim et al., 2010; Kim et al., 2015). Genetic perturbations that disrupt both the DNA cis-regulatory element (as in a deletion of a genomic locus) as well as the transcription of the lncRNA fail to separate functional roles of the DNA sequence as opposed to the RNA product and data from these types of mutations must be interpreted in light of this fact.

Fortunately, the timely advent of CRISPR/Cas9 and other genome editing technologies have proven an efficient approach in unlinking lncRNA transcripts from their cis-regulatory elements- either by inserting a premature polyadenylation cassette (Gutschner et al., 2011; Latos et al., 2012; Yin et al., 2015; Paralkar et al., 2016) or disruption of splice sites (Engreitz et al., 2016), an approach we have utilized to separate *Playrr* from its underlying DNA element e926 (Chapter 2). Importantly, through precise genome editing, many have demonstrated the RNA transcript itself

may be dispensable while transcriptional regulation of nearby genes is mediated via the DNA cis-regulatory element, as was the case with the lncRNA, *Lockd*, and the adjacent gene promoter, *Cdkn1* (Paralkar et al., 2016) as well as the lncRNA *Haunt* and the HoxA cluster (Yin et al., 2015). Importantly, the nature of the genetic mutation needs to be carefully considered when evaluating correlative phenotypes and expression effects.

Lastly, elucidating biological relevance *in vivo* needs to be carefully interpreted. An illustrative case in the field is that of *HOTAIR*, a lncRNA arising from the *HOXC* genomic locus (Rinn et al., 2007). Seminal work from the lab of Howard Chang discovered *HOTAIR* and demonstrated that knockdown of *HOTAIR* was associated with upregulation of *HOXD* genes in *trans* (Rinn et al., 2007). Subsequently, this group reported recapitulation of this finding *in vivo* with a targeted deletion of the orthologous locus in mouse and reported both -*trans* deprepression of *Hoxd* genes along with homeotic transformations consistent with *Hoxd* genes' known roles in axial skeleton patterning (Li et al, 2013). However, a reanalysis with this same mouse model in the lab of Denis Duboule found no *trans* regulation of the *HoxD* cluster though the authors observed minor cis-effects on transcription of HoxC (Amandio et al., 2016). Importantly, critical review and comparison of the contradictory findings in both phenotype and expression were ultimately attributed to differences in mouse strain background and tissue types used in transcriptome profiling (Li et al., 2013; Amandio et al., 2016; Selleri 2016). The resolution of this initially controversial debate over *HOTAIR* highlighted the importance of careful

interpretation of lncRNA models and the pursuit of functional studies using *in vivo* models with robust and penetrant phenotypes (Selleri 2016).

Nevertheless, despite the experimental challenges in functionally validating RNA transcripts, there is a growing body of work in the field demonstrating the exciting and diverse biological roles for lncRNAs at the cellular and organismal levels while also providing accompanying lessons to improve upon previous studies. In my dissertation work we have leveraged the efficiency of CRISPR/Cas9 genome editing to engineer novel noncoding mutations at the *Pitx2* locus and investigate a biological role for the lncRNA *Playrr*. Chapter 2 details the creation of two CRISPR/Cas9 genome edited alleles that effectively allowed us to parse out the function of *Playrr* from its associated DNA element e926: 1) *Playrr* a deletion of the DNA cis-regulatory element e926 and the *Playrr* transcriptional start site and 2) a splice site mutation at the first exon-intron junction of *Playrr*. In Chapter 3, I utilize our splice site mutation mouse model to specifically characterize the role of *Playrr* in vivo in the context of heart development and cardiac arrhythmias.

1.2.4 The embryo as a platform for studying cis-regulatory transcriptional regulation

Development is a beautifully dynamic yet highly stereotypical process by which a single cell divides, differentiates, migrates and undergoes complex coordinated behaviors to generate the entirety of diverse functional tissues of the embryo. This process of morphogenesis is orchestrated by the genome and as such development represents an extraordinarily important model to study the transcriptional regulation of

gene regulatory networks that have cascading cellular, tissue, and whole organismal level readouts. Importantly, many studies have demonstrated that transcriptional regulation in development, from patterning to specification and differentiation, requires the integrated, coordinate input of multiple regulatory elements at a given genomic locus (Montavon et al, 2011; Marinic et al 2013). Indeed, several developmental loci have been critical in pioneering the field of transcriptional regulation by providing models for understanding genomic regulation of transcription in an in vivo developmental context. At the *HoxD* locus, a switch of regulatory interactions between *Hoxd* cluster genes and topological domains in which regulatory elements in a centromeric and telomeric gene deserts flanking the HoxD cluster mediate proximal-distal patterning of the vertebrate limb (Montavon et al., 2011; Andrey et al., 2013). Subsequent work investigating chromatin structure at the HoxD locus using 3C based assays, DNA FISH, and high resolution imaging have not only provided chromatin level mechanisms of precise spatiotemporal expression of Hoxd cluster genes at this locus, but also demonstrate the power of relating chromatin structure and transcriptional regulation in a developmental context (Andrey et al., 2013; Fabre et al., 2015; Fabre et al., 2017). Furthermore, studies at the *Fgf8* locus have revealed how enhancer-promoter specificity is achieved when multiple interdependent regulatory elements are interspersed with unrelated genes (Marinic et al., 2013). Together these studies have demonstrated that the functional transcriptional output resulting from the combinatorial integration of cis-regulatory information depends upon the chromatin context (Marinic et al, 2013).

These pioneering models have proven invaluable for understanding the relationships between chromatin structure, noncoding regulatory elements, and biological function. Nevertheless, further diversity of loci need to be studied to begin to make any generalizable inferences about the mechanisms of gene regulation. In my dissertation work I demonstrate the use of the *Pitx2* developmental locus to investigate tissue specific expression in the context of intestinal morphogenesis (Chapter 2) as well as cardiac development and disease (Chapter 3). My findings lend evidence to suggest that precise spatiotemporal expression during development is accomplished by the coordinate activity of multiple distinct regulatory elements whose function as a coherent unit depends on their relative positioning within locus structure rather than autonomous, intrinsic activity of independent regulatory elements.

1.2.5 *Pitx2* in development and disease

In addition to providing an excellent platform for studying genomic mechanisms of transcriptional regulation, investigating the control of master regulators such as *Pitx2* is highly relevant for human health and disease. Pitx2 is a key developmental transcription factor with complex gene expression involving multiple isoforms and therefore pleiotropic roles in development and postnatal life. Asymmetrically expressed *Pitx2c* isoform is responsible for left-right (LR) patterning and specification of left-sided cellular identity during the development of the heart, lungs, spleen, and intestines. In contrast, bilaterally expressed *Pitx2a* and *Pitx2b* play distinct and overlapping roles in specifying craniofacial structures and pituitary cell types (Lin et al., 1999; Liu et al., 2003; Hamada et al., 2014). Perturbation of proper *Pitx2*

expression therefore affects multiple organ systems. Loss of asymmetric *Pitx2c* disrupts LR patterning and can result in congenital heart disease, intestinal malrotation and volvulus, and other organ defects with life-threatening consequences in newborns (Hamada et al., 2014; Ryan et al., 1993; Davis et al., 2008; Kurpios et al., 2008; Welsh et al., 2013).

Mutations in *Pitx2* were first identified as causal links to Axenfeld-Reiger syndrome (ARS), an autosomal dominant disorder characterized by mental retardation, ocular and dental abnormalities, and umbilical hernias (Semina et al., 1996). Screening of ARS patients has identified individuals who possess no mutations in *Pitx2* coding sequences but harbor large deletions that encompass an adjacent gene desert devoid of coding genes, suggesting that regulatory elements in the noncoding gene desert are necessary and/or sufficient for modulating precise spatiotemporal Pitx2 expression (Volkmann et al., 2011).

Finally, recent studies have provided a clear link between loss of *Pitx2*c expression in the aging heart and predisposition to atrial fibrillation (AF), the most common sustained cardiac arrhythmia in otherwise healthy adults (Martin, 2015) that predisposes patients to risk of stroke, heart attack, and death. Importantly, *Pitx2* is not only essential for LR heart morphogenesis during development but also has maintained postnatal levels within the adult left atrium (Shiratori et al, 2006; Ai et al, 2006; Wang et al, 2010). Loss of *Pitx2* and alterations in *Pitx2* dosages in mice and humans is associated with congenital cardiac defects, increased susceptibility to AF, and development of bilateral sinoatrial nodes (SAN) (Kitamura et al, 1999; Wang et al, 2010; Wang et al, 2014; Ammirabile et al, 2011). The SAN is the cardiac

pacemaker that dictates heart rhythm under normal physiological conditions. Its dysregulation in sinus node dysfunction (SND), another clinically significant atrial arrhythmia, is an important comorbidity found in AF patients where both conditions can predispose and exacerbate each other (Chang et al., 2013). There are several major pathophysiological mechanisms that contribute to AF development, including genetic predisposition, and understanding of genetic pathways underlying both AF and SND can lead to intervention targets for both of these important atrial arrhythmias. However, despite several GWAS implicating noncoding genetic variants at the *Pitx2* locus as the most significantly associated with AF (Gudbjartsson, et al., 2007; Sinner et al., 2011, Ellinor et al., 2012; Christopherson et al., 2017) the links between noncoding elements at the *Pitx2* locus, *Pitx2* dosage, and AF remains unclear. In Chapter 3, I explore the hypothesis that noncoding elements at the *Pitx2* locus play a role in transcriptional regulation of *Pitx2* and cardiac arrhythmias by investigating the relationships between the noncoding lncRNA *Playrr,* transcriptional modulation of *Pitx2* dosage, and AF in our *Playrr*^{Ex1sj} mouse model.

With so many roles in development and disease, studying precise spatiotemporal *Pitx2* expression by investigating the role of cis-regulatory elements at the locus will allow for a clearer understanding of multiple human diseases. Thus, the major goal of my dissertation is to leverage CRISPR/Cas9 genome engineering technology to determine the function of noncoding elements at the *Pitx2* locus in the context of the developing GI tract, the heart, and in AF. In doing so, I hope to shed light on the role of noncoding elements in transcriptional regulation of *Pitx2* and

simultaneously create noncoding mutant mouse models for clinically relevant Pitx2-related disease.

1.3 Conclusion

The transcriptional program implemented by a regulatory landscape represents the combinatorial inputs of DNA sequence information, cis-regulatory enhancers, and potentially RNA mediated mechanisms with chromatin architectural constraints. Ultimately, my dissertation research touches upon the functional aspects of all of these elements involved in transcriptional regulation, and excitingly, implicates a pathophysiological role for the long noncoding RNA at the *Pitx2* locus in susceptibility to human cardiac arrhythmias.

REFERENCES

Ai, D., Liu, W., Ma, L., Dong, F., Lu, M. F., Wang, D., … Martin, J. F. (2006). Pitx2 regulates cardiac left-right asymmetry by patterning second cardiac lineage-derived myocardium. *Developmental Biology*, *296*(2), 437–449. https://doi.org/10.1016/j.ydbio.2006.06.009

Amândio, A. R., Necsulea, A., Joye, E., Mascrez, B., & Duboule, D. (2016). Hotair Is Dispensible for Mouse Development. *PLOS Genetics*, *12*(12), e1006232. https://doi.org/10.1371/journal.pgen.1006232

Ammirabile, G., Tessari, A., Pignataro, V., Szumska, D., Sutera Sardo, F., Benes, J., … Campione, M. (2012). Pitx2 confers left morphological, molecular, and functional identity to the sinus venosus myocardium. *Cardiovascular Research*, *93*(2), 291–301. https://doi.org/10.1093/cvr/cvr314

Andrey, G., Montavon, T., Mascrez, B., Gonzalez, F., Noordermeer, D., Leleu, M., … Duboule, D. (2013). A switch between topological domains underlies HoxD genes collinearity in mouse limbs. *Science (New York, N.Y.)*, *340*(6137), 1234167. https://doi.org/10.1126/science.1234167

Arents, G., Burlingame, R. W., Wang, B. C., Love, W. E., & Moudrianakis, E. N. (1991). The nucleosomal core histone octamer at 3.1 A resolution: a tripartite protein assembly and a left-handed superhelix. *Proceedings of the National Academy of Sciences of the United States of America*, *88*(22), 10148–10152. https://doi.org/10.1073/PNAS.88.22.10148

Bridger JM, Volpi EV. Fluorescence in situ hybridization (FISH): protocols and applications. Totowa: Humana Press; 2010

Busslinger, G. A., Stocsits, R. R., Van Der Lelij, P., Axelsson, E., Tedeschi, A., Galjart, N., & Peters, J. M. (2017). Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. *Nature*, *544*(7651), 503–507. https://doi.org/10.1038/nature22063

Cabili, M. N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., & Rinn, J. L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes & Development*, *25*(18), 1915–1927. https://doi.org/10.1101/gad.17446611

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M. C., Maeda, N., … Nishiguchi, S. (2005). The transcriptional landscape of the mammalian genome. *Science*, *309*(5740), 1559–1563. https://doi.org/10.1126/science.1112014

Carter, D., Chakalova, L., Osborne, C. S., Dai, Y. feng, & Fraser, P. (2002). Long-range chromatin regulatory interactions in vivo. *Nature Genetics*, *32*(4), 623–626. https://doi.org/10.1038/ng1051

Cerase, A., Pintacuda, G., Tattermusch, A., & Avner, P. (2015). Xist localization and function: new insights from multiple levels. *Genome Biology*, *16*(1), 166. https://doi.org/10.1186/s13059-015-0733-y

Chang, H.-Y., Lin, Y.-J., Lo, L.-W., Chang, S.-L., Hu, Y.-F., Li, C.-H., … Chen, S.-A. (2013). Sinus node dysfunction in atrial fibrillation patients: the evidence of regional atrial substrate remodelling. *EP Europace*, *15*(2), 205–211. https://doi.org/10.1093/europace/eus219

Core, L. J., Waterfall, J. J., & Lis, J. T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science (New York, N.Y.)*, *322*(5909), 1845–1848. https://doi.org/10.1126/science.1162228

Crick, F. (1970). Central dogma of molecular biology. *Nature*, *227*(5258), 561–563. https://doi.org/10.1038/227561a0

Davis, N. M., Kurpios, N. A., Sun, X., Gros, J., Martin, J. F., & Tabin, C. J. (2008). The Chirality of Gut Rotation Derives from Left-Right Asymmetric Changes in the Architecture of the Dorsal Mesentery. *Developmental Cell*. https://doi.org/10.1016/j.devcel.2008.05.001

Dekker, J., Rippe, K., Dekker, M., & Kleckner, N. (2002). Capturing chromosome conformation. *Science (New York, N.Y.)*, *295*(5558), 1306–1311. https://doi.org/10.1126/science.1067799

Dekker, J., & Heard, E. (2015). Structural and functional diversity of Topologically Associating Domains. *FEBS Letters*, *589*(20PartA), 2877–2884. https://doi.org/10.1016/j.febslet.2015.08.044

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., … Guigó, R. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Research*, *22*(9), 1775–1789. https://doi.org/10.1101/gr.132159.111

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., … Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, *485*(7398), 376–380. https://doi.org/10.1038/nature11082

Dixon, J. R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J. E., Lee, A. Y., differentiation. *Nature*, *518*(7539), 331–336. https://doi.org/10.1038/nature14222

Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., … Gingeras, T. R. (2012). Landscape of transcription in human cells. *Nature*, *489*(7414), 101–108. https://doi.org/10.1038/nature11233

Ellinor PT, Lunetta KL, Albert CM, Glazer NL, Ritchie MD, Smith AV, Arking DE, Muller-Nurasyid M, Krijthe BP, Lubitz SA, Bis JC, Chung MK, Dorr M, Ozaki K, Roberts JD, Smith JG, Pfeufer A, Sinner MF, Lohman K, Ding J, Smith NL, Smith JD, Rienstra M, Rice KM, Van Wagoner DR, Magnani JW, Wakili R, Clauss S, Rotter JI, Steinbeck G, Launer LJ, Davies RW, Borkovich M, Harris TB, Lin H, Volker U, Volzke H, Milan DJ, Hofman A, Boerwinkle E, Chen LY, Soliman EZ, Voight BF, Li G, Chakravarti A, Kubo M, Tedrow UB, Rose LM, Ridker PM, Conen D, Tsunoda T, Furukawa T, Sotoodehnia N, Xu S, Kamatani N, Levy D, Nakamura Y, Parvez B, Mahida S, Furie KL, Rosand J, Muhammad R, Psaty BM, Meitinger T, Perz S, Wichmann HE, Witteman JC, Kao WH, Kathiresan S, Roden DM, Uitterlinden AG, Rivadeneira F, McKnight B, Sjogren M, Newman AB, Liu Y, Gollob MH, Melander O, Tanaka T, Stricker BH, Felix SB, Alonso A, Darbar D, Barnard J, Chasman DI, Heckbert SR, Benjamin EJ, Gudnason V & Kaab S (2012). Meta-analysis identifies six new susceptibility loci for atrial fibrillation. Nat Genet 44, 670–675.

ENCODE Project Consortium, T. E. P. (2011). A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biology*, *9*(4), e1001046. https://doi.org/10.1371/journal.pbio.1001046

Engreitz, J. M., Haines, J. E., Perez, E. M., Munson, G., Chen, J., Kane, M., … Lander, E. S. (2016). Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature*, *539*(7629), 452–455. https://doi.org/10.1038/nature20149

Espinoza, H.M., Cox, C.J., Semina, E.V., and Amendt, B.A. (2002) A molecular basis for differential developmental anomalies in Axenfeld–Rieger syndrome. *Hum. Mol. Genet.* **11**, 743–753.

Fabre, P. J., Benke, A., Joye, E., Nguyen Huynh, T. H., Manley, S., & Duboule, D. (2015). Nanoscale spatial organization of the HoxD gene cluster in distinct transcriptional states. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(45), 13964–13969. https://doi.org/10.1073/pnas.1517972112

Fabre, P. J., Leleu, M., Mormann, B. H., Lopez-Delisle, L., Noordermeer, D., Beccari, L., & Duboule, D. (2017). Large scale genomic reorganization of topological domains at the HoxD locus. *Genome Biology*, *18*(1), 149. https://doi.org/10.1186/s13059-017-1278-z

Gasser, S. M. Visualizing chromatin dynamics in interphase nuclei. *Science* 296, 1412–1416 (2002).

Gibcus, J. H., & Dekker, J. (2013). The Hierarchy of the 3D Genome. *Molecular Cell*, *49*(5), 773–782. https://doi.org/10.1016/J.MOLCEL.2013.02.011

Grote, P., Wittler, L., Hendrix, D., Koch, F., Währisch, S., Beisaw, A., … Herrmann, B. G. (2013). The Tissue-Specific lncRNA Fendrr Is an Essential Regulator of Heart and Body Wall Development in the Mouse. *Developmental Cell*, *24*(2), 206–214. https://doi.org/10.1016/J.DEVCEL.2012.12.012

Gudbjartsson DF, Arnar DO, Helgadottir A, Gretarsdottir S, Holm H, Sigurdsson A, Jonasdottir A, Baker A, Thorleifsson G, Kristjansson K, Palsson A, Blondal T, Sulem P, Backman VM, Hardarson GA, Palsdottir E, Helgason A, Sigurjonsdottir R, Sverrisson JT, Kostulas K, Ng MC, Baum L, So WY, Wong KS, Chan JC, Furie KL, Greenberg SM, Sale M, Kelly P, MacRae CA, Smith EE, Rosand J, Hillert J, Ma RC, Ellinor PT, Thorgeirsson G, Gulcher JR, Kong A, Thorsteinsdottir U & Stefansson K (2007). Variants conferring risk of atrial fibrillation on chromosome 4q25. Nature 448, 353–357.

Gutschner, T., Baas, M., & Diederichs, S. (2011). Noncoding RNA gene silencing through genomic integration of RNA destabilizing elements using zinc finger nucleases. *Genome Research*, *21*(11), 1944–1954. https://doi.org/10.1101/gr.122358.111

Guttman, M., Amit, I., Garber, M., French, C., Lin, M. F., Feldser, D., … Lander, E. S. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*, *458*(7235), 223–227. https://doi.org/10.1038/nature07672

Hamada, H., and Tam, P.P. (2014) Mechanisms of left-right asymmetry and patterning: driver, mediator and responder. *F1000Prime Rep.* **6**, 1-9.

Heitz E. Das Heterochromatin der Moose.I. Jahrb Wiss Bot. 1928;69:762–818.

Heo, J. B., & Sung, S. (2011). Vernalization-Mediated Epigenetic Silencing by a Long Intronic Noncoding RNA Vernalization-Mediated Epigenetic Silencing by a Long Intronic Noncoding RNA. *Science*, *331*(6013), 76–79. Retrieved from http://science.sciencemag.org/

Kaab S, Darbar D, van Noord C, Dupuis J, Pfeufer A, Newton-Cheh C, Schnabel R, Makino S, Sinner MF, Kannankeril PJ, Beckmann BM, Choudry S, Donahue BS, Heeringa J, Perz S, Lunetta KL, Larson MG, Levy D, MacRae CA, Ruskin JN, Wacker A, Schomig A, Wichmann HE, Steinbeck G, Meitinger T, Uitterlinden AG, Witteman JC, Roden DM, Benjamin EJ & Ellinor PT (2009). Large scale replication and meta-analysis of variants on chromosome 4q25 associated with atrial fibrillation. Eur Heart J 30, 813–819.

Kallen, A. N., Zhou, X.-B., Xu, J., Qiao, C., Ma, J., Yan, L., … Huang, Y. (2013). The Imprinted H19 LncRNA Antagonizes Let-7 MicroRNAs. *Molecular Cell*, *52*(1), 101–112. https://doi.org/10.1016/j.molcel.2013.08.027

Kim, D.-H., & Sung, S. (2017). Vernalization-Triggered Intragenic Chromatin Loop Formation by Long Noncoding RNAs. *Developmental Cell*, *40*(3), 302–312.e4. https://doi.org/10.1016/j.devcel.2016.12.021

Kim, T.-K., Hemberg, M., Gray, J. M., Costa, A. M., Bear, D. M., Wu, J., … Greenberg, M. E. (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature*, *465*(7295), 182–187. https://doi.org/10.1038/nature09033

Kim, T.-K., Hemberg, M., & Gray, J. M. (2015). Enhancer RNAs: a class of long noncoding RNAs synthesized at enhancers. *Cold Spring Harbor Perspectives in Biology*, *7*(1), a018622. https://doi.org/10.1101/cshperspect.a018622

Kitamura, K., Miura, H., Miyagawa-Tomita, S., Yanazawa, M., Katoh-Fukui, Y., Suzuki, R., … Yokoyama, M. (1999). Pitx2 deficiency and abnormal morphogenesis. *Development*, *126*, 5749–5758. Retrieved from http://dev.biologists.org/content/develop/126/24/5749.full.pdf

Klattenhoff, C. A., Scheuermann, J. C., Surface, L. E., Bradley, R. K., Fields, P. A., Steinhauser, M. L., … Boyer, L. A. (2013). Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. *Cell*, *152*(3), 570–583. https://doi.org/10.1016/j.cell.2013.01.003

Kollmar, M., Kollmar, L., Hammesfahr, B., & Simm, D. (2015). diArk--the database for eukaryotic genome and transcriptome assemblies in 2014. *Nucleic Acids Research*, *43*(Database issue), D1107-12. https://doi.org/10.1093/nar/gku990

Kopp, F., & Mendell, J. T. (2018). Functional Classification and Experimental Dissection of Long Noncoding RNAs. *Cell*, *172*(3), 393–407. https://doi.org/10.1016/j.cell.2018.01.011

Kornberg, R. D., & Lorch, Y. (1999). Twenty-Five Years of the Nucleosome, Fundamental Particle of the Eukaryote Chromosome. *Cell*, *98*(3), 285–294. https://doi.org/10.1016/S0092-8674(00)81958-3

Krijger, P. H. L., & de Laat, W. (2016). Regulation of disease-associated gene expression in the 3D genome. *Nature Reviews Molecular Cell Biology*, *17*(12), 771–782. https://doi.org/10.1038/nrm.2016.138

Kurpios, N. A., Ibañes, M., Davis, N. M., Lui, W., Katz, T., Martin, J. F., … Tabin, C. J. (2008). The direction of gut looping is established by changes in the extracellular

matrix and in cell:cell adhesion. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.0803578105

Latos, P. A., Pauler, F. M., Koerner, M. V., Senergin, H. B., Hudson, Q. J., Stocsits, R. R., … Barlow, D. P. (2012). Airn Transcriptional Overlap, But Not Its lncRNA Products, Induces Imprinted Igf2r Silencing. *Science*, *338*(6113), 1469–1472. https://doi.org/10.1126/science.1228110

Li, L., Liu, B., Wapinski, O. L., Tsai, M.-C., Qu, K., Zhang, J., … Chang, H. Y. (2013). Targeted Disruption of Hotair Leads to Homeotic Transformation and Gene Derepression. *Cell Reports*, *5*(1), 3–12. https://doi.org/10.1016/j.celrep.2013.09.003

Lin, C.R., Kioussi, C., O'Connell, S., Briata, P., Szeto, D., Liu, F., J. Izpisúa-Belmonte, C., and Rosenfeld, M.G. (1999) Pitx2 regulates lung asymmetry, cardiac positioning and pituitary and tooth morphogenesis. *Nature* **401**, 279-282.

Liu, W., Selever, J., Lu, M.F., and Martin, J.F. (2003) Genetic dissection of Pitx2 in craniofacial development uncovers new functions in branchial arch morphogenesis, late aspects of tooth morphogenesis and cell migration. *Development* **130**, 6375–6385.

Logan, M., Pagán-Westphal, S. M., Smith, D. M., Paganessi, L., & Tabin, C. J. (1998). The transcription factor pitx2 mediates situs-specific morphogenesis in response to left-right asymmetric signals. *Cell*, *94*(3), 307–317. https://doi.org/10.1016/S0092-8674(00)81474-9

Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, *389*(6648), 251–260. https://doi.org/10.1038/38444

Mahadevan, A., Welsh, I. C., Sivakumar, A., Gludish, D. W., Shilvock, A. R., Noden, D. M., … Kurpios, N. A. (2014). The left-right Pitx2 pathway drives organ-specific arterial and lymphatic development in the intestine. *Developmental Cell*. https://doi.org/10.1016/j.devcel.2014.11.002

Marigorta, U. M., Rodríguez, J. A., Gibson, G., & Navarro, A. (2018). Replicability and Prediction: Lessons and Challenges from GWAS. *Trends in Genetics : TIG*, *34*(7), 504–517. https://doi.org/10.1016/j.tig.2018.03.005

Marinić, M., Aktas, T., Ruf, S., & Spitz, F. (2013). An integrated holo-enhancer unit defines tissue and gene specificity of the Fgf8 regulatory landscape. *Developmental Cell*, *24*(5), 530–542. https://doi.org/10.1016/j.devcel.2013.01.025

Montavon, T., Soshnikova, N., Mascrez, B., Joye, E., Thevenet, L., Splinter, E., … Duboule, D. (2011). A regulatory archipelago controls Hox genes transcription in digits. *Cell*, *147*(5), 1132–1145. https://doi.org/10.1016/j.cell.2011.10.023

Morris, K. V., & Mattick, J. S. (2014, June 29). The rise of regulatory RNA. *Nature Reviews Genetics*. Nature Publishing Group. https://doi.org/10.1038/nrg3722

Paralkar, V. R., Taborda, C. C., Huang, P., Yao, Y., Kossenkov, A. V., Prasad, R., … Weiss, M. J. (2016). Unlinking an lncRNA from Its Associated cis Element. *Molecular Cell*, *62*(1), 104–110. https://doi.org/10.1016/J.MOLCEL.2016.02.029

Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S., & Brockdorff, N. (1996). Requirement for Xist in X chromosome inactivation. *Nature*, *379*(6561), 131–137. https://doi.org/10.1038/379131a0

Pollard, T. D. & Earnshaw, W. C. *Cell Biology* (Saunders, Philadelphia, 2002).

Ransohoff, J. D., Wei, Y., & Khavari, P. A. (2018). The functions and unique features of long intergenic non-coding RNA. *Nature Reviews. Molecular Cell Biology*, *19*(3), 143–157. https://doi.org/10.1038/nrm.2017.104

Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Brugmann, S. A., … Chang, H. Y. (2007). Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Noncoding RNAs. *Cell*, *129*(7), 1311–1323. https://doi.org/10.1016/J.CELL.2007.05.022

Rusk, N. (2009). When ChIA PETs meet Hi-C. *Nature Methods*, *6*(12). https://doi.org/10.1038/nmeth1209-863

Ryan, A.K., Blumberg, B., Rodriguez-Esteban, C., Yonei-Tamura, S., Tamura, K., Tsukui, T., de la Peña, J., Sabbagh, W., Greenwald, J., Choe, S., Norris, D.P., Robertson, E.J., Evans, R.M., Rosenfeld, M.G., and Izpisúa Belmonte, J.C. (1998) Pitx2 determines left–right asymmetry of internal organs in vertebrates. *Nature* **394**, 545–551.

Schnabel RB, Yin X, Gona P, Larson MG, Beiser AS, McManus DD, Newton-Cheh C, Lubitz SA, Magnani JW, Ellinor PT, Seshadri S, Wolf PA, Vasan RS, Benjamin EJ & Levy D (2015). 50 year trends in atrial fibrillation prevalence, incidence, risk factors, and mortality in the Framingham Heart Study: a cohort study. Lancet 386, 154–162.

Selleri, L., Bartolomei, M. S., Bickmore, W. A., He, L., Stubbs, L., Reik, W., & Barsh, G. S. (2016). A Hox-Embedded Long Noncoding RNA: Is It All Hot Air? *PLOS Genetics*, *12*(12), e1006485. https://doi.org/10.1371/journal.pgen.1006485

Semina, E.V., Reiter, R., Leysens, N.J., Alward, W.L., Small, K.W., Datson, N.A., Siegel-Bartelt, J., Bierke-Nelson, D., Bitoun, P., Zabel, B.U., Carey, J.C., and Murray, J.C. (1996) Cloning and characterization of a novel bicoid-related homeobox

transcription factor gene, RIEG, involved in Rieger syndrome. *Nat. Genet.* **14**, 392–399.

Struhl, K. (2007). Transcriptional noise and the fidelity of initiation by RNA polymerase II. *Nature Structural & Molecular Biology*, *14*(2), 103–105. https://doi.org/10.1038/nsmb0207-103

Tao, Y., Zhang, M., Li, L., Bai, Y., Zhou, Y., Moon, A. M., … Martin, J. F. (2014). Pitx2, an atrial fibrillation predisposition gene, directly regulates ion transport and intercalated disc genes. *Circulation. Cardiovascular Genetics*, *7*(1), 23–32. https://doi.org/10.1161/CIRCGENETICS.113.000259

Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F., & De Laat, W. (2002). Looping and interaction between hypersensitive sites in the active β-globin locus. *Molecular Cell*, *10*(6), 1453–1465. https://doi.org/10.1016/S1097-2765(02)00781-5

Tsai, M.-C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J. K., Lan, F., … Chang, H. Y. (2010). Long Noncoding RNA as Modular Scaffold of Histone Modification Complexes. *Science*, *329*(5992), 689–693. https://doi.org/10.1126/science.1192002

Ulianov, S. V, Khrameeva, E. E., Gavrilov, A. A., Flyamer, I. M., Kos, P., Mikhaleva, E. A., … Razin, S. V. (2016). Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. *Genome Research*, *26*(1), 70–84. https://doi.org/10.1101/gr.196006.115

Volkmann, B.A., Zinkevich, N.S., Mustonen, A., Schilter, K.F., Bosenko, D.V., Reis, L.M., Broeckel, U., Link, B.A., and Semina, E.V. (2011) Potential novel mechanism for Axenfeld-Rieger syndrome: deletion of a distant region containing regulatory elements of PITX2. *Invest. Ophthalmol. Vis. Sci.* **52**, 1450–1459.

Wang, J., Klysik, E., Sood, S., Johnson, R. L., Wehrens, X. H. T., & Martin, J. F. (2010). Pitx2 prevents susceptibility to atrial arrhythmias by inhibiting left-sided pacemaker specification. *Proceedings of the National Academy of Sciences*, *107*(21), 9753–9758. https://doi.org/10.1073/pnas.0912585107

Welsh, I. C., Thomsen, M., Gludish, D. W., Alfonso-Parra, C., Bai, Y., Martin, J. F., & Kurpios, N. A. (2013). Integration of left-right Pitx2 transcription and Wnt signaling drives asymmetric gut morphogenesis via Daam2. *Developmental Cell*. https://doi.org/10.1016/j.devcel.2013.07.019

Welsh, I. C., Kwak, H., Chen, F. L., Werner, M., Shopland, L. S., Danko, C. G., … Kurpios, N. A. (2015). Chromatin Architecture of the Pitx2 Locus Requires CTCF-

and Pitx2-Dependent Asymmetry that Mirrors Embryonic Gut Laterality. *Cell Reports*, *13*(2), 337–349. https://doi.org/10.1016/j.celrep.2015.08.075

Wendt, K. S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., … Peters, J. M. (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature*, *451*(7180), 796–801. https://doi.org/10.1038/nature06634

Wu, H., Yang, L., & Chen, L. L. (2017). The Diversity of Long Noncoding RNAs and Their Generation. *Trends in Genetics*. https://doi.org/10.1016/j.tig.2017.05.004

Yin, Y., Yan, P., Lu, J., Song, G., Zhu, Y., Li, Z., … Shen, X. (2015). Opposing Roles for the lncRNA Haunt and Its Genomic Locus in Regulating HOXA Gene Activation during Embryonic Stem Cell Differentiation. *Cell Stem Cell*, *16*(5), 504–516. https://doi.org/10.1016/j.stem.2015.03.007

Zhao, J., Ohsumi, T. K., Kung, J. T., Ogawa, Y., Grau, D. J., Sarma, K., … Lee, J. T. (2010). Genome-wide Identification of Polycomb-Associated RNAs by RIP-seq. *Molecular Cell*, *40*(6), 939–953. https://doi.org/10.1016/j.molcel.2010.12.01

CHAPTER 2

EMBRYONIC GUT LATERALITY IS MIRRORED BY ASYMMETRIC CHROMATIN ARCHITECTURE AT THE PITX2 LOCUS DEPENDENT ON PITX2 AND CTCF

Authors: Ian C. Welsh, Hojoong Kwak, Frances L. Chen, Melissa Werner, Lindsay S. Shopland, Charles G. Danko, John T. Lis, Min Zhang, James F. Martin and Natasza A. Kurpios

AUTHOR CONTRIBUTIONS

I.C.W. and N.A.K. designed the study. I.C.W. performed all the experiments with additional contributions from L.S.S. for FISH guidance and data analysis; H.K for performing GRO-seq with guidance from J.T.L; C.G.D. for dREG analysis with guidance from H.K.; *F.L.C. for double label RNA in situ hybridization (Fig. 1B) and CRISPR deletion of e926 (Fig. 3AC);* M.Z, and J.F.M for HiC and ChIA-PET computation analysis and quantifications (Fig. 6B). I.C.W. and N.A.K. wrote the manuscript with input from all coauthors.

This chapter was published in *Cell Reports*:

Welsh, I. C., Kwak, H., Chen, F. L., Werner, M., Shopland, L. S., Danko, C. G., … Kurpios, N. A. (2015). Chromatin Architecture of the Pitx2 Locus Requires CTCF- and Pitx2-Dependent Asymmetry that Mirrors Embryonic Gut Laterality. *Cell Reports*, *13*(2), 337–349. https://doi.org/10.1016/j.celrep.2015.08.075

2.1 Preface

This chapter represents my significant contributions to a co-authored publication in *Cell Reports*. Additional follow up studies I independently performed on the biological role of the lncRNA *Playrr* in the context of *Pitx2* transcriptional regulation in gut morphogenesis are detailed in Chapter 4 and Appendix A.

2.2 SUMMARY

Three-dimensional (3D) chromatin organization is fundamental for cell type-specific gene expression. The transcription factor Pitx2 is expressed on the left side of the early vertebrate embryo to pattern left-right (L-R) organs including the dorsal mesentery (DM), whose asymmetric cell behavior directs gut looping chirality. However, despite the critical importance of organ laterality, chromatin-level regulation of Pitx2 remains undefined. Here we show that genes immediately neighboring Pitx2 on chicken chromosome 4 are expressed strictly on the right side of the DM, opposite left-sided Pitx2. Pitx2 represses right-sided genes, including a conserved long noncoding RNA we have named Playrr (Pitx2 locus asymmetric regulated RNA). Using CRISPR/CAS9 genome editing of Playrr, 3D fluorescent in situ hybridization (FISH) and variations of chromatin conformation capture (3C), we demonstrate that mutual antagonism between Pitx2 and Playrr expression during LR organogenesis is coordinated by asymmetric chromatin interactions dependent on Pitx2 and the CCCTC-binding factor CTCF. Collectively, we demonstrate that transcriptional and morphological asymmetries driving gut looping are mirrored by chromatin architectural asymmetries at the Pitx2 locus. We propose a model where Pitx2 auto-

regulation directs chromatin topology to coordinate L- R transcription of this locus essential for asymmetric organogenesis.

2.3 INTRODUCTION

The external bilateral symmetry of vertebrates conceals the highly conserved left-right (L-R) asymmetries of the internal organs essential for their normal function and efficient packing within the body cavity. L-R patterning initiates early during gastrulation via transient signaling of the transforming growth factor β-related protein Nodal, which results in persistent expression of the homeobox transcription factor Pitx2 throughout the left lateral mesoderm (Logan et al. 1998; Shiratori et al. 2001). Subsequently, restricted Pitx2 expression specifies the left identity of individual organ primordia (Ryan et al. 1998; Zorn and Wells 2009). The control of laterality by Pitx2 represents a remarkably ancient function, as this pathway is required for normal asymmetric morphogenesis even in basal deuterostomes such as sea urchin and non-bilaterians such as hydra (Duboc et al. 2005; Watanabe et al. 2014).

Tissue-specific gene expression provides the basis for genetic control of morphogenesis and is central to human health and disease. The activity of intercellular signaling and transcription factors is integrated by cis-regulatory elements encoded in the genome to coordinate spatial gene expression and direct morphogenesis. However, genes and regulatory elements that must physically interact to drive tissue-specific expression are commonly distributed across large genomic intervals (Zuniga et al. 2004; Marinić et al. 2013; Lettice et al. 2002; Anderson et al. 2014). While it is now appreciated that genes cluster in the nucleus and form 3D domains of considerable

(500 kb – 1 Mb) chromosomal sequence, how genes are regulated within this structural context is unknown (Shopland et al. 2006). Importantly, most information on gene regulation at the level of 3D nuclear structure has been obtained from cells in culture (Sutherland and Bickmore 2009). Evaluating 3D chromatin structure in situ, in the 3D tissues in which cells differentiate, is important for understanding the role of chromatin 3D structure in tissue-specific gene regulation.

Indicative of the inherent complexity, a number of strategies to translate the 3D organization of the genome into cell type specific gene expression have been enumerated (reviewed in de Laat and Duboule 2013). Highlighting outstanding questions regarding the source of specificity of transcriptional control in vertebrates, such distributed regulatory landscapes can also influence "by-stander" genes within the genomic interval, while at other loci regulation is strictly independent (Spitz et al. 2003; Cajiao et al. 2004; Anderson et al. 2014; Marinić et al. 2013). While these few examples are instrumental, additional experimentally tractable models are needed to dissect, in the context of tissues and organs that a locus controls, the mechanisms through which 3D organization of the linear genome coordinates spatial gene expression.

Regulation of Pitx2 in vertebrates is complex, involving multiple isoforms with unique and overlapping spatial expression and function. Pitx2a and Pitx2b are generated by alternative splicing and are expressed bilaterally, while Pitx2c is derived from an alternative promoter, and is exclusively left-sided (Shiratori et al. 2006; Liu et al. 2001). Pitx2 null mice exhibit mid gestation lethality and global organ laterality

defects (Lu et al. 1999). Moreover, mutations in the human PITX2 gene are associated with Axenfeld- Rieger Syndrome (ARS) characterized by mental retardation, mandibular and ocular birth defects, as well as ventral body wall defects and umbilical hernias (Semina et al. 1996). Importantly, screening of ARS patients has identified individuals who possess no mutations in Pitx2 coding sequences but who harbor large lesions within an adjacent gene desert devoid of coding genes, suggesting an essential cis-regulatory role for elements within the desert in driving Pitx2 expression (Flomen et al. 1998; Volkmann et al. 2011; Reis et al. 2012). In addition to ARS, single-nucleotide polymorphisms (SNPs) most significantly associated with PITX2-linked atrial fibrillation (AF), the most prevalent form of sustained cardiac arrhythmia in the human population, map to the gene desert proximal to PITX2 (Lubitz et al. 2010; Kent et al. 2002). Pitx2 heterozygous mice are predisposed to AF indicating that altered levels of Pitx2 lead to AF, further suggesting that regulatory elements within the gene desert can influence Pitx2 expression. Despite the wealth of available knowledge about early signaling events controlling laterality, the specific genomic mechanisms governing the expression of its master effector remain unexplored.

To this end, we make use of the dorsal mesentery (DM), a bridge of mesodermal tissue suspending the gut that possesses a unique binary (L vs. R) molecular and cellular asymmetry directed by left-sided Pitx2 (Fig. 1A, top). The left compartment of the DM condenses while the right expands causing the DM to deform and tilt the gut tube leftward (Davis et al. 2008; Kurpios et al. 2008). This tilt provides a bias for asymmetric gut rotation, disruption of which randomizes gut looping (Davis et al. 2008) (Fig. 1A, top). To define the molecular composition of the DM, we completed a

laser capture microdissection and microarray analysis of the left (Pitx2-positive) and right (Pitx2- negative) halves of the chicken DM at the critical time when L-R DM asymmetries are apparent (HH21, akin to mouse embryonic [E] day 10.5) (Hamburger and Hamilton 1951; Welsh et al. 2013). We report here our further analyses of these data, which strikingly reveal that genes immediately neighboring Pitx2 and positioned either proximally and distally to a large conserved gene desert flanking Pitx2 are expressed in a right-specific pattern opposite to left-specific Pitx2. Moreover, positioned at the proximal end of the gene desert we identified the conserved sequence element e926, with left- specific enhancer activity in the DM of transgenic mouse embryos but right-specific activity in the heart. In contrast, our genome-wide global run-on sequencing, GRO-seq and dREG (discriminative Regulatory Element detection from GRO-seq) (Core et al. 2008; Danko et al. 2014) analysis of left and right DM samples found that e926 functions endogenously as a promoter for Playrr, a conserved long noncoding RNA (lncRNA) exclusively transcribed in the right DM. We employed fluorescent in situ hybridization (FISH) to correlate this binary L-R expression of locus genes with nuclear architecture and discovered highly evolutionary conserved long-range chromatin looping in the left and right DM. Importantly, within this invariant 3D topology, we identified subtle but conserved L-R differences in nuclear proximity of e926/Playrr and Pitx2 that are dependent on Pitx2. Moreover, both HiC and Chromatin Interaction Analysis by Paired- End Tag Sequencing (ChIA-PET) confirmed preferential long-range interactions between e926/Playrr and Pitx2 in vitro. Collectively, these data represent the first report of chromatin-level asymmetry during L-R organogenesis and suggest a model where

tissue- specific cis-regulatory topology establishes L-R transcription among higher vertebrates.

2.4 RESULTS

2.4.1 Right-sided expression of genes at the Pitx2 locus is opposite to Pitx2 on the left

We previously performed laser capture microdissection and microarray analysis of the left and right chicken DM when L-R DM asymmetries are first observed (Welsh et al., 2013). Our initial objective was to identify cellular effectors in the left DM that which may therefore represent Pitx2 targets responsible for the cellular behavior in the left DM. However, most recently, focusing on the right side of the DM, we identified that glutamyl aminopeptidase A (Enpep) is the most differentially expressed gene in the right DM, with ~17-fold higher expression (Fig. 1A, graph). Not surprisingly, we showed that Pitx2 is the most differentially expressed gene on the left side of the DM, with ~19-fold higher expression in the left vs. right DM consistent with the central role of this gene in regulating asymmetric organogenesis (Welsh et al. 2013) (Fig. 1A, graph). Remarkably, we noted that these two genes exhibiting the highest fold differences in left vs. right- sided expression are located immediately adjacent to each other in the genome. On chicken chromosome 4, Pitx2 is flanked proximally by a large (~600kb) gene desert and 27kb distally by the convergently transcribed Enpep (Fig. 1B). We confirmed these microarray findings using RNA in situ hybridization and demonstrated that tissue asymmetry in the DM is mirrored by differences in expression of these linked genes across the L-R axis (Fig. 1B HH21).

Intrigued, we asked whether additional genes neighboring the Pitx2 locus also exhibit asymmetric expression in the DM and identified two additional genes with right- side restricted gene expression: fatty acid elongase, (Elovl6), the immediate neighbor of Enpep, and the ortholog of the uncharacterized human gene C4ORF32, Loc422694 (cC4orf32) positioned at the proximal boundary of the gene desert (Fig. 1A-C). This novel pattern of asymmetric expression was not exclusive to the DM. We observed asymmetric expression of linked genes well prior to DM formation in the lateral mesoderm (precursor to the DM, Fig. 1B HH17), and in other asymmetric organs, such as the heart, where Pitx2 plays an essential role (Fig. 1B, HH12) (Franco and Campione 2003).

Furthermore, highlighting the importance of the Pitx2 locus early during embryogenesis, its gene content, order, and orientation are highly conserved from humans to frogs, suggesting that functional and regulatory constraints maintain such synteny (Kikuta et al. 2007; Nobrega et al. 2003) (Fig. 1C). Identification of asymmetric regulatory element e926 at the Pitx2 locus Regulation of asymmetric Pitx2c expression requires the activity of the ASE enhancer element located in the last intron of Pitx2 (Shiratori et al., 2006). The functional activity of this element is highly conserved and homologous ASE sequences from fish, frog, chicken, mouse, and human are able to drive left-specific reporter gene activity in mice (Shiratori et al. 2001, 2006).

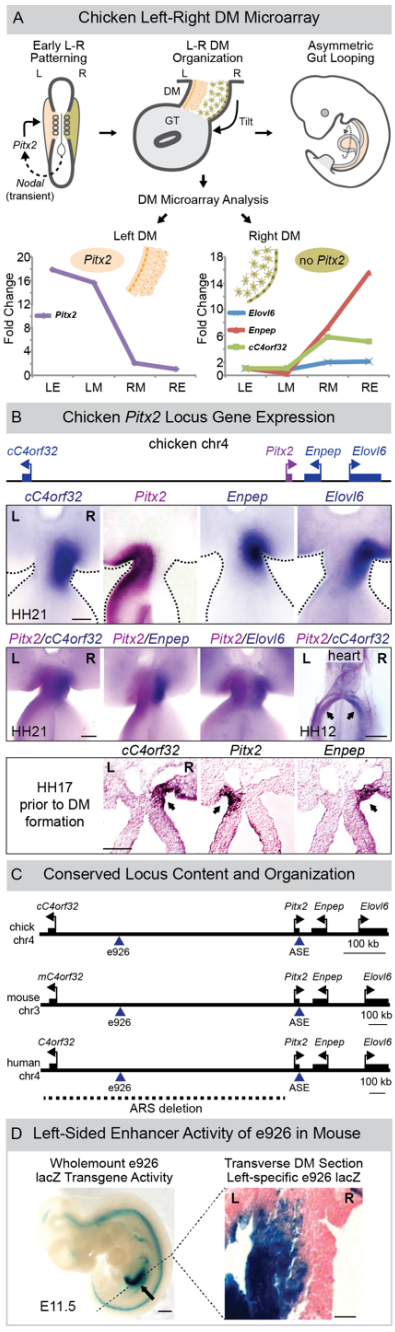


Figure 2-1. L-R asymmetric gene expression at the Pitx2 locus (A) Nodal-induced Pitx2 initiates gut looping directed by L-R changes in the DM cellular architecture (tan, left; green, right). DM microarray (LE, left epithelium; LM, left mesenchyme; RM, right mesenchyme; RE, right epithelium) identifies genes linked to Pitx2 with right side- restricted expression, validated in (B) via whole mount double in situ hybridization (Pitx2, magenta; right-specific genes, blue) at HH21 (DM), at HH17 (lateral mesoderm, DM precursor, arrows) and at HH12 (sinus venosus of the primitive heart, arrows). (C) Pitx2 locus conservation in chicken, mouse, and human (dashed line, human ARS deletion). Note, human locus is shown inverted relative to its orientation on chr4. (D) e926 directs left-specific reporter gene expression (LacZ) in the left DM of transgenic mouse embryos. See also Figure S1.

We hypothesized that additional enhancer elements exist within the conserved gene desert that contribute to the robust L-R expression of genes at the Pitx2 locus. Such deserts, devoid of protein coding genes, commonly harbor regulatory elements (REs) critical for coordinating spatiotemporal expression of nearby genes. Indeed, translocations and deletions within the desert flanking Pitx2 found in ARS patients provide strong evidence for the cis-regulatory role at the Pitx2 locus (Fig. 1C, human) (Rainger et al. 2014; Volkmann et al. 2011; Reis et al. 2012). The Vista Enhancer Database is a genome-wide collection of computationally and experimentally predicted enhancer elements that have been screened for tissue-specific enhancer activity via transgenic assays in E11.5 mouse embryos (Visel et al. 2007). We searched the Vista Enhancer Database and identified hs926, a human derived sequence located in the proximal third of the desert whose sequence is highly conserved among human, mouse, and chicken (Fig. 1C). Element 926 (e926, 1614 base pairs) showed reproducible enhancer like activity in the midgut of E11.5 transgenic reporter embryos (Fig. 1D, n=4/7). Upon sectioning of three e926 transgenic embryos (kindly provided by the Vista Enhancer group), we found that e926 drove robust lacZ reporter activity specifically in the left DM of the midgut and duodenum (Fig. 1D and S1 n=3/3). Interestingly, we also noted right-specific e926 enhancer activity in the heart (Fig. S1, n=2/3). Thus, when assayed via transgenesis, e926 functions as a transcriptional enhancer capable of responding to both left and right asymmetric regulatory environments within the embryo, including the left DM.

2.4.2 Right-specific regulatory activity of e926 in vivo

Recent findings demonstrate that enhancers may act differently when tested in trans as single elements, vs. within the cis-context of the endogenous locus (Marinić et al. 2013; Ruf et al. 2011; West et al. 2002). Hence, to characterize e926 function in the DM, we conducted a genome-wide global run-on sequencing, GRO-seq, of left vs. right derived DM samples (Fig. 2A) (Core et al. 2008). GRO-seq allows mapping of nascent transcription by engaged RNA polymerase II (RNA Pol II) and collects strand-specific reads that identify regions of divergent transcription shown to be a distinct signature of active REs (Core et al. 2008; Melgar et al. 2011; Core et al. 2014). Moreover, a computational tool termed discriminative Regulatory Element detection from GRO-seq (dREG) recognizes the pattern of divergent transcription at active REs allowing characterization of the transcriptional regulatory state at the level of both nascent transcription and activation of associated REs (Fig. 2A, green) (Danko et al. 2014).

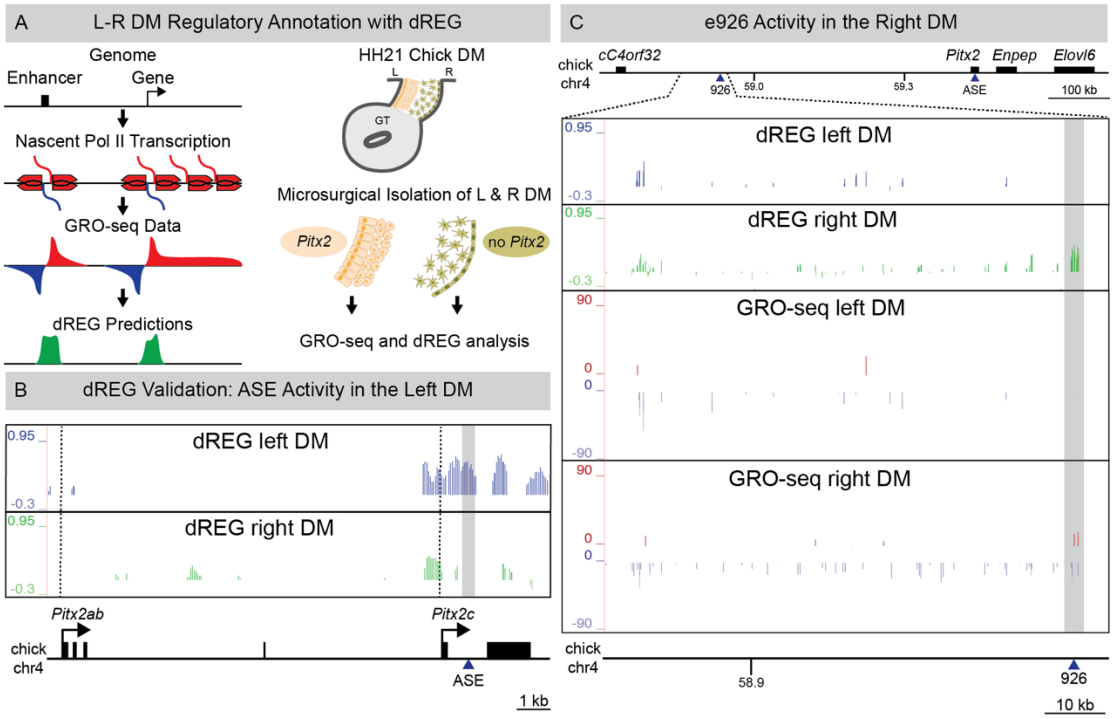


Figure 2-2. L-R DM GRO-seq/dREG analyses (A) Left: Overview of GRO-seq and dREG. Right: GRO-seq libraries were prepared from HH21 L-R DM samples (B) dREG peaks corresponding to ASE are exclusive to the left DM (grey box). Dashed vertical lines mark Pitx2ab and Pitx2c TSS. (C) dREG peaks overlapping e926 are right-specific (grey box) in contrast to left-specific transgenic reporter gene activity driven by e926. GRO-seq detects asymmetric transcription from the minus strand at the proximal gene desert in the right DM. See also Figure S2 and S3.

Consistent with our microarray data, both GRO-seq and dREG analysis confirmed the binary L-R gene expression from the Pitx2 locus, readily apparent as differential transcription at the proximal and distal gene ends of desert in the left or right DM, respectively (Fig. S2). For example, GRO-seq reads mapped to the asymmetric Pitx2c at the distal end demonstrate transcription in the left but not right DM samples, while dREG peaks overlapping the ASE enhancer were exclusively observed in the left DM samples (Fig. 2B, grey box; Fig. S3). In contrast, GRO-seq and dREG detected extensive transcription of the proximal desert preferentially in the right DM (Fig. 2C, Fig. S2) including dREG peaks directly overlapping the conserved e926 enhancer (Fig. 2C, grey box). Although this confirms e926 as an asymmetrically responsive cis-regulatory element, it stands in striking contrast to the transient transgenesis assay suggesting the left-specific activity of e926 in the DM (Fig. 1D).

2.4.3 Identification of a conserved lncRNA, Playrr, transcribed from e926 on the right dREG identifies active REs as sites of divergent transcription occurring at both enhancers and promoters (Danko et al. 2014). Therefore, an alternative explanation for the discrepancy in asymmetric activity of e926 is that this regulatory element does not function as an enhancer but may act instead as a promoter in the right DM. However, there are currently no annotated genes in the chicken genome within the gene desert distal to cC4orf32 and proximal to Pitx2. Interestingly, we noted that GRO-seq reads extend a considerable distance proximally from e926, suggesting the presence of a gene transcribed from the minus strand towards cC4orf32 (Fig. 2C, GRO-seq right DM). Supporting this hypothesis, the syntenic region of the mouse genome is annotated with D030025E07Rik, a long noncoding RNA (lncRNA) with a

transcriptional start site (TSS) contained within e926 that is also transcribed from the minus strand of mouse chromosome 3 (Fig. 3A). To characterize the expression profile of this lncRNA in the chicken and mouse DM we cloned D030025E07Rik and chicken ESTs corresponding to the transcribed region identified through GRO-seq. In both species, expression of this lncRNA was restricted to the right side of the DM consistent with the GRO-seq data (Fig. 3B). These results support the presence of a conserved lncRNA expressed on the right side, contralateral to the left-sided Pitx2 expression. We refer to this novel lncRNA as Playrr (Pitx2 locus asymmetric regulated RNA).

2.4.4 Mutual antagonism of Pitx2 and Playrr expression in vivo

We found conserved Pitx2 binding sites within e926 suggesting that Pitx2 regulates Playrr expression (Fig. S1B). To test the effect of Pitx2 on Playrr expression in vivo, we used a genetic approach and examined Playrr expression in Pitx2-/- mouse DM at E10.5. We found that in the absence of Pitx2 on the left, Playrr is bilaterally expressed in the DM (Fig. 3B). TaqMan based QRT-PCR confirmed the upregulation of Playrr expression in Pitx2-/- embryos but did not detect altered expression of other genes at the locus. This indicates that in the left DM, Pitx2 specifically represses Playrr expression (Fig. S4C).

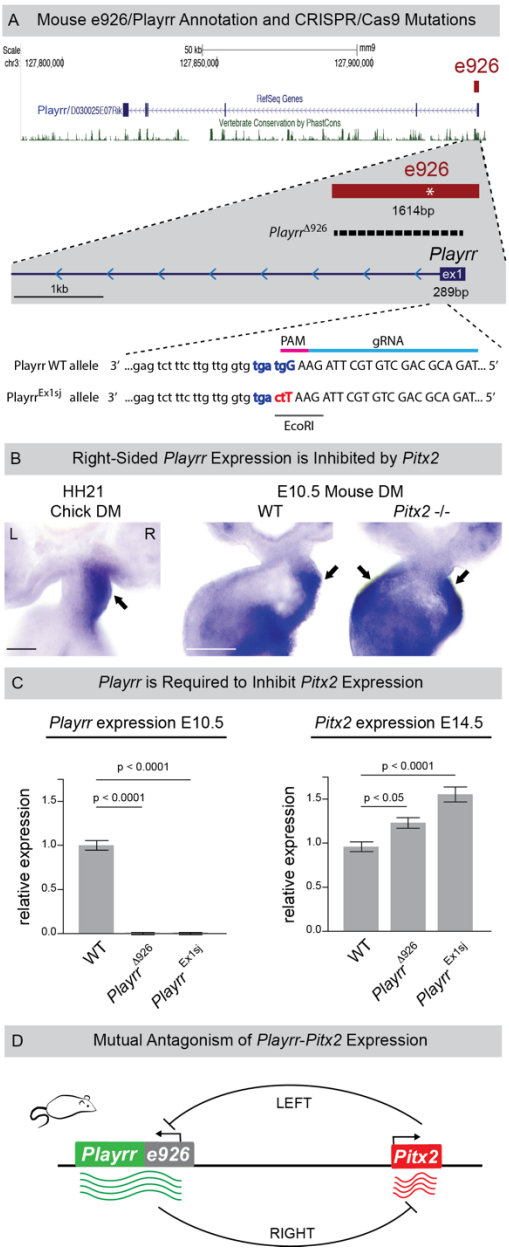


Figure 2-3. Identification of a conserved lncRNA, Playrr, asymmetrically transcribed from e926. (A) Annotation of mouse lncRNA Playrr (D030025E07Rik). e926 contains the TSS and exon 1 of Playrr and a conserved Pitx2 binding motif (asterisk). CRISPR/Cas9 mutation of e926/Playrr editing via deletion of e926 (Playrr$^{\Delta 926}$, 1420bp dashed line) or targeted disruption of Playrr splicing (Playrr^{Ex1sj}) by a 3bp change (red letters) that converts the U1 snRNP splice site recognition sequence (blue letters) to an EcoRI site. Exon 1 (upper case), intron 1 sequence (lower case), relative guide RNA (gRNA, light blue) and PAM motif (pink) were used to produce Playrr^{Ex1sj}. (B) Whole mount in situ hybridization showing right-specific Playrr in WT HH21 chicken and E10.5 mouse embryos vs. bilateral Playrr expression in Pitx2-null mouse DM. (C) TaqMan QRT-PCR expression in Playrr$^{\Delta 926}$ and Playrr^{Ex1sj} mutants relative to WT of Playrr (left) and Pitx2 (right) at E10.5 or E14.5, respectively. (D) Summary of genetic data showing mutual antagonism of Playrr and Pitx2 expression. See also Figure S1 and S4.

We next used CRISPR/Cas9 mediated genome editing to functionally dissect the role of the binary e926/Playrr element on Pitx2 locus expression (Fig. 3A). First, we used a pair of guide RNAs targeting the proximal and distal ends of e926 to simultaneously delete e926 and disrupt transcription of Playrr (Playrr$^{\Delta 926}$). Conversely, to disrupt Playrr with minimal impact on e926, we mutated 3 base pairs of the U1 snRNP recognition sequence (ggtagt) at the 3' end of Playrr exon 1 to disrupt Playrr splicing (Playrr^{Ex1sj}, Fig. 3A), leading to intron retention and exosomal degradation (Almada et al. 2013). QRT-PCR confirmed that both deletion of e926 (Playrr$^{\Delta 926}$) and targeted disruption of Playrr (Playrr^{Ex1sj}) result in loss of detectable levels of Playrr RNA in E10.5 embryos (Fig. 3C). Importantly, Pitx2 expression was found to be significantly upregulated in the visceral organ primordia of E14.5 of both Playrr$^{\Delta 926}$ and Playrr^{Ex1sj} mutant embryos compared to WT littermates (Fig. 3C, E14.5). Thus, our data demonstrate mutual antagonism between Pitx2 and Playrr expression and establish that function of the lncRNA Playrr is required to modulate Pitx2 expression levels (Fig. 3D).

2.4.5 Asymmetric chromatin looping of the Pitx2 locus in the chicken DM

The contralateral expression of Pitx2 and Playrr, their mutual antagonism, and evolutionary conservation of their arrangement in chromosomal sequence, suggest that their regulation may be coordinated by chromosome structure. Thus, to understand how binary L-R transcription relates to nuclear chromatin architecture we employed multi- color 3D DNA fluorescent in situ hybridization on chicken DM sections (3D

tissue- FISH). We hypothesized that looping interactions at the Pitx2 locus organize chromatin differentially in nuclei on the left vs. right. We used DNA probes (15-20 kb) to label cC4orf32 (Cy5, blue) at the proximal boundary of the gene desert, Playrr (DIG, green) within the gene desert, and Pitx2 (Cy3, red) at the distal end of the gene desert (Fig. 4A). Quantification of each pairwise interprobe distance (i.e. cC4orf32-Playrr, cC4orf32- Pitx2, and Playrr-Pitx2) was used to compare how these loci are positioned within the 3D space of the nucleus relative to each other and to the intervening gene desert (Table 1).

In the linear genome, cC4orf32 and Playrr are nearest to each other, separated by ~162kb compared to the ~578kb separating cC4orf32 and Pitx2 or ~416kb separating Playrr and Pitx2 (Fig. 4A). However, our FISH revealed that within the 3D nuclear space of the DM, both cC4orf32 and Playrr were significantly closer to Pitx2 than to each other (Table 1A). This interaction is striking as it brings genes that are separated over a large linear distance (416 kb and 578 kb) into closer proximity than genes separated by just 160 kbs. Our data therefore establish that long-range looping of the Pitx2 locus positions the proximal and distal ends of the gene desert in close proximity within the 3D space of the nucleus.

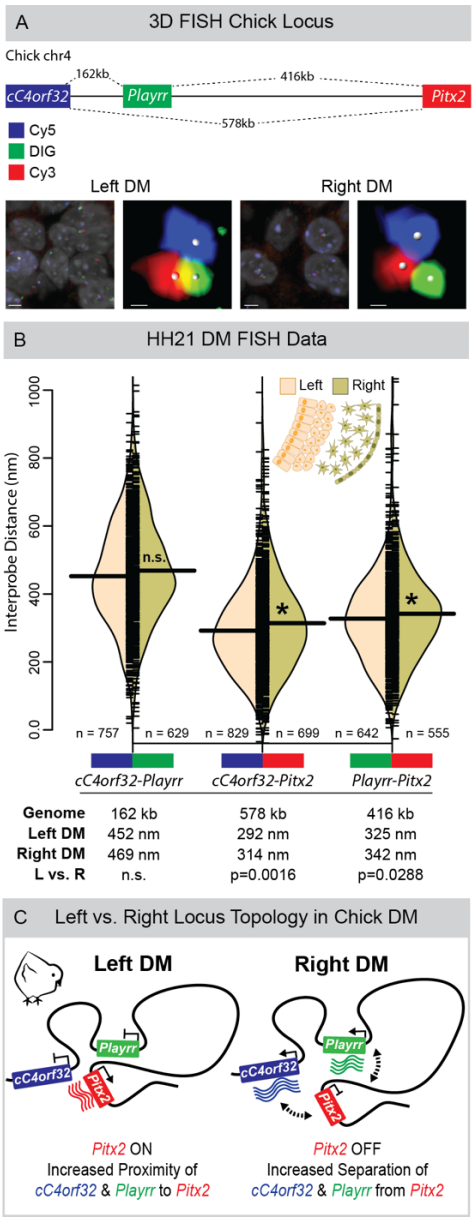


Figure 2-4. Asymmetric chromatin interactions of the Pitx2 locus in the chicken L-R DM (A) Linear genomic distances in the chicken separating cC4orf32, Playrr, and Pitx2, marked with Cy5, DIG, and Cy3 labeled DNA probes, respectively. (B) Split bean plots comparing distributions of interprobe distances for each probe pair between the left and right DM (WT, HH21). Individual measurements are plotted as short horizontal ticks, the density trace and mean are plotted as the filled curve and large horizontal bar, respectively. Table shows summary of distances separating each probe pair in the linear genome or nucleus. (C) Model of Pitx2 locus organization and resulting gene expression in the chicken DM. See also Figure S4.

We found the global architecture of the locus was similarly organized in the left and right DM (Fig. 4BC). Importantly however, we identified subtle but statistically significant L-R differences in interprobe distances for both cC4orf32-Pitx2 and Playrr-Pitx2, demonstrating that both cC4orf32 and Playrr are positioned significantly closer to Pitx2 in the left DM compared to the right (Fig. 4B, Fig. S4A, Table 1A). L-R differences in the distributions of these loci were highly reproducible across five replicate experiments (Fig. S4). Hence, mirroring the L-R asymmetric gene expression and cellular architecture within the DM, proximity of cC4orf32, Playrr, and Pitx2 in nuclei of the left DM is associated with preferential transcription of Pitx2. Conversely, in the right DM, where cC4orf32 and Playrr are expressed, they are further separated from Pitx2 (Fig. 4C). Thus, L-R differences in 3D looping of the Pitx2 locus are characterized by subtle local shifts, rather than global differences, in the relative position of individual genes within a constant overall locus topology.

2.4.6 Spatial proximity of Playrr and Pitx2 are a conserved feature of locus topology To address the degree to which asymmetric nuclear architecture of the chicken Pitx2 locus is conserved, we used the same FISH labeling scheme on mouse DM sections and analyzed pairwise interprobe distances for 5730508B09Rik (mC4orf32, the mouse ortholog of C4ORF32, Cy5, blue) at the proximal end of the desert, Playrr (DIG, green) within the desert, and Pitx2 (Cy3, red) at the distal end of the desert (Fig. 5A). In the mouse genome, the linear genomic distance separating the FISH probe pairs mC4orf32- Playrr, mC4orf32-Pitx2, and Playrr-Pitx2 is 380kb, 1.36Mb, and 980kb respectively (Table 1B). Consistent with our analysis of chicken

DM nuclei, long-range looping across the mouse gene desert positions both mC4orf32 and Playrr in significantly closer proximity to Pitx2 than to each other (Fig. 5C, Table 1B). Remarkably, although in the mouse Playrr is located nearly twice the genomic distance proximal to Pitx2 compared to chicken, the interprobe distance measured for Playrr-Pitx2 was nearly identical to that measured in chicken (Table 1B). Moreover, the difference between Playrr-Pitx2 spatial proximity in the left vs. the right DM was maintained in the mouse, further highlighting our observation that proximity of Playrr and Pitx2 in nuclei on the left is associated with preferential Pitx2 transcription (Fig. 5C, Table 1B). In contrast, we found that the position of mC4orf32 relative to both Playrr and Pitx2 is farther than observed in chicken and not different in nuclei on the left vs. right (Fig. 5CF, Table 1B). Interestingly, accompanying this subtle shift in positioning relative to Pitx2, is a shift from asymmetric to bilateral expression of C4orf32 in chick vs. mouse respectively (Fig. S6A). Thus, small differences in the position of genes relative to each other have direct consequences on their asymmetric expression. Collectively, L-R asymmetry in nuclear proximity of Playrr with Pitx2 is a conserved feature of Pitx2 locus topology.

2.4.7 Altered L-R patterning disrupts asymmetric Pitx2 locus topology

In mice lacking Pitx2, the left DM fails to condense, all L-R DM asymmetry is lost and stereotypical gut looping is randomized (a ‘double-right’ phenotype) (Fig. 5D) (Davis et al. 2008; Kurpios et al. 2008). On a molecular level, Pitx2-null embryos exhibit a loss of left-specific expression of Pitx2 target genes, such as Islet1 (Fig. 5D) and show bilateral expression of right-sided genes, such as Tbx18 (Fig. 5D) and Playrr

(Fig. 3C). To investigate whether such molecular and cellular defects due to loss of Pitx2 impact the subtle L-R differences of locus topology, we performed 3D tissue-FISH on Pitx2-null mouse DM. As expected, we found that the global organization of the locus was not impacted by loss of Pitx2 (Fig. 5E, Table 1C). Importantly, the conserved proximity of Playrr and Pitx2 in the left WT DM was lost in Pitx2-null embryos (Fig. 5E, Table 1C). Hence, in the absence of Pitx2, Playrr and Pitx2 are significantly farther apart in the left DM compared to the WT left DM, and indistinguishable from the interprobe distance in the WT right DM (Fig. 5CE, Fig. S4, Table 1BC). Thus, a double-right phenotype of the Pitx2-null DM is accompanied by isomerized nuclear architecture at the Pitx2 locus (Fig. 5F). These data indicate that whereas Pitx2 is dispensable for the global architecture of the locus it is required for subtle asymmetries involving the position of individual loci with respect to the overall locus topology (Fig. 5F).

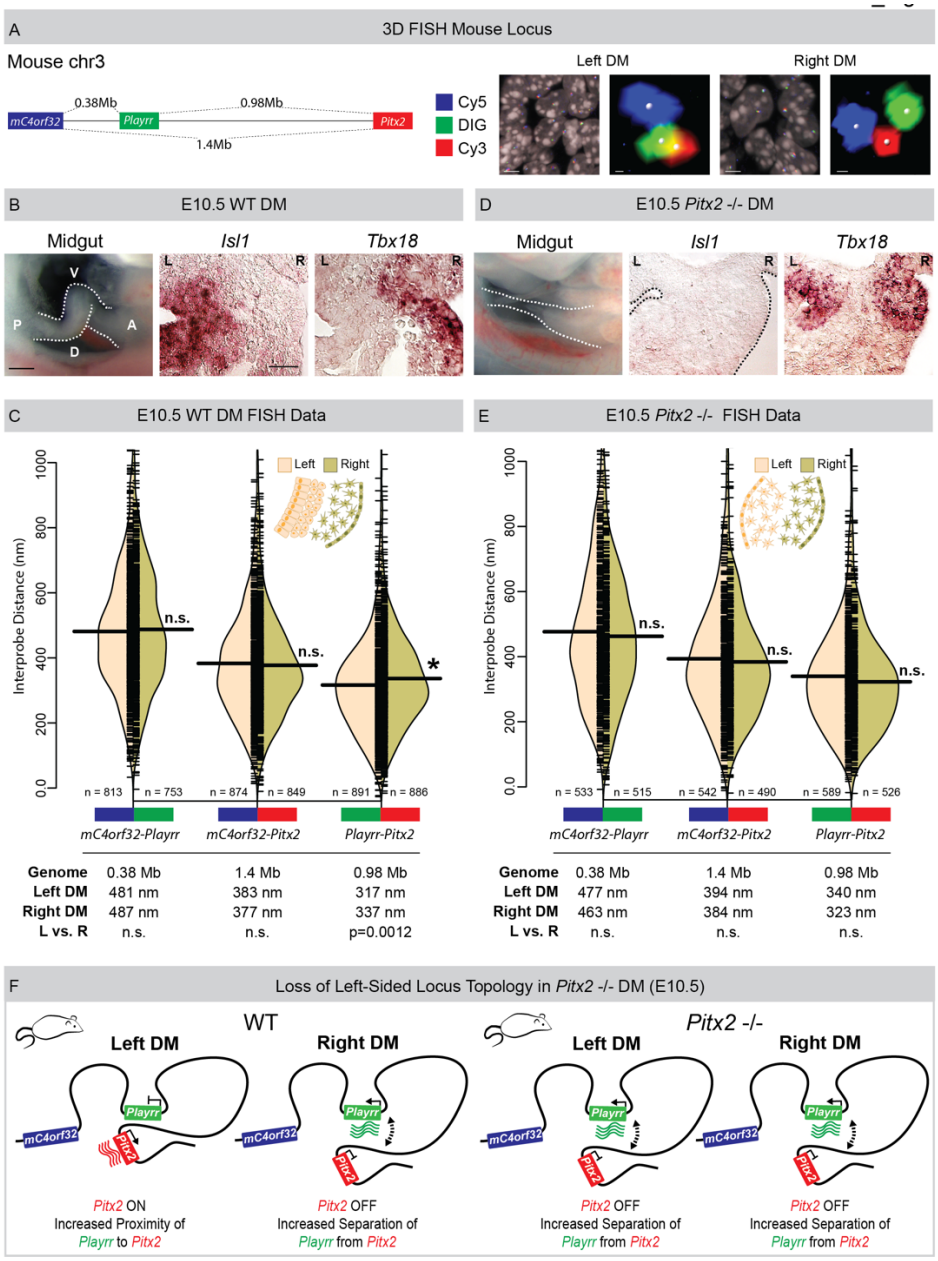


Figure 2-5. Spatial proximity of Playrr-Pitx2 chromatin interactions are dependent on Pitx2. (A) Linear genomic distances in the mouse separating mC4orf32, Playrr, and Pitx2, marked with Cy5, DIG, and Cy3 labeled DNA probes, respectively. (B) Chiral looping of the WT midgut is accompanied by left-specific Isl1 and right-specific Tbx18 expression. (C) Split bean plots of left and right DM (WT E10.5). Table shows summary of distances separating each probe pair in the linear genome or nucleus. (D) Altered midgut looping and right-isomerized DM in Pitx2-null embryos is evidenced by loss of left-specific Isl1 and bilateral Tbx18 expression. (E) Split bean plots of left and right DM (Pitx2-null E10.5). (F) Model of Pitx2 locus organization and resulting gene expression in the mouse DM. See also Figure S4.

2.4.8 Analysis of Playrr-Pitx2 chromatin interactions in mouse ES cells

The landscape of chromatin interactions associated with the transcriptional and epigenetic profile of mouse embryonic stem cells (mESCs) has been extensively characterized. Significantly, RNA Pol II ChIP-seq data (Kagey et al. 2010) demonstrate that mESCs exclusively express the asymmetric Pitx2c isoform associated with L-R patterning (Fig. 6A) while mC4orf32 is expressed at low levels and Enpep and Playrr are not expressed at all. Accompanying Pitx2c expression in these cells, p300 and Mediator binding indicate activation of the asymmetric ASE enhancer (Fig. S7) (Creyghton et al. 2010). Hence, mESCs mimic the transcriptional and regulatory profile of the Pitx2 locus in the left DM. We therefore leveraged available chromatin interaction data sets to analyze locus topology associated with Pitx2c expression in mESCs (Shen et al. 2012; Dowen et al. 2014).

Experimental techniques to detect looping interactions and define chromatin topology include HiC, a genome wide variation of chromatin conformation capture (3C), and chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) (de Wit and de Laat 2012). Both HiC and Smc1 ChIA-PET demonstrate long-range interactions are enriched across the Pitx2 locus compared to adjacent regions both upstream and downstream of the locus, consistent with a topologically associating domain (TAD) spanning the region. Importantly, HiC and Smc1 ChIA-PET data suggest division of the Pitx2 locus TAD into further sub-TAD domains that correlate well with the position of our DNA probes used for DM FISH analysis in vivo (Fig. 6B).

In order to determine if preferential proximity of Playrr to Pitx2 vs. mC4orf32 accompanies Pitx2c expression in mESCs, as we observed in the mouse DM, we quantified HiC & ChIA-PET reads linking the mC4orf32 and Playrr containing sub-TADs with each other and to Pitx2 (Fig. 6B, bottom). Significantly, although Playrr is positioned over 2.5 times farther from Pitx2 than from mC4orf32, the Playrr sub-TAD interacts nearly 4-fold more frequently with Pitx2 than with the adjacent mC4orf32 sub- TAD (Fig. 6B, bottom).

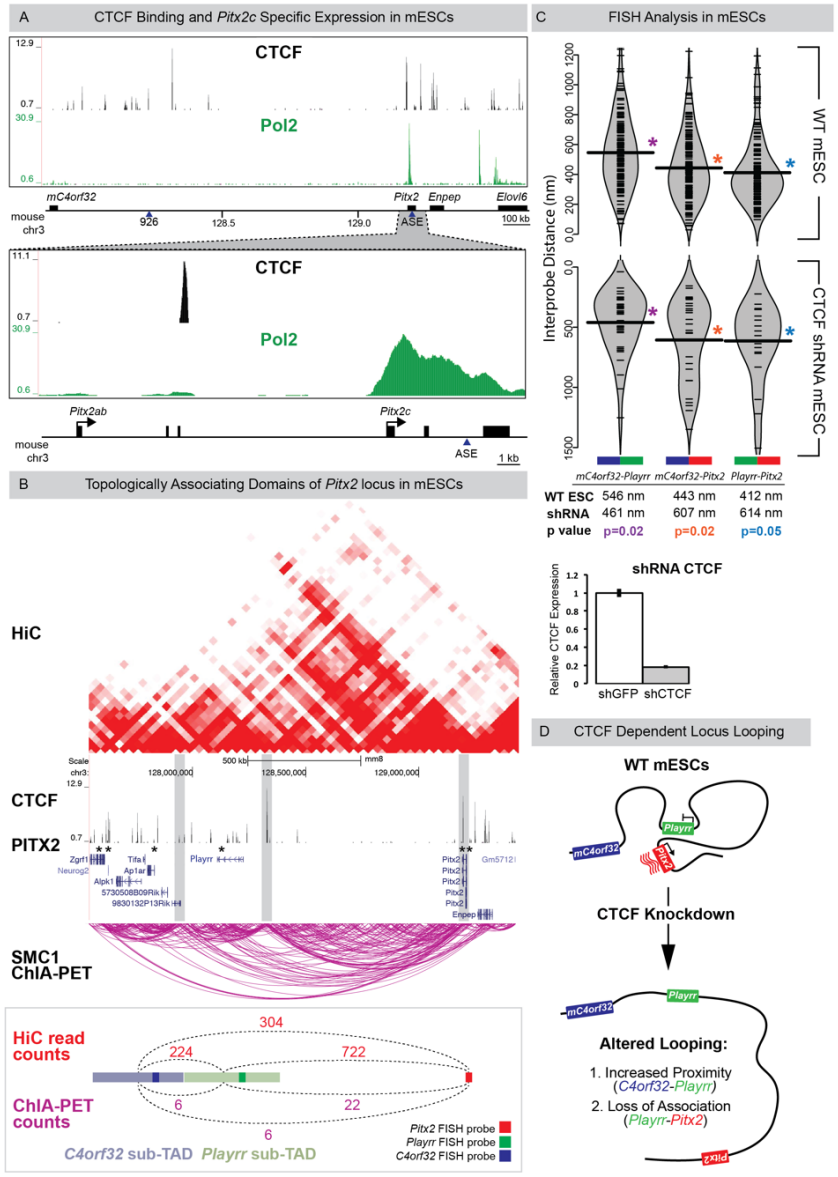


Figure 2-6. Playrr-Pitx2 chromatin interactions in mouse ES cells are dependent on CTCF. (A) Top: CTCF binding at the Pitx2 locus (CTCF ChIP-seq); Bottom: mESCs preferentially express the asymmetric Pitx2c isoform (RNA Pol II ChIP-seq). (B) A topologically associating domain (TAD) spans the Pitx2 locus (HiC, red; Smc1 ChIA- PET, magenta). Vertical grey bars highlight boundaries between sub-TADs. CTCF binding sites (ChIP-seq) at both Pitx2 and Playrr are associated with Pitx2 binding (Pitx2-FLAG ChIP-seq, asterisks). Quantification of HiC (red) and Smc1 ChIA-PET (magenta) interactions between the genomic interval containing the Pitx2 FISH probe and the Playrr sub-TAD (light green) compared to the mC4orf32 sub-TAD (light blue). (C) FISH analysis of Pitx2 locus topology in WT (top) or shRNA mediated CTCF knockdown (bottom) mESCs; table shows summary of distances separating each probe pair in the nucleus. (D) Summary of Pitx2 locus topology in mESCs; see also Table 1 for interprobe distance measurements. See also Figure S7.

Finally, we performed FISH on WT mESCs to directly confirm that locus topology as measured via HiC and ChIA-PET is consistent with the results of our in vivo FISH analysis. We found that the 3D organization of the Pitx2 locus in the nuclei of mESCs is nearly identical to that of the DM of the mouse, where Playrr-Pitx2 are in closest proximity and mC4orf32-Playrr are furthest separated (Fig. 6C top, Table 1D). Thus, both molecular analysis via chromatin looping assays and direct visualization via FISH confirm the preferential association of Playrr with Pitx2 that accompanies asymmetric Pitx2 expression. CTCF is essential for maintaining long-range Playrr-Pitx2 interaction CCCTC-binding factor, CTCF, is a sequence-specific architectural protein known to be involved in long-range loop formation whose interaction with additional architectural proteins such as Cohesin and Mediator shape constitutive and dynamic cell- type specific chromatin organization (Splinter et al. 2006; Phillips-Cremins et al. 2013).

CTCF ChIP-seq data from mESCs identifies numerous significant CTCF binding peaks across the region spanned by mC4orf32 and Playrr at the proximal end of the gene desert (Fig. 6B) (Kagey et al. 2010). Smc1 ChIA-PET and CTCF ChIP-seq both highlight that Playrr-Pitx2 interaction is mediated via CTCF site within the Pitx2 gene body and a CTCF site located upstream of e926/Playrr (Fig. 6AB). Significantly, further validating a role for Pitx2-dependent interaction of Pitx2 with Playrr, in vivo Pitx2-FLAG ChIP-seq data from the Martin lab establishes that the CTCF binding sites at both Pitx2 and Playrr are also associated with Pitx2 binding (Fig. 6B) (Tao et al. 2014).

To test this hypothesis we used lentiviral shRNA to knockdown CTCF in mESCs (Fig. 6C). Compared to control ESCs, knockdown of CTCF abrogated long-range looping across the gene desert and disrupted chromatin organization of the locus (Fig. 6C,D, Table 1DE). Significantly, compared to WT mESCs, CTCF knockdown decreased the distance separating mC4orf32-Playrr (mean separation 546nm vs. 461nm, WT vs CTCF knock down respectively) while the close proximity of Playrr to Pitx2 was lost (mean separation 412nm vs. 614nm, WT vs CTCF knock down, respectively). These data demonstrate that CTCF mediated chromatin looping is essential for maintaining long- range interactions at the Pitx2 locus, a prerequisite for subsequent Pitx2-dependent regulatory interactions across the L-R axis (Fig. 6D).

2.5 DISCUSSION

Spatiotemporal Pitx2 expression controls transcriptional regulatory programs essential for normal development and the establishment of asymmetric organ morphology. In the DM, left-sided Pitx2 regulates molecular and cellular asymmetry required to initiate intestinal looping and gut vasculogenesis (Davis et al. 2008; Kurpios et al. 2008; Welsh et al. 2013; Mahadevan et al. 2014). Remarkably, we show here that the binary L-R organization of the DM at the tissue level is mirrored in the binary transcriptional output and chromatin level organization of the Pix2 locus in the left and right DM (Fig. 7).

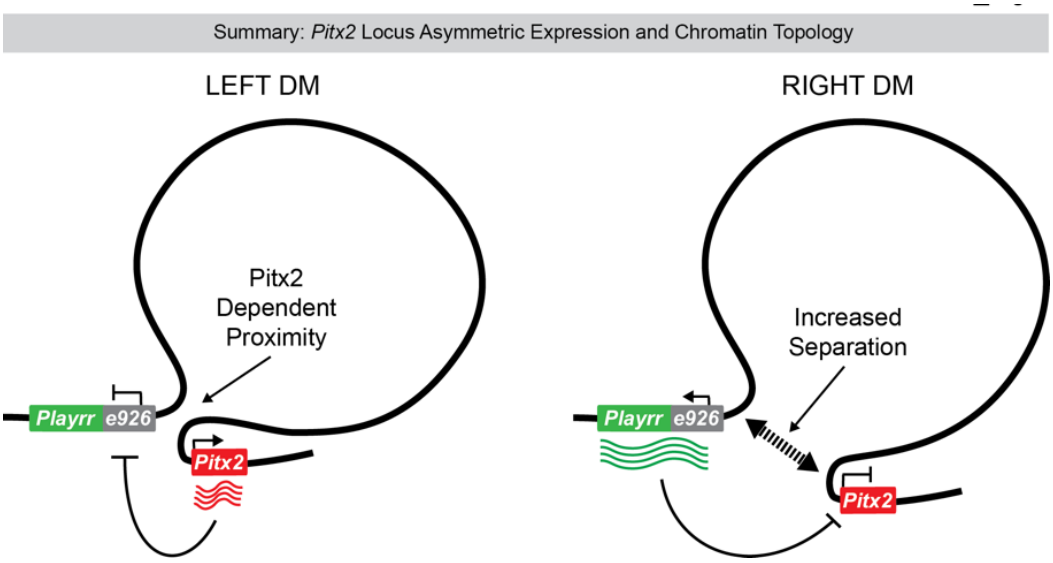


Figure 2-7. Summary Model: Pitx2 mediated asymmetry across multiple scales of biological organization. Pitx2 is required to establish asymmetry 1) at the level of chromatin in nuclei of cells in either left or right DM that is in turn required for 2) Pitx2 locus expression and subsequent expression of Pitx2 target genes controlling L-R differences in cell behavior in order to ultimately 3) coordinate the shape and placement of individual organs within the body.

Leveraging the organization of the DM, our 3D tissue FISH analysis characterizes how the large conserved chromosome interval harboring the Pitx2 locus is organized within the 3D space of the nucleus in order to effect L-R patterning and asymmetric gene expression. The 3D topology of the Pitx2 locus, as measured optically by a common set of 3 reference points, demonstrates both global and local scale organizational principles and is consistent with HiC and Smc1 ChIA-PET data available from mESs, an in vitro model for left-specific expression from the Pitx2 locus. We establish that in chicken and mouse tissue, long-range chromatin looping across a large gene desert is a constitutive feature of locus organization dependent on CTCF. Conversely, consistent with L-R differences in cis-regulatory interactions within this invariant 3D topology, we show that Playrr-Pitx2 are in closer proximity in the left DM where Pitx2 is preferentially expressed but positioned further apart in the

right DM where Playrr is expressed and Pitx2 is not (Fig. 7). While L-R differences in locus topology, as measured by mean interprobe distance, are notably small, these differences are highly reproducible across biological replicates. Moreover, in the absence of Pitx2, mean Playrr-Pitx2 interprobe distance in the left DM is similar to that in the right DM of both Pitx2-/- and WT mice, supporting that small L-R differences are Pitx2-dependent and hence biologically meaningful. Finally, our FISH analysis only marks the relative position of asymmetrically expressed genes in the left or right DM and it is likely that the subtle L-R proximity differences are actually secondary to interactions between promoters and sequences that we have not labeled. Indeed, both HiC data and Smc1 ChIA-PET in mESCs demonstrate that the preferential interaction between Playrr and Pitx2 likely involves looping between CTCF bound sites just distal to e926/Playrr and within the Pitx2 gene body (Fig. 6B).

The preassembly of constitutive long range chromatin looping necessary for proper spatial expression has been documented at other critical developmental loci (Shopland et al. 2006; Montavon et al. 2011; Jin et al. 2013; de Laat and Duboule 2013). Such preassembly of permissive regulatory topologies has been suggested to poise cells for rapid changes in expression. For example, simple shifts in regulatory contacts, analogous to allosteric conformational changes that alter protein activity, may promote transcriptional activation via release of paused RNA Pol II (Liu et al. 2013; Ghavi-Helm et al. 2014; de Laat and Duboule 2013). This suggests that L-R differences in Pitx2 vs. Playrr transcription may be achieved by regulating the position of their promoters and associated REs within nuclear foci containing RNA Pol II and tissue-specific co-factors that target release of paused RNA Pol II already assembled at their

promoters (Razin et al. 2011; Schaukowitch et al. 2014). Implementation of such regulatory strategy, where 3D topology facilitates efficient RNA Pol II trafficking, may provide a mechanism for robust establishment of asymmetric expression in response to very transient Nodal signaling critical for Pitx2 induction (Shiratori and Hamada 2006).

Supporting this hypothesis, our microarray and RNA in situ hybridization data demonstrate the exclusive expression of Mllt3 in the left DM, a transcriptional co-activator associated with rapid establishment of cell fate and lineage commitment (data not shown) (Flajollet et al. 2013; Pina et al. 2008). Significantly, Mllt3 directly interacts with and activates P-TEFb, a kinase targeting Ser2 in the C-terminal domain (CTD) of RNA Pol II, a critical step for release of paused RNA Pol II and transcriptional elongation (Shim et al. 2002; Biswas et al. 2011; Bitoun et al. 2007). Mllt3 also complexes with Dot1L, a developmentally essential histone methyltransferase that interacts with the phosphorylated CTD of RNA Pol II and is solely responsible for histone H3K79 methylation at active enhancers and transcribed genes. Interestingly, we observed H3K79 methylation flanking the ASE enhancer in mESCs expressing Pitx2c (Fig. S7) (Steger et al. 2008; Kim et al. 2012; Jones et al. 2008; Bonn et al. 2012) further supporting the involvement of Mllt3 at the Pitx2 locus.

Maintenance of the global organization of a locus across a range of cellular contexts is a defining characteristic of topologically associating domains (TADs), a feature of genomic structure that emerges at the megabase scale characterized by preferential association of long-range chromatin interactions (Phillips-Cremins et al. 2013; Nora et

al. 2013). The structure of TADs correlate with blocks of genomic synteny and thus changes in their structure are likely evolutionarily constrained (Dixon et al. 2012; Nora et al. 2013). The boundaries of TADs are enriched for binding of CTCF (Dixon et al. 2012) and we show here that CTCF is required for maintenance of the megabase scale topology of the Pitx2 locus in mESCs. High-resolution long-range chromatin interaction maps provided by HiC analysis of mESCs (Shen et al. 2012) support the presence of a TAD extending proximally from Pitx2 across the gene desert that span the region that we have characterized via 3D FISH (Shen et al. 2012) (Fig. 6B).

However, we observed changes in spatial gene expression between chicken and mouse consistent with significant differences in positioning of the proximal but not distal gene desert. Interestingly, in mESCs, binding of CTCF downstream of exon 3 of the bilaterally expressed Pitx2a isoform overlaps a conserved CTCF recognition sequence in chicken and suggests a conserved role in chromatin looping for CTCF binding this site (Fig. 6B) (Phillips-Cremins et al. 2013). In contrast, CTCF binding proximally in the vicinity of cC4orf32/mC4orf32 appears to lack conservation or are potentially divergent. Interestingly, although the gene desert is conserved in zebrafish, they lack an ortholog of C4ORF32, further supporting evolutionary divergence of the proximal gene desert (Fig. S6C) (Volkmann et al. 2011). Furthermore, while Elovl6 is present, the zebrafish Pitx2 locus lacks an Enpep ortholog, and although both chick and mouse loci possess Enpep, this gene is not expressed in the mouse DM compared to its right-sided expression in chicken. This suggests evolution has acted to continuously refine regulatory interactions at the Pitx2 locus while maintaining critical control over spatial expression of Pitx2. Consistent with this hypothesis, the HiC interaction map of mESC

shows a very distinct interaction boundary separating Pitx2 from Enpep, supporting that these adjacent genes are partitioned into separate regulatory domains (Fig. 6B). Thus, while the Pitx2 locus in chicken and mouse exhibits conserved 3D chromatin architecture, species-specific differences in this architecture have consequences on spatial gene expression and may result from the acquisition or turnover of CTCF binding.

Characterization of distal regulatory elements in the genome and identification of their target promoters is requisite for advancing our understanding of genomic regulatory logic. This study demonstrates the utility of combining GRO-seq with dREG analysis in vivo to characterize nascent transcription and identifying the activity of REs within a single sample. An unanticipated finding from this approach was the identification of Playrr as a novel, asymmetrically expressed lncRNA derived from the conserved e926 sequence element. Rather than matching the left-specific enhancer activity of e926 observed in transgenic embryos, we show that Playrr is exclusively expressed in the right DM. Thus, mapping of active REs in situ via dREG analysis avoids potentially confounding data that may arise from testing putative cis-regulatory sequences in isolation of their endogenous context (Marinić et al. 2013).

A coherent understanding of the biological roles of lncRNAs is still lacking, although a role in transcriptional regulation is an emerging theme. Studies demonstrate that lncRNAs can function either locally at their site of transcription or at distal sites in either cis or trans in order to control gene expression (Rinn and Chang 2012; Vance and Ponting 2014). Recent analysis of the trans-acting lncRNA Fendrr, which interacts

with the polycomb repressive complex 2 (PRC2), establishes a role for lncRNAs in regulation of Pitx2 developmental expression (Grote et al. 2013). Although the impact on L-R patterning was not addressed, mutation of Fendrr causes loss of Pitx2 silencing and altered differentiation of the lateral mesoderm, resulting in an embryonic lethal phenotype consistent with Pitx2 misexpression (Grote et al. 2013; Grote and Herrmann 2013).

We now show that Playrr, a lncRNA asymmetrically expressed from the Pitx2 locus itself, also acts to modulate Pitx2 expression. Interestingly, expression of Fendrr is restricted to gastrulation stages of development while Playrr is expressed later during organogenesis. This suggests that negative regulation and modulation of Pitx2 expression levels require the sequential activity of multiple lncRNAs, analogous to the two-step process of induction and maintenance of Pitx2 expression on the left (Shiratori et al. 2001). Significantly, although the mechanism is unknown, evidence for auto-regulation of asymmetric gene expression by Pitx2 has been documented in other developmental contexts including the heart and limb (Shiratori et al. 2006, 2014). The contralateral expression and mutual antagonism of Playrr and Pitx2 expressed from the same locus suggest e926/Playrr is an integrated regulatory module that provides such autoregulation. Finally, as the function of many lncRNAs is presumed to depend on structural conformation that allows interaction with DNA, RNA, or regulatory proteins (Rinn and Chang 2012; Nitsche et al. 2015; Johnsson et al. 2014), further characterization of our Playrr^{Ex1sj} mutant provides a novel means to address lncRNA structure and function and the mechanisms through which Playrr influences Pitx2 locus expression.

Pitx2 is expressed in dynamic spatial patterns essential for normal development and individual Pitx2 isoforms play distinct dosage dependent roles in diverse tissues (Liu et al. 2001; Waite et al. 2013). Notably, looping of the midgut requires high levels of Pitx2c expression and is particularly sensitive to ectopic bilateral expression of Pitx2c (Liu et al. 2001). Thus reciprocal interactions between Pitx2 and the binary e926/Playrr regulatory module likely provide an organ intrinsic mechanism to tightly regulate Pitx2 dosage.

Consistent with the pleiotropic function of Pitx2, mutations of human PITX2 result in a spectrum of serious birth defects associated with Axenfeld-Rieger Syndrome and predispose otherwise healthy individuals to cardiac fibrillation and arrhythmia (Semina et al. 1996; Tao et al. 2014; Wang et al. 2014). Conservation of the expansive gene desert at the locus and the number of noncoding mutations identified in human ARS patients likely represent the considerable regulatory demand necessary to orchestrate complex Pitx2 expression. Therefore, identification of mechanisms that direct precise spatiotemporal expression of Pitx2 is as critical as defining its downstream regulatory targets. The binary L-R organization of the DM, accessible for both targeted and whole animal genetic manipulation, combined with the unique pattern of asymmetric gene expression mirrored by L-R 3D chromatin organization of the Pitx2 locus, represents a powerful experimental system to advance understanding of the integrated regulation and essential function of the evolutionarily conserved Pitx2. Furthermore, our work begins to shed light on the cis-regulatory mechanisms

and etiology of ARS and Pitx2 associated AF offering unprecedented potential for developing mouse models of these important human diseases.

2.6 EXPERIMENTAL PROCEDURES

2.6.1 Animals

Mouse embryos were collected from timed matings with the morning of the plug defined as E0.5. Pitx2 mutant mice, $Pitx2^{hd}$ allele (Lu et al., 1999), were used for analyses. Fertile eggs (White Leghorn) obtained from the Cornell Poultry Research Farm were incubated at 38°C and staged (Hamburger and Hamilton, 1992). Experiments adhered to guidelines of the Institutional Animal Care and Use Committee of Cornell University, under the Animal Welfare Assurance on file with the Office of Laboratory Animal Welfare.

2.6.2 Histology

Embryos were dissected in PBS and fixed in 4% PFA/PBS or Bouin's fixative overnight, dehydrated in ethanol and embedded in paraffin, sections were collected on Superfrost Plus slides. 6μm paraffin sections were stained with Hematoxylin &Eosin and included a brief wash with 1% HCl/70% EtOH followed by NH4OH.

2.6.3 Laser capture microdissection (LCM) and microarray

Detailed methods and full microarray results are in preparation to be published elsewhere. Briefly, cryosections of unfixed/flash frozen WT HH21 chicken DM were

arrayed on membrane slides (Leica, 11505189). The asymmetric morphology of the left and right DM was used to discriminate and capture separately each of four compartments. Three separate biological experiments of four cell compartments each were performed, giving 12 microarray samples in total (biological triplicate). A contrast stain (HistoGene Kit, Arcturus, KIT0419) was applied just prior to LCM. RNA was isolated using the PicoPure RNA Isolation kit (Arcturus, KIT0202), and cDNA was prepared using the WT- Ovation Pico kit (NuGen). Affymetrix cRNA target labeling reactions were carried out per manufacturer instructions, and GCOS output files were loaded into GeneSpringGX 7.3 or GeneSpring 13 software packages (Agilent) for expression analyses. The fidelity of our dissections and analyses was validated using genes with known asymmetric expression profiles (Pitx2, left; Tbx18, right) (Welsh et al. 2013). Of 38,535 probe sets on the Affymetrix Chicken microarray, 18,001 were called Present or Marginal in all 12 replicate samples analyzed. A paired t-test with equal variances yielded 3,524 probe sets (in LM vs RM samples), which were then filtered on fold change and flags to obtain genes highly specific to each mesenteric cell compartment. The microarray data has been deposited in the NIH GEO (Accession number pending).

2.6.4 RNA in situ hybridization

250μm thick embryo slices for whole mount RNA in situ hybridization (ISH) were collected with a McIlwain tissue chopper (Campden Instruments), fixed in 4% PFA/PBS ON, dehydrated, and stored in 100% methanol prior to processing. Whole mount ISH followed standard protocols as previously described (Welsh et al. 2013).

2.6.5 Statistical analysis

Measurement data were analyzed with R and Mann–Whitney–Wilcoxon test was used to compare interprobe distances in FISH experiments, and data was plotted with the beanplot package in R (Kampstra 2008). Box plots of FISH data in supplemental figure 4 were generated using JMP Pro 11. Student's t-test was used for comparison of QRT-PCR data, error bars show ± S.E.M.

2.6.6 L-R GRO-seq and dREG analysis

Whole embryos (n=250) were chopped into 250μm transverse slices using a McIlwain tissue chopper, followed by manual microdissection of the left and right DM. Collected tissues were pooled, snap frozen in liquid nitrogen and stored at -80°C until processing for GRO-seq as previously described (Core et al. 2008). We also performed GRO-seq on whole embryos (HH21, n=2), on embryonic heads (HH12, n=250) and on the left and right hemisected embryos (HH12, n=250), to monitor enrichment of asymmetric reads and further define the regulatory landscape of the Pitx2 locus across several developmental contexts. dREG analysis of all GRO-seq data

sets was implemented in R and is available via GitHub (https://github.com/Danko-Lab/dREG).

2.6.7 Cloning, plasmids, and oligonucleotides

Full-length cDNAs and probes for RNA ISH were cloned using TA cloning (Invitrogen) and oligo-dT primed cDNA reverse transcription (Superscript III, Invitrogen) from RNA pooled from HH19 and HH21 whole chicken, or from E8.5-18.5 whole mouse embryos. Cloned DNA was sequence-verified.

2.6.8 CRISPR/Cas9 targeting of e926/Playrr

The following Guide RNAs were designed using available online tools (http://crispr.mit.edu/) (Hsu et al. 2013) and cloned into BbsI cut pX330: Playrr, TAGACGCAGCTGTGCTTAGAAGG; e926 proximal GTGGCGGACTCATGTTAAAAAGG; e926 distal GTGATTCCCACCACGCTTTGAGG. For Playrr targeting, the 156bp ssODN used as a repair template carries a 3bp change to disrupt the 5' splice site of Playrr exon 1 and introduce an EcoRI restriction site for genotyping (see Fig. 3). sgRNA was in vitro transcribed (Ambion MEGAshortscriptT7 kit) from PCR generated template and purified (Ambion MEGAclear kit). Each sgRNA (50ng/ul), ssODN (10ng/ul) and Cas9 mRNA (100ng/ul) were injected into F1 hybrid (C57BL/6J x FvB/N) 1 cell embryos by the Cornell Stem Cell and Transgenic Core Facility. Injected embryos were cultured to the 2 cell stage prior to transfer to recipient females. Two independent lines were established and analyzed for each mutation.

2.6.9 Quantitative reverse-transcriptase PCR (QRT-PCR)

Embryonic tissue was isolated in cold PBS, and stored in RNAlater until RNA was extracted with a Qiagen RNeasy miniprep kit. For cultured cells, RNA was extracted using 4-6 x 10^6 cells. 2µg of RNA was reverse transcribed using the ABI high-capacity cDNA archive kit and diluted to 20ng/µl. The following TaqMan gene expression assays were used for relative quantification using an ABI7500 realtime PCR system: Actb, Mm00607939_s1; 5730508B09Rik (mC4orf32), Mm02375228_s1; D030025E07Rik (Playrr), Mm03937997_m1; CTCF, Mm00484027_m1; Enpep, Mm00468278_m1; GAPDH, Mm99999915_g1; Pitx2ab, Mm00660192_g1; Pitx2c, Mm00440826_m1; pan- Pitx2, Mm01316994_m1.

2.6.10 3D DNA fluorescent in situ hybridization (FISH)

The following BACs spanning either the chicken or mouse Pitx2 locus were used to generate FISH probes targeting Pitx2 locus genes: chicken – CH261-95I8, CH261-66N5, CH261-187K8, CH261-34B16, CH261-110J5, CH261-91C24, CH261-134M23; mouse – RP23-306C6, RP23-328J13, RP23-225C17, RP24-98F15, RP23-266N9, RP23-106J9, RP23-150F9, RP24-156B21, RP24-100G2, RP23-307L21. Restriction enzyme digestion or PCR amplification using purified BAC DNA was used to subclone DNA templates for probe production. The genomic distances analyzed in this study are resolvable via interphase FISH (50-100 kb); to further

maximize resolution we designed probes to be less than 25kb (Ronot and Usson 2001; Trask et al. 1991). Genomic intervals spanned by FISH probes and probe sizes: chicken (galGal4): cC4orf32 (23.8kb), chr4:56,825,232- 56,849,123; e926 (16kb), chr4:56,989,104-57,005,206; Pitx2 (17.5kb), chr4:57,398,355- 57,415,925; mouse (mm10): mC4orf32 (15.5kb), chr3:127,846,052-127,861,644; e926 (16kb), chr3:128,221,875-128,237,854; Pitx2 (21kb), chr3:129,199,824-129,221,272. Probes were labeled with dUTP-Cy3, -Cy5, or -DIG, via nick translation of 1μg of DNA (Roche).

For 3D FISH, embryos were fixed overnight in 4% paraformaldehyde and embedded in paraffin following standard protocol. 6μm sections were collected on Suprerfrost plus slides, dried overnight at 37^{o}C, baked at 60^{o}C for 20 minutes, cooled, and then dewaxed in xylene with 2 washes for 10 minutes each, followed by 2 washes in 100% ethanol, and 1 wash in 70% ethanol, 5 minutes each. Sections were then treated for 5 minutes with 0.2N NaOH in 70% ethanol to remove RNAs, washed twice in 70% ethanol, rehydrated, and washed 10 minutes in 0.1M citrate buffer at 80^{o}C. Sections were washed in water, equilibrated in 2X SSC for 5 minutes, denatured for 2.5 minutes at 79^{o}C in 70% formamide/2X SSC. Sections were hybridized overnight 37^{o}C with 50ng of probe, 10μg species specific Cot-1 DNA, 10μg salmon sperm DNA, and 10μg tRNA. The following day sections were washed with 2X SSC/50% formamide at 37^{o}C, 2X SSC 37^{o}C, 1X SSC at room temperature, for 15 minutes each. DIG labeled probes were detected with AF488 anti-DIG conjugated antibodies diluted 1:500 in 4X SSC/1% BSA.

For FISH on cultured V6.4 mouse ESCs cells were seeded onto coverslips, fixed in 4% paraformaldehyde, then permeabilized following published protocols (Kurz et al., 1996). Hybridization followed the same protocol with the exception that the citrate unmasking step was omitted.

2.6.11 Quantification of HiC and Smc1 ChIA-PET interactions

Previously published HiC (Dixon et al. 2012, GEO accession GSE35156) or ChIA-PET (Dowen et al. 2014, GEO accession GSE57911) data sets were used to analyze chromatin interactions in mESCs linking the following genomic intervals (mm9): mC4orf32 sub-TAD, chr3:127279136-127674264; Playrr sub-TAD, chr3:127674265-128090964, Pitx2 FISH probe, chr3:128902742-128924190. The boundaries of sub-TADs were defined based on CTCF peaks (Roadmap data). To quantify HiC interactions, the virtual 4C tool (http://promoter.bx.psu.edu/hi-c/) was used with rs47546564, rs31670515, and Pitx2 as anchor points for the mC4orf32 sub-TAD, Playrr sub-TAD, and Pitx2 FISH probe, respectively. Read counts across each genomic interval connected to an individual anchor point were summed. To quantify ChIA-PET interactions, we counted the number of PETs that had one end overlapping with Pitx2, mC4orf32 or mPlayrr FISH probes and the other end overlapping with individual sub-TADs by at least 1 bp. The count results were similar from two Smc1 ChIA-PET replicates; we therefore took the sum of the two and visualized the data in WashU genome browser.

2.6.12 Image acquisition and analysis for 3D FISH

Slides were imaged on a Zeiss 710 scanning laser microscope using a 63X/1.4 NA Plan APOCHROMAT oil immersion objective and Z-series data were acquired using the optimal step size. Imaris image analysis software (Bitplane) was used to quantify interprobe distances. Following image filtering using baseline subtraction, FISH signals for each channel were defined using the built in spot function which segments the signal using a Gaussian filter based background subtraction method and calculates the center of image mass of each segmented spot. The position of each spot object in (x, y, z) is equal to the valueofthecenterofhomogeneous massoftheobject. Distance measurements between the center of mass for each signal was determined pairwise for each probe pair using the measurement point tool and verified by 2 independent investigators.

2.6.13 Lentiviral production and transduction

293FT cells were transfected with pLKO.1 lentiviral plasmids containing either shRNAs targeting CTCF or GFP (control), along with the packaging plasmids pLP1, pLP2, and pLP/VSV-g, using Lipofectamine 2000. The following day, media was replaced with ES media minus LIF and cells were incubated overnight. Viral media was collected and centrifuged at 3000 rpm to pellet debris and then filtered through a 0.45μm PVFD filter and frozen until use. Virus containing media was supplemented with additional FBS, LIF, and 6μg/ml polybrene, prior to use.

WT V6.4 ES cells, expanded on gamma irradiated feeder cells, were trypsinized, plated for 30 minutes onto gelatin-free tissue culture plates to remove feeders prior to transduction. Following seeding ES cells at 2 x 10^6 cells/100mm plate and incubation

for 24 hours, media was exchanged with ES media supplemented with 6μg/ml polybrene and plates were incubated for an additional15 minutes at 37^{o}C prior to exchanging media with virus containing ES media. After transduction, media was exchanged for normal ES cell media and cells were incubated for an additional 24 hours prior to selection with 2μg/ml puromycin. After 3 days of selection, surviving ES cells were trypsinized and plated for FISH or collected for RNA.

2.7 ACKNOWLEDGEMENTS

We thank Dr. Len Pennacchio and Jennifer Akiyama of the VISTA enhancer browser (http://enhancer.lbl.gov) for providing archived transgenic embryos, Dr. Stephen Dalton for sharing shRNA constructs, David Gludish for technical advice regarding shRNA experiments and for microarray analyses, and Dr. John Schimenti for helpful discussions and stocks of V6.5 mES cells. Jason Crossley, Michelle Miller, and Sriharsha Ponna provided technical assistance. We are grateful to Aravind Sivakumar and other members of the Kurpios lab for feedback on the manuscript. We also thank Cornell University Biotechnology Resource Center (BRC) Imaging Facility and Carol Bayles for assistance with confocal microscopy (NIH 1S10RR025502-01) and Rob Munroe and Chris Abratte of the Cornell Stem Cell and Transgenic Core Facility.

REFERENCES

Anderson E, Devenney PS, Hill RE, Lettice LA. 2014. Mapping the Shh long-range regulatory domain. Development. 2014 Oct;141(20):3934-3943.

Attanasio C, Nord AS, Zhu Y, Blow MJ, Li Z, Liberton DK, Morrison H, Plajzer-Frick I, Holt A, Hosseini R, et al. 2013. Fine tuning of craniofacial morphology by distant- acting enhancers. Science 342: (6157):1241006.

Biswas D, Milne TA, Basrur V, Kim J, Elenitoba-Johnson KSJ, Allis CD, Roeder RG. 2011. Function of leukemogenic mixed lineage leukemia 1 (MLL) fusion proteins through distinct partner protein complexes. Proc Natl Acad Sci 108: 15751–15756.

Bitoun E, Oliver PL, Davies KE. 2007. The mixed-lineage leukemia fusion partner AF4 stimulates RNA polymerase II transcriptional elongation and mediates coordinated chromatin remodeling. Hum Mol Genet 16: 92–106.

Bonn S, Zinzen RP, Girardot C, Gustafson EH, Perez-Gonzalez A, Delhomme N, Ghavi- Helm Y, Wilczyński B, Riddell A, Furlong EEM. 2012. Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. Nat Genet 44: 148–156.

Cajiao I, Zhang A, Yoo EJ, Cooke NE, Liebhaber SA. 2004. Bystander gene activation by a locus control region. EMBO J 23: 3854–3863.

Core LJ, Martins AL, Danko CG, Waters CT, Siepel A, Lis JT. 2014. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. Nat Genet 46: 1311–1320.

Core LJ, Waterfall JJ, Lis JT. 2008. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science 322: 1845–1848.

Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, et al. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A 107: 21931–21936.

Danko CG, Hyland SL, Core LJ, Martins AL, Waters CT, Lee HW, Cheung VG, Kraus WL, Lis JT, Siepel A. 2014. Accurate Identification of Active Transcriptional Regulatory Elements from Global Run-On and Sequencing Data. bioRxiv. http://www.biorxiv.org/content/early/2014/11/12/011353.abstract.

Davis NM, Kurpios NA, Sun X, Gros J, Martin JF, Tabin CJ. 2008. The Chirality of Gut Rotation Derives from Left-Right Asymmetric Changes in the Architecture of the Dorsal Mesentery. Dev Cell 15: 134–145.

De Laat W, Duboule D. 2013. Topology of mammalian developmental enhancers and their regulatory landscapes. Nature 502: 499–506.

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485: 376–380.

Drissen R, Palstra R-J, Gillemans N, Splinter E, Grosveld F, Philipsen S, de Laat W. 2004. The active spatial organization of the beta-globin locus requires the transcription factor EKLF. Genes Dev 18: 2485–2490.

Duboc V, Röttinger E, Lapraz F, Besnardeau L, Lepage T. 2005. Left-right asymmetry in the sea urchin embryo is regulated by nodal signaling on the right side. Dev Cell 9: 147–158.

Flajollet S, Rachez C, Ploton M, Schulz C, Gallais R, Métivier R, Pawlak M, Leray A, Issulahi AA, Héliot L, et al. 2013. The Elongation Complex Components BRD4 and MLLT3/AF9 Are Transcriptional Coactivators of Nuclear Retinoid Receptors. PLoS One 8(6):e64880.

Flomen RH, Vatcheva R, Gorman PA, Baptista PR, Groet J, Barišić I, Ligutic I, Nižetić D. 1998. Construction and Analysis of a Sequence-Ready Map in 4q25: Rieger Syndrome Can Be Caused by Haploinsufficiency ofRIEG, but Also by Chromosome Breaks ≈90 kb Upstream of This Gene. Genomics 47: 409–413.

Franco D, Campione M. 2003. The role of Pitx2 during cardiac development: Linking left-right signaling and congenital heart diseases. Trends Cardiovasc Med 13: 157–163.

Ghavi-Helm Y, Klein FA, Pakozdi T, Ciglar L, Noordermeer D, Huber W, Furlong EEM. 2014. Enhancer loops appear stable during development and are associated with paused polymerase. Nature. 512(7512):96-100.

Grote P, Herrmann BG. 2013. The long non-coding RNA Fendrr links epigenetic control mechanisms to gene regulatory networks in mammalian embryogenesis. RNA Biol 10: 1579–1585.

Grote P, Wittler L, Hendrix D, Koch F, Währisch S, Beisaw A, Macura K, Bläss G, Kellis M, Werber M, et al. 2013. The Tissue-Specific lncRNA Fendrr Is an Essential Regulator of Heart and Body Wall Development in the Mouse. Dev Cell 24: 206– 214.

Hamburger V, Hamilton HL. 1951. A series of normal stages in the development of the chick embryo. J Morphol 88: 49–92.

Jin F, Li Y, Dixon JR, Selvaraj S, Ye Z, Lee AY, Yen C-A, Schmitt AD, Espinoza C a, Ren B. 2013. A high-resolution map of the three-dimensional chromatin interactome in human cells. Nature 503: 290–294.

Jones B, Su H, Bhat A, Lei H, Bajko J, Hevi S, Baltus GA, Kadam S, Zhai H, Valdez R, et al. 2008. The histone H3K79 methyltransferase Dot1L is essential for mammalian development and heterochromatin structure. PLoS Genet 4(9):e1000190.

Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, Ebmeier CC, Goossens J, Rahl PB, Levine SS, et al. 2010. Mediator and cohesin connect gene expression and chromatin architecture. Nature 467: 430–435.

Kampstra P. 2008. Beanplot : A Boxplot Alternative for Visual Comparison of Distributions. J Stat Softw 28: 1–9.

Kikuta H, Laplante M, Navratilova P, Komisarczuk AZ, Engström PG, Fredman D, Akalin A, Caccamo M, Sealy I, Howe K, et al. 2007. Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates. Genome Res 17: 545–555.

Kim SK, Jung I, Lee H, Kang K, Kim M, Jeong K, Kwon CS, Han YM, Kim YS, Kim D, et al. 2012. Human histone H3K79 methyltransferase DOT1L methyltransferase binds actively transcribing RNA polymerase II to regulate gene expression. J Biol Chem 287: 39698–39709.

Kurpios NA, Ibañes M, Davis NM, Lui W, Katz T, Martin JF, Belmonte JCI, Tabin CJ. 2008. The direction of gut looping is established by changes in the extracellular matrix and in cell:cell adhesion. Proc Natl Acad Sci 105 : 8499–8506.

Lettice LA, Horikoshi T, Heaney SJH, van Baren MJ, van der Linde HC, Breedveld GJ, Joosse M, Akarsu N, Oostra BA, Endo N, et al. 2002. Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. Proc Natl Acad Sci U S A 99: 7548–7553.

Liu C, Liu W, Lu MF, Brown NA, Martin JF. 2001. Regulation of left-right asymmetry by thresholds of Pitx2c activity. Development 128: 2039–2048.

Liu W, Ma Q, Wong K, Li W, Ohgi K, Zhang J, Aggarwal AK, Rosenfeld MG. 2013. Brd4 and JMJD6-associated anti-pause enhancers in regulation of transcriptional pause release. Cell 155: 1581–1595.

Logan M, Pagán-Westphal SM, Smith DM, Paganessi L, Tabin CJ. 1998. The transcription factor pitx2 mediates situs-specific morphogenesis in response to left-right asymmetric signals. Cell 94: 307–317.

Lu MF, Pressman C, Dyer R, Johnson RL, Martin JF. 1999. Function of Rieger syndrome gene in left-right asymmetry and craniofacial development. Nature 401: 276–278.

Mahadevan A, Welsh ICC, Sivakumar A, Gludish DWW, Shilvock ARR, Noden DMM, Huss D, Lansford R, Kurpios NAA. 2014. The Left-Right Pitx2 Pathway

Drives Organ-Specific Arterial and Lymphatic Development in the Intestine. Dev Cell 31: 690–706.

Marinić M, Aktas T, Ruf S, Spitz F. 2013. An Integrated Holo-Enhancer Unit Defines Tissue and Gene Specificity of the Fgf8 Regulatory Landscape. Dev Cell 24: 530–542.

Melgar MF, Collins FS, Sethupathy P. 2011. Discovery of active enhancers through bidirectional expression of short transcripts. Genome Biol 12:(11):R113.

Montavon T, Soshnikova N, Mascrez B, Joye E, Thevenet L, Splinter E, De Laat W, Spitz F, Duboule D. 2011. A regulatory archipelago controls hox genes transcription in digits. Cell 147: 1132–1145.

Nobrega MA, Ovcharenko I, Afzal V, Rubin EM. 2003. Scanning Human Gene Deserts for Long-Range Enhancers. Sci 302 : (5644):413.

Nora EP, Dekker J, Heard E. 2013. Segmental folding of chromosomes: A basis for structural and regulatory chromosomal neighborhoods? BioEssays 35: 818–828.

Phillips-Cremins JE, Sauria MEG, Sanyal A, Gerasimova TI, Lajoie BR, Bell JSK, Ong CT, Hookway TA, Guo C, Sun Y, et al. 2013. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell 153: 1281–1295.

Pina C, May G, Soneji S, Hong D, Enver T. 2008. MLLT3 Regulates Early Human Erythroid and Megakaryocytic Cell Fate. Cell Stem Cell 2: 264–273.

Rainger JK, Bhatia S, Bengani H, Gautier P, Rainger J, Pearson M, Ansari M, Crow J, Mehendale F, Palinkasova B, et al. 2014. Disruption of SATB2 or its long-range cis-regulation by SOX9 causes a syndromic form of Pierre Robin sequence. Hum Mol Genet 23(10):2569-2579.

Razin S V., Gavrilov AA, Pichugin A, Lipinski M, Iarovaia O V., Vassetzky YS. 2011. Transcription factories in the context of the nuclear and genome organization. Nucleic Acids Res 39: 9085–9092.

Reis LM, Tyler RC, Volkmann Kloss BA, Schilter KF, Levin A V, Lowry RB, Zwijnenburg PJG, Stroh E, Broeckel U, Murray JC, et al. 2012. PITX2 and FOXC1 spectrum of mutations in ocular syndromes. Eur J Hum Genet. 20(12): 1224-1233.

Rinn JL, Chang HY. 2012. Genome Regulation by Long Noncoding RNAs. Annu Rev Biochem 81: 145–166.

Ryan AK, Blumberg B, Rodriguez-Esteban C, Yonei-Tamura S, Tamura K, Tsukui T, de la Peña J, Sabbagh W, Greenwald J, Choe S, et al. 1998. Pitx2 determines left-right asymmetry of internal organs in vertebrates. Nature 394: 545–551.

Schaukowitch K, Joo J-Y, Liu X, Watts JK, Martinez C, Kim T-K. 2014. Enhancer RNA Facilitates NELF Release from Immediate Early Genes. Mol Cell 56: 29–42.

Semina E V, Reiter R, Leysens NJ, Alward WL, Small KW, Datson NA, Siegel-Bartelt J, Bierke-Nelson D, Bitoun P, Zabel BU, et al. 1996. Cloning and characterization of a novel bicoid-related homeobox transcription factor gene, RIEG, involved in Rieger syndrome. Nat Genet 14: 392–399.

Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, Kuan S, Wagner U, Dixon J, Lee L, Lobanenkov V V., et al. 2012. A map of the cis-regulatory sequences in the mouse genome. Nature 488: 116–120.

Shim EY, Walker AK, Shi Y, Blackwell TK. 2002. CDK-9/cyclin T (P-TEFb) is required in two postinitiation pathways for transcription in the C. elegans embryo. Genes Dev 16: 2135–2146.

Shiratori H, Hamada H. 2006. The left-right axis in the mouse: from origin to morphology. Development 133: 2095–2104.

Shiratori H, Sakuma R, Watanabe M, Hashiguchi H, Mochida K, Sakai Y, Nishino J, Saijoh Y, Whitman M, Hamada H. 2001. Two-step regulation of left-right asymmetric expression of Pitx2: Initiation by nodal signaling and maintenance by Nkx2. Mol Cell 7: 137–149.

Shiratori H, Yashiro K, Shen MM, Hamada H. 2006. Conserved regulation and role of Pitx2 in situs-specific morphogenesis of visceral organs. Development 133: 3015–3025.

Shopland LS, Lynch CR, Peterson KA, Thornton K, Kepper N, Von Hase J, Stein S, Vincent S, Molloy KR, Kreth G, et al. 2006. Folding and organization of a contiguous chromosome region according to the gene distribution pattern in primary genomic sequence. J Cell Biol 174: 27–38.

Spitz F, Gonzalez F, Duboule D. 2003. A global control region defines a chromosomal regulatory landscape containing the HoxD cluster. Cell 113: 405–417.

Splinter E, Heath H, Kooren J, Palstra RJ, Klous P, Grosveld F, Galjart N, De Laat W. 2006. CTCF mediates long-range chromatin looping and local histone modification in the ??-globin locus. Genes Dev 20: 2349–2354.

Steger DJ, Lefterova MI, Ying L, Stonestrom AJ, Schupp M, Zhuo D, Vakoc AL, Kim J- E, Chen J, Lazar MA, et al. 2008. DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously coupled with gene transcription in mammalian cells. Mol Cell Biol 28: 2825–2839.

Sutherland H, Bickmore WA. 2009. Transcription factories: gene expression in unions? Nat Rev Genet 10: 457–466.

Tao Y, Zhang M, Li L, Bai Y, Zhou Y, Moon AM, Kaminski HJ, Martin JF. 2014. Pitx2, an atrial fibrillation predisposition gene, directly regulates ion transport and intercalated disc genes. Circ Cardiovasc Genet 7: 23–32.

Van de Werken HJG, Landan G, Holwerda SJB, Hoichman M, Klous P, Chachik R, Splinter E, Valdes-Quezada C, Öz Y, Bouwman BAM, et al. 2012. Robust 4C-seq data analysis to screen for regulatory DNA interactions. Nat Methods 9: 969–972.

Vance KW, Ponting CP. 2014. Transcriptional regulatory functions of nuclear long noncoding RNAs. Trends Genet 30: 348–355.

Visel A, Minovitsky S, Dubchak I, Pennacchio LA. 2007. VISTA Enhancer Browser - A database of tissue-specific human enhancers. Nucleic Acids Res 35 (Database issue):D88-92.

Volkmann BA, Zinkevich NS, Mustonen A, Schilter KF, Bosenko D V., Reis LM, Broeckel U, Link BA, Semina E V. 2011. Potential novel mechanism for Axenfeld-Rieger syndrome: Deletion of a distant region containing regulatory elements of PITX2. Investig Ophthalmol Vis Sci 52: 1450–1459.

Waite MR, Skidmore JM, Micucci JA, Shiratori H, Hamada H, Martin JF, Martin DM. 2013. Pleiotropic and isoform-specific functions for Pitx2 in superior colliculus and hypothalamic neuronal development. Mol Cell Neurosci 52: 128–139.

Wang J, Bai Y, Li N, Ye W, Zhang M, Greene SB, Tao Y, Chen Y, Wehrens XHT, Martin JF. 2014. Pitx2-microRNA pathway that delimits sinoatrial node development and inhibits predisposition to atrial fibrillation. Proc Natl Acad Sci U S A 111: 9181–9186.

Watanabe H, Schmidt HA, Kuhn A, Hoger SK, Kocagoz Y, Laumann-Lipp N, Ozbek S, Holstein TW. 2014. Nodal signalling determines biradial asymmetry in Hydra. Nature 515 (7525):112-115

Welsh IC, Thomsen M, Gludish DW, Alfonso-Parra C, Bai Y, Martin JF, Kurpios N a. 2013. Integration of left-right Pitx2 transcription and Wnt signaling drives asymmetric gut morphogenesis via Daam2. Dev Cell 26: 629–644.

Williamson I, Berlivet S, Eskeland R, Boyle S, Illingworth RS, Paquette D, Dostie J, Bickmore WA. 2014. Spatial genome organization: contrasting views from chromosome conformation capture and fluorescence in situ hybridization. Genes Dev 28 : 2778–2791.

Zorn AM, Wells JM. 2009. Vertebrate endoderm development and organ formation. Annu Rev Cell Dev Biol 25: 221–251.

Zuniga A, Michos O, Spitz F, Haramis A-PG, Panman L, Galli A, Vintersten K, Klasen C, Mansfield W, Kuc S, et al. 2004. Mouse limb deformity mutations disrupt a

global control region within the large regulatory landscape required for Gremlin expression. Genes Dev 18: 1553–1564.

CHPT 2 SUPPLEMENTARY FIGURES

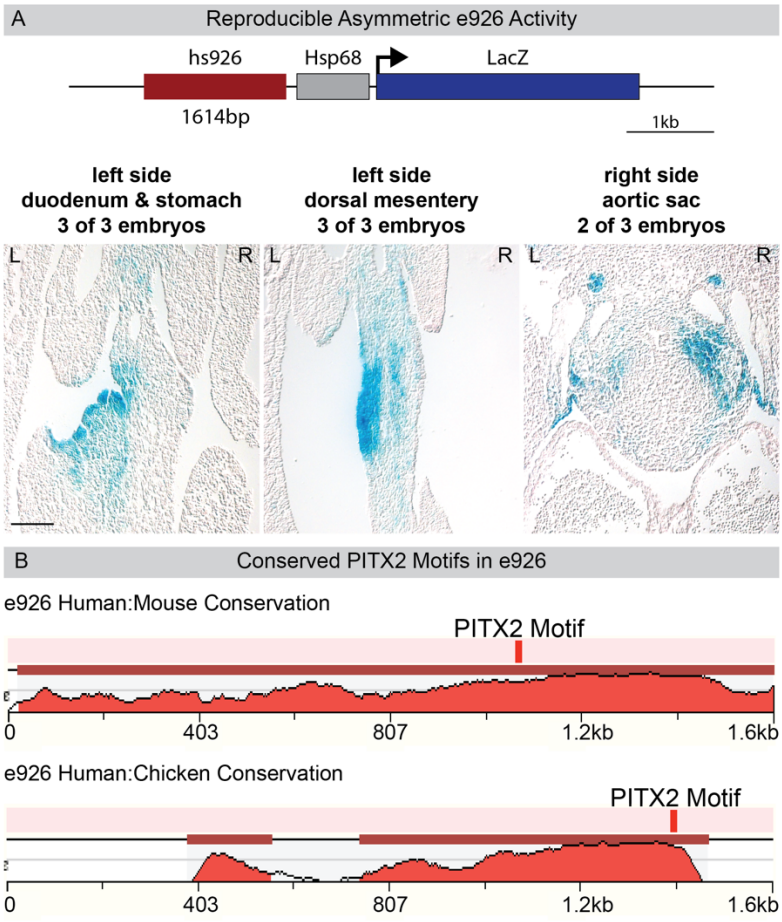


Fig 2-S1. Reproducible domains of asymmetric enhancer activity of hs926 by in vivo transgenic analysis. Related to Figure 1 and 3. (A) Human hs926 sequence cloned into the Hsp68-LacZ reporter construct (top) shows reproducible enhancer activity in asymmetric domains of transgenic embryos (bottom). (B) Prediction of *Pitx2* binding sites within e926 conserved between human and mouse (top) or human and chicken (bottom) (http://rvista.dcode.org/).

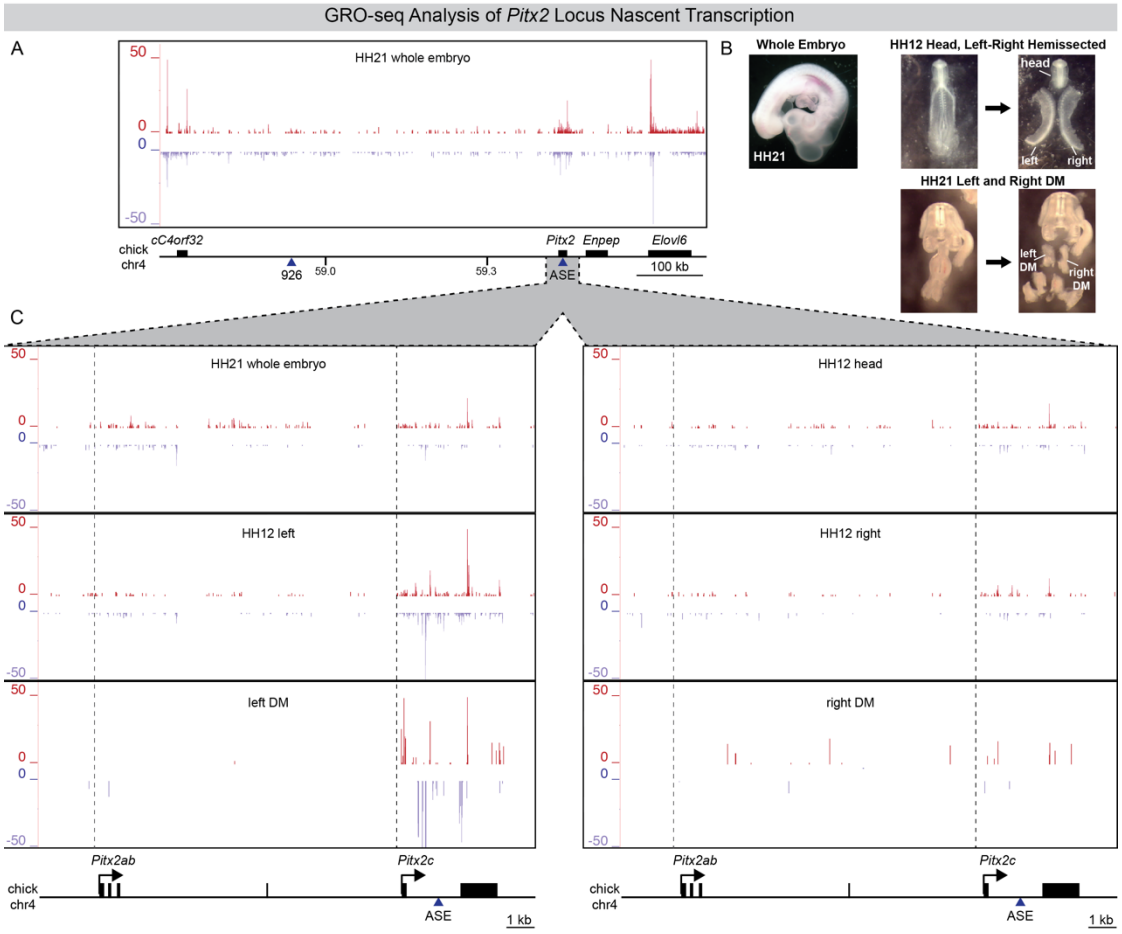


Figure 2-S2. Annotation of asymmetric *Pitx2* locus transcription in vivo using GRO- seq. Related to Figure 2. (A) GRO-seq of HH21 whole embryo characterizes nascent transcription across the *Pitx2* locus. (B) Whole embryo, HH12 head, and HH12 left & right hemisected samples were collected in addition to HH21 left and right DM samples to monitor enrichment and specificity of asymmetric transcription. (C) Transcription of bilaterally expressed *Pitx2ab* isoforms is observed in HH21 whole embryo and HH12 head. Elevated *Pitx2* expression is observed in the HH12 left hemisected sample compared to the right, while only the left DM sample detects exclusive *Pitx2c* expression. Dashed vertical lines mark *Pitx2ab* and *Pitx2c* TSS.

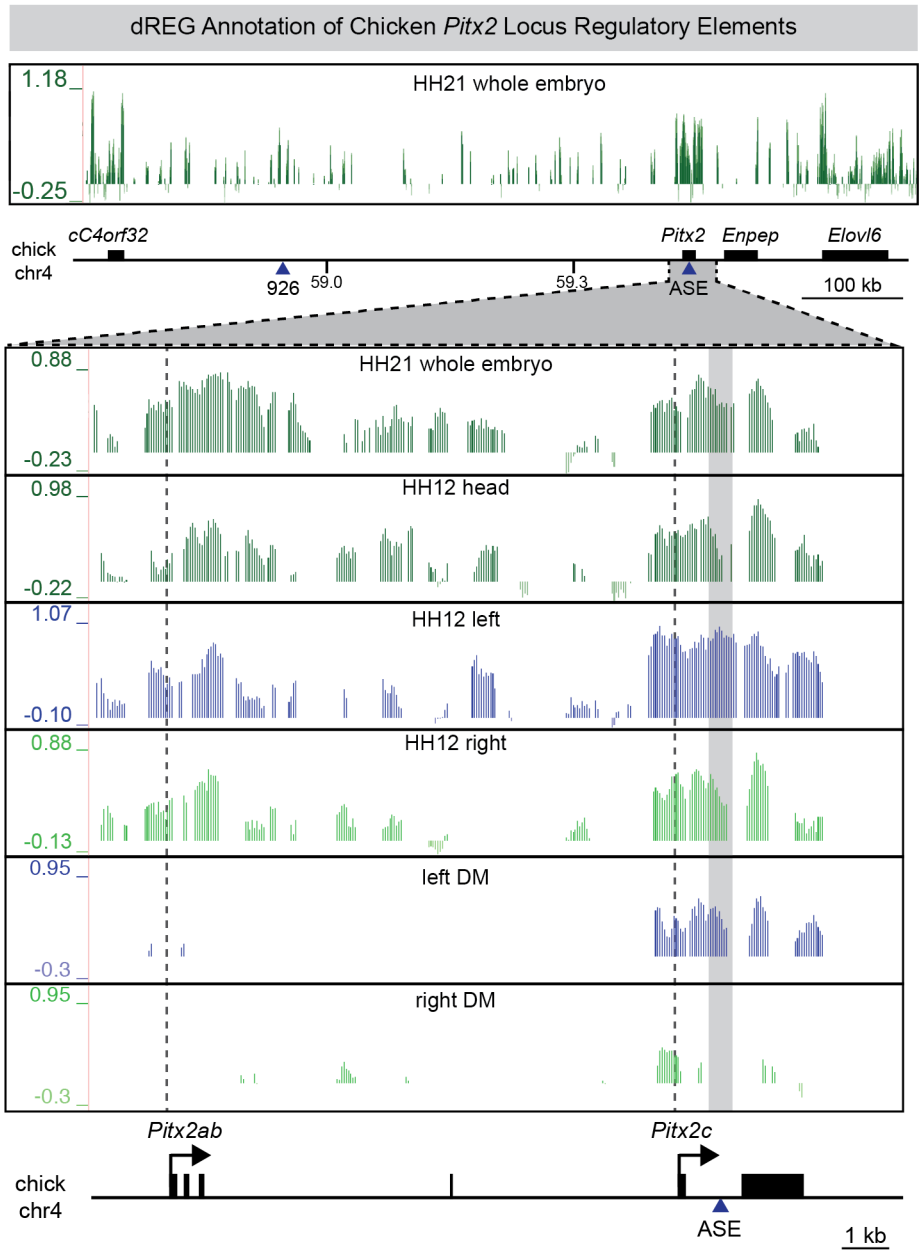


Figure 2-S3. Annotation of active regulatory elements at the *Pitx2* locus in vivo using dREG. Related to Figure 2. Top: dREG annotation of the *Pitx2* locus from GRO-seq of HH21 whole embryo sample identifies a considerable number of distal regulatory elements distributed across the gene desert in addition to peaks more closely associated with coding genes. Bottom: dREG peaks in the left DM sample demonstrates enrichment of asymmetric *Pitx2c* specific peaks and in the right DM shows absence of ASE activity. Dashed vertical lines mark *Pitx2ab* and *Pitx2c* TSS; grey box marks the ASE.

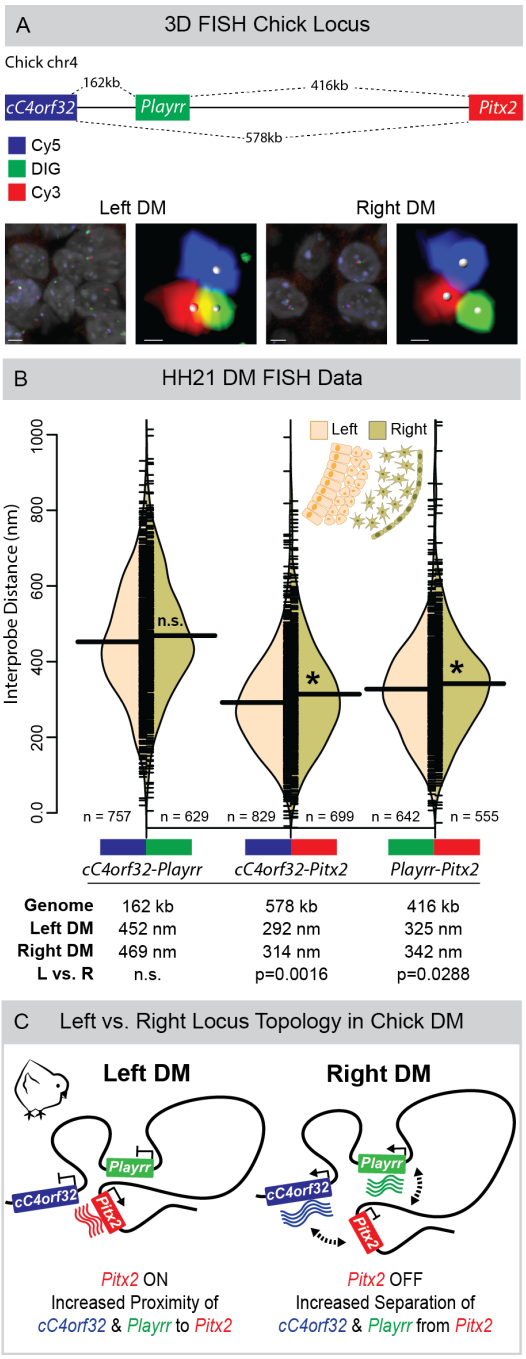


Figure 2-S4. Reproducible L-R differences in chicken and mouse FISH data. Related to Figure 4 and 5. (A) Increased proximity in the left DM between *cC4orf32-Pitx2* and *Playrr-Pitx2* in chicken, or *Playrr-Pitx2* in mouse, is highly reproducible across biological replicates. Box plots show the lower quartile, median, upper quartile, and whiskers show ±1.5 times the interquartile range. Contour plotting the density of the underlying distribution of the data are color coded for left (tan) and right (green) DM. (B) Box plots showing *Playrr-Pitx2* interprobe distances in the left DM of *Pitx2* -/- embryos is indistinguishable from the right DM of WT embryos. (C) TaqMan QRT-PCR of *Playrr* and *Pitx2* isoform expression levels in *Pixt2* -/- compared to WT.

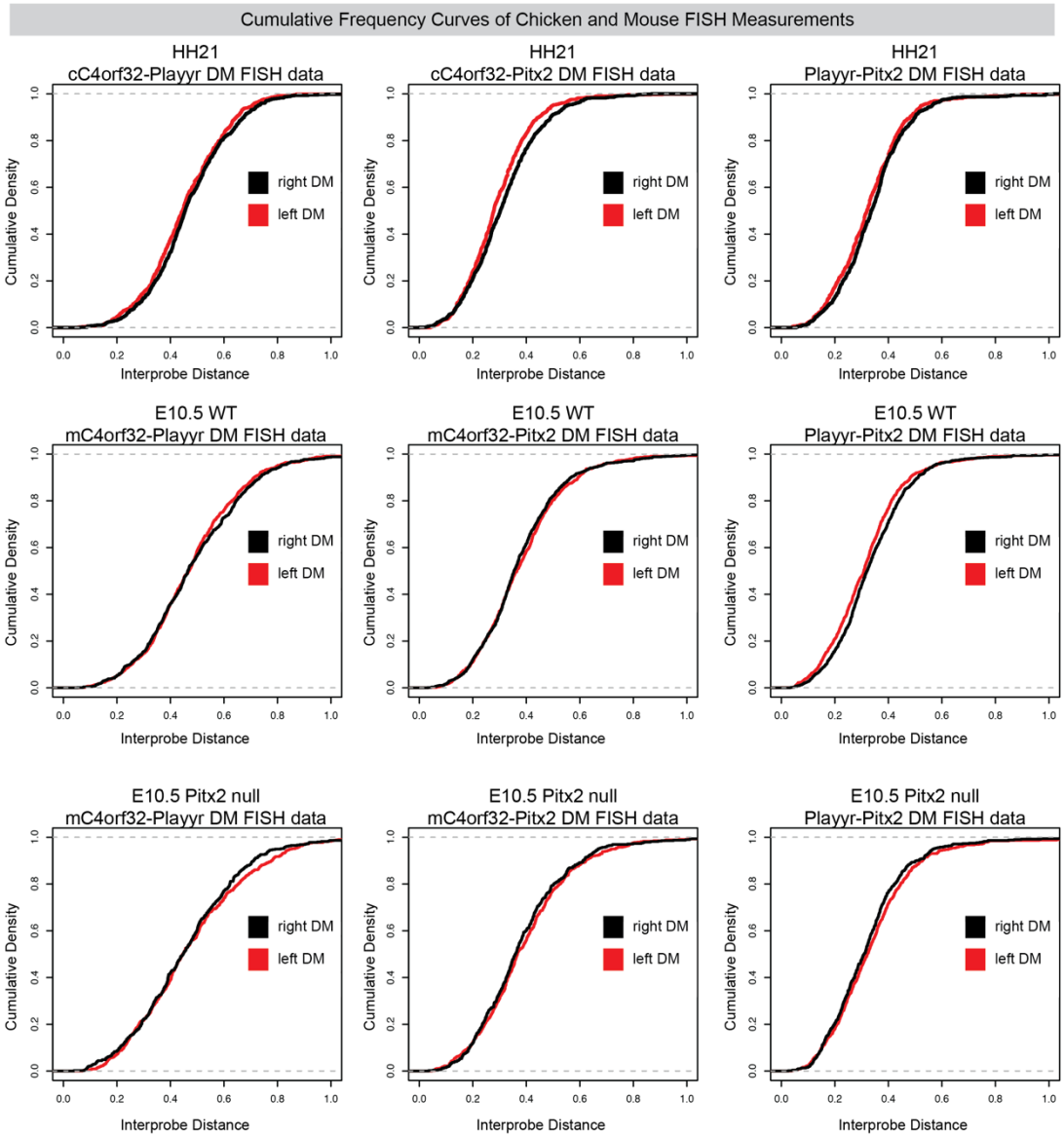


Figure 2-S5. Cumulative density curves of chicken and mouse FISH data. Related to Figure 4 and 5. Empirical cumulative distribution frequency (ECDF) plots generated in R show cumulative density (y-axis) of FISH interprobe distance (x-axis) for each probe pair in the left or right DM measured in chicken and mouse.

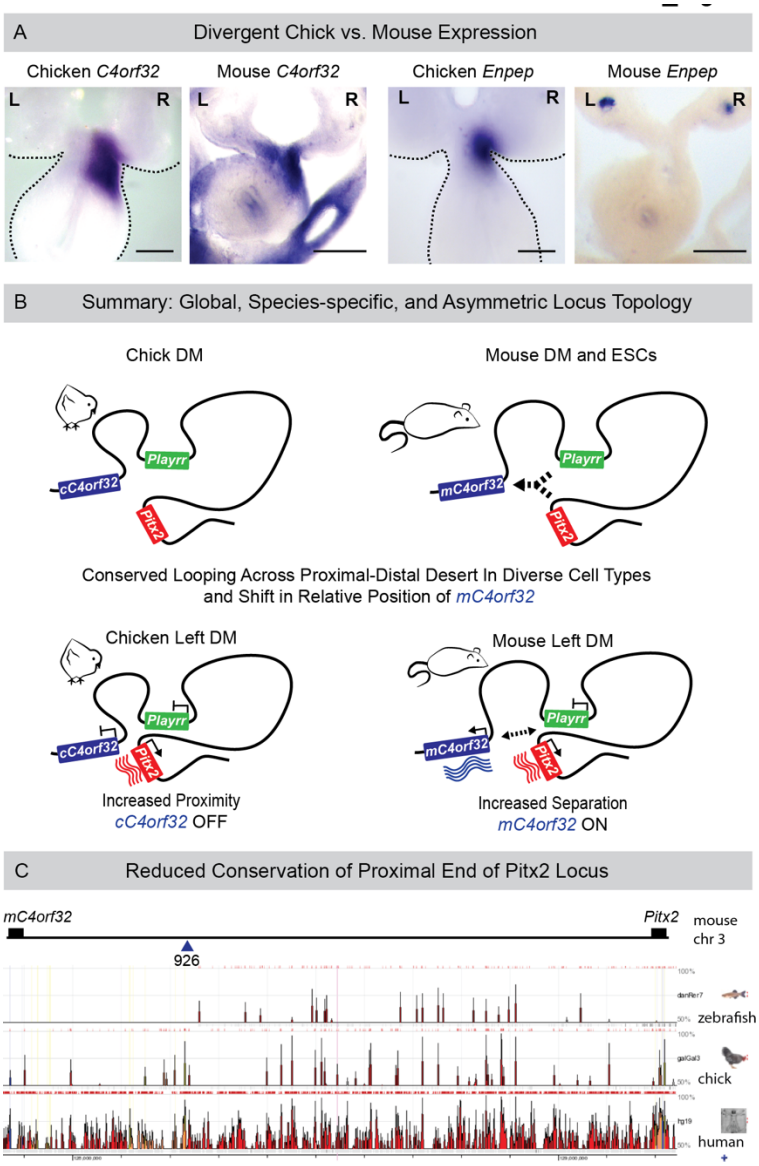


Figure 2-S6. Conserved and divergent features of the *Pitx2* locus in chicken and mouse. Related to Figure 4, 5, and 6 (A) Increased separation of *mC4orf32* from *Playrr* and *Pitx2* in mice (compare Fig. 4 and 5), and TAD structure of the *Pitx2* locus in mESCs (Fig. 6) showing that *Pitx2* and *Enpep* are positioned within separate regulatory domains is accompanied by divergent chicken vs. mouse *C4orf32* and *Enpep* DM expression (B) Summary models of conserved and divergent locus topology. (C) Comparison of conserved sequences from the mouse *Pitx2* locus with zebrafish, chicken, and human, supports that evolutionary divergence is greatest at the proximal end of the gene desert. Note that Zebrafish lack orthologs of *C4ORF32*, e926/*Playrr, ENPEP*.

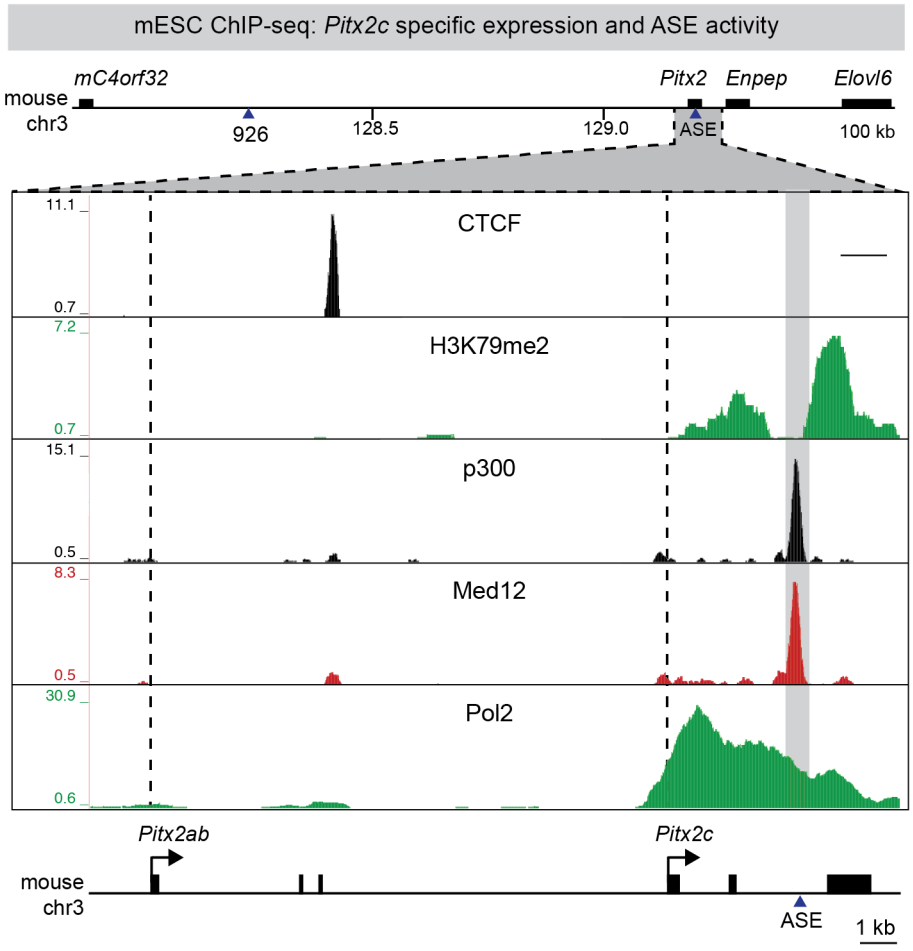


Figure 2-S7. ChIP-seq analysis of mouse embryonic stem cells (mESCs). Related to Figure 6. RNA Pol II ChIP-seq data demonstrate that exclusive expression of the asymmetric *Pitx2c* isoform in mESCs is accompanied by binding at the ASE enhancer by positive regulators of transcriptional activation such as Med12, p300 and chromatin marks including H3K79 methylation deposited by the DOT1L transcriptional elongation complex via association with RNA Pol II.

CHAPTER 3

This chapter represents a first author manuscript in preparation to be submitted in 2018.

CRISPR/CAS9 GENOME EDITING AT THE ATRIAL FIBRILLATION (AF)-ASSOCIATED *PITX2* LOCUS: THE ROLE OF THE LONG NONCODING RNA *PLAYRR* IN CARDIAC ARRHYTHMIA

Frances L. Chen[1], Eva M. Oxford[1], John P. Leach[2], Christina Cong[1], Ian C. Welsh[3], Erin Daugherity[4,5], Jeffrey T. Pea[3], Sophie A. Kupliec-Weglinski[1], James F. Martin[2], and Natasza A. Kurpios[1]*

Author Affiliations:

[1]Department of Molecular Medicine, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA

[2]Department of Molecular Physiology and Biophysics, Baylor College of Medicine, Houston, TX 77030, USA

[3]Department of Orofacial Sciences, School of Dentistry, University of California San Francisco, San Francisco, CA 94143, USA

[4]Department of Biomedical Sciences, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA

[5]Center for Animal Resources and Education, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA

Author Contributions:

F.L.C, E.M.O, and N.A.K designed the study. F.L.C performed all experiments and data analysis with additional contributions from: E.M.O for analysis of AliveCor ECG; E.M.O with guidance from E.K.D for study design and conducting 24 hour telemetry and restraint stress ECG experiments (Fig. 4 & Fig. 5); C.C. for WISH and sectioning (Fig. 2D); ICW for WISH (Fig. 2A); J.P.L and J.F.M for programmed stimulation and pacing induced AF data (Fig. 8); J.T.P and S.A.K for assistance with AliveCor ECG (Fig. 3B). F.L.C and N.A.K wrote the manuscript with input from coauthors.

3.1 Abstract

Atrial Fibrillation (AF) is the most common arrhythmia in humans and increases risk of stroke, cardiac arrest, and death. AF has a significant genetic contribution to disease risk, but the genetic mechanisms leading to AF are unclear. The most significantly associated AF risk loci map to variants in a noncoding gene desert upstream of the *Pitx2* locus on human 4q25, suggesting a role for cis-regulation of *Pitx2* in the development of AF. *Pitx2* is required for left-right (LR) heart morphogenesis and loss of *Pitx2* leads to increased susceptibility to AF, development of bilateral sinoatrial nodes (SAN), and sinus node dysfunction (SND), a bradyarrhythmia and risk factor for the development of AF. Here we investigate a role for the long noncoding (lnc) RNA *Playrr* in modulating *Pitx2* transcription in the context of heart morphogenesis and cardiac arrhythmia.

Playrr is a conserved enhancer-associated lncRNA located within the *Pitx2* locus gene desert and expressed asymmetrically (*on the right*) to *Pitx2 (on the left)* in the developing mouse and chicken embryo, including in the embryonic SAN domain in the mouse. To investigate the biological role of *Playrr*, we used CRISPR/Cas9 genome editing in mice to target the *Playrr* RNA transcript while leaving its associated DNA cis regulatory element intact. By adapting a surface ECG device (AliveCor) and telemetry ECG we have detected cardiac arrhythmias in *Playrr* mutant mice. We demonstrate that adult *Playrr* mutant mice exhibit bradycardia and irregular R-R intervals that are indicative of SND. Finally, programmed stimulation of *Playrr* mutant mice reveals that loss of *Playrr* predisposes to pacing-induced AF. These phenotypes are strikingly similar to those previously observed in *Pitx2* heterozygous mice and suggest that *Playrr* modulates *Pitx2* expression leading to cardiac conduction abnormalities.

3.2 Introduction

3.2.1 Atrial fibrillation and sinus node dysfunction

Atrial Fibrillation (AF) is the most common arrhythmia worldwide in humans, with its prevalence increasing by 5% per year in patients over 65 years (Benjamin et al., 2009). In AF, ectopic electrical activity (ie an initiating *trigger*) originating outside the sinoatrial node (SAN), the dominant pacemaker of the heart, interacts with altered atrial tissue (*substrate*) to cause uncontrolled, abnormal firing of electrical impulses and subsequent quivering (*fibrillation*) of the atria. This irregular and rapid arrhythmia can then lead to dangerous thromboembolic events and other heart-related

complications (Camm et al., 2010). In fact, AF is the most important risk factor for stroke and predisposes patients to a number of other serious and potentially fatal conditions, including dementia and heart failure (Lloyd-Jones et al., 2004; Benjamin et al., 2009).

Sinus node dysfunction (SND) is defined by abnormal impulse formation and propagation from the SAN, resulting in bradycardia, syncope, and pacemaker implantation (Roberts-Thomson, Sanders, & Kalman, 2007). SND is a major risk factor for AF and vice versa; in patients with SND, over 50% subsequently develop AF, while in patients with AF, subsequent pathological remodeling of the atrial substrate has been associated with SND (Chang et al., 2013). Additionally, SND is associated with AF progression (Roberts-Thomson et al., 2007; Zhao 2014; Liu et al., 2014). Interestingly, SND is most often accompanied by changes to atrial substrate in the *right* atrium, where the SAN resides.

Both SND and AF are arrhythmias that occur and act upon abnormal atrial triggers and substrates, and can thereby initiate and perpetuate each other (John and Kumar, 2016; Lloyd-Jones et al., 2004). The molecular and physiological mechanisms linking these two arrhythmias are complex and remain unclear. However, understanding the genetic basis of these two important diseases provides a powerful framework for discovering and therapeutically targeting the molecular mechanisms underlying cardiac arrhythmias.

3.2.2 *Pitx2* plays multiple roles in cardiac development and postnatal function

The *Pitx2* transcription factor is a central driver of LR asymmetric cardiac morphogenesis and loss of asymmetric *Pitx2* expression in the left lateral plate mesoderm (LPM) of the early embryo results in LR laterality defects, including cardiac malformations (Logan et al., 1998; Ryan et al., 1998). *Pitx2* plays a critical role in the asymmetric development of the SAN exclusively in the right atrium. The SAN develops from a *Tbx18+/Nkx2.5-* genetic lineage located in the splanchnic mesoderm-derived sinus venosus (SV) myocardium, the embryonic venous inflow tract (Mommersteeg et al., 2007). During development, asymmetric *Pitx2* expression in the left SV directly represses the SAN genetic program on the left, and loss of *Pitx2* in the embryo results in right atrial isomerism with bilateral SAN (**Fig. 1**) (Mommersteeg et al., 2007; Wang et al., 2010; Amirabile et al., 2012; Wang et al., 2014). In both humans and mice, *Pitx2* expression is maintained in the adult left atrium (Tao et al., 2014; Kirchhof et al., 2011; Chinchilla et al., 2011) and directly regulates genes integral for cardiac function, including intercalated disc (ID) proteins and ion channel genes essential for proper cardiac conduction and electromechanical coupling (Tao et al., 2014). Studies in multiple independent mutant mouse models have shown that both partial reduction and complete loss of cardiac *Pitx2* expression leads to cardiac arrhythmias, including SND and pacing-induced AF (Wang et al., 2010; Tao et al 2014; Kirchof et al., 2011; Chinchila et al., 2011). Importantly, both decreased and increased *Pitx2* levels have been linked to cardiac tissue of human AF patients (Chinchilla et al., 2011; Perez-Hernandez et al., 2016). Finally, there is a dose-dependent relationship between *Pitx2* transcriptional levels and that of AF

susceptibility genes (*Zfhx3, Kccn3, Wnt8a*) and calcium handling genes (*Atp2a2, Casq2, Plb*) (Lozano-Velasco et al., 2015), indicating that *Pitx2* gene dosage plays critical roles in susceptibility to atrial arrhythmias.

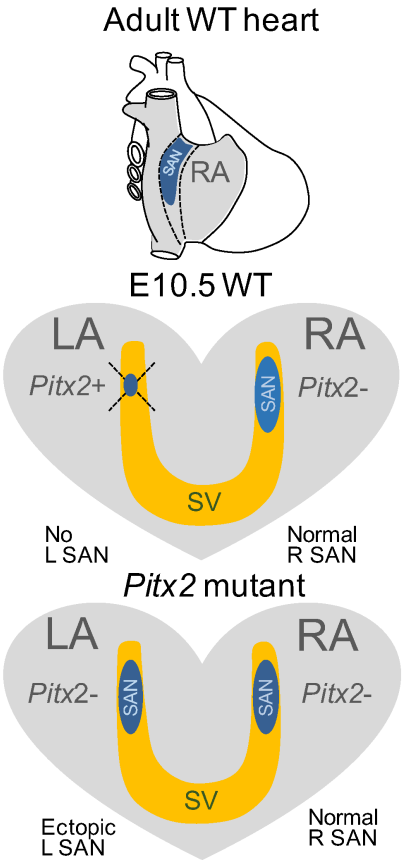


Fig. 3-1: Schematic of SAN location in adult heart and LR asymmetric patterning during development. The single right sided SAN is located at the junction of the right atrium and superior vena cava in the adult. *Pitx2* expressed in the left sinus venosus (SV) domain prevents activation of SAN genetic program and restricts SAN development to a single, asymmetric right sided SAN in wild type embryos. In *Pitx2* mutant embryos, loss of *Pitx2* on the left leads to ectopic L SAN, SND, and AF. SAN: sinoatrial node; SV: sinus venosus; LA: left atrium; RA: right atrium.

3.2.3 The most significantly associated AF risk variants map to noncoding sequence on 4q25 at the *Pitx2* locus

Consistent with the critical roles *Pitx2* plays in cardiac structure and function, multiple genome wide association studies have consistently mapped the most significantly associated risk variants for AF to human 4q25 locus within the noncoding gene desert just 170 kb upstream of *Pitx2* (Gudbjartsson, et al., 2007; Sinner et al., 2011, Ellinor et al., 2012; Christopherson et al., 2017). Additionally, recent work has demonstrated regulatory activity of this noncoding region, and using 3C based assays, showed physical interaction of this region with the *Pitx2c* promoter (Aguirre et al., 2015). These findings suggest that noncoding variants regulating *Pitx2* transcription may provide a cis-regulatory mechanism for understanding the genomic basis for AF. However, to date there are no published functional studies demonstrating that noncoding elements located within the *Pitx2* gene desert transcriptionally regulate *Pitx2* and/or are causal for AF. Having available the tools and in vivo models to investigate the functional relevance of these noncoding elements represents a critical opportunity in potentially bridging transcriptional regulation of *Pitx2* and cardiac arrhythmias.

3.2.4 Use of *Playrr-Ex1sj* mutant mouse model to investigate the genetic mechanisms of *Pitx2* linked arrhythmias

Here we sought to identify the transcriptional and developmental functions *Playrr*. In doing so, we aimed to investigate the mechanisms by which noncoding regulatory elements at the *Pitx2* locus may mediate transcriptional regulation of *Pitx2* necessary

for cardiac function. We discovered that *Playrr* is expressed in a right-specific pattern in the atria opposite the left-specific *Pitx2*, including in the SAN domain. We performed transcriptional and physiological analyses in our novel CRISPR/Cas9 generated mutant mouse model, $Playrr^{Ex1sj}$, a splice-site mutation that selectively targets the *Playrr* lncRNA transcript without perturbing its underlying DNA cis regulatory element (e926) (Welsh et al., 2015). Using both surface ECG screening and telemetry ECG, we demonstrate that $Playrr^{Ex1sj}$ mutant mice have resting and stressed cardiac arrhythmias evident of SND, a major predisposing risk factor for AF. Finally, we demonstrate that $Playrr^{Ex1sj}$ mutant mice, like loss of function *Pitx2* mutants, are predisposed to pacing induced AF in comparison to control mice. To our knowledge, this work represents the first study implicating a biological function for a lncRNA in modulation of cardiac rhythm.

3.3 Results

3.3.1 *Playrr* is expressed in cardiac domains opposite and asymmetric to Pitx2, in a highly tissue-specific region of SAN patterning

We previously discovered a conserved lncRNA, *Playrr* (Welsh et al., 2015) transcribed from a large intergenic gene desert located approximately 968 kb upstream of the *Pitx2* locus, on mouse chromosome 3. We found that the expression of this lncRNA was right-sided, asymmetric and complementary to that of *Pitx2* in the lateral plate mesoderm, the precursor tissue that gives rise to the sinus venous, the site of SAN development (**Fig. 2A**). In order to investigate cardiac-specific expression and

function of *Playrr*, I performed qRT-PCR specifically at E10.5 during LR patterning of the heart and at E14.5 embryos, when LR atrial development is complete. I found that *Playrr* expression was enriched in the heart (16.9 fold enrichment compared to E10.5 whole embryos) and expressed at similar levels when comparing both stages. Additionally, I confirmed loss of *Playrr* expression in $Playrr^{Ex1sj}$ mutant embryonic organs at both embryonic stages (**Fig. 2B**). In order to identify *Playrr* expression domains relevant to *Pitx2* expression and LR patterning of the SAN, I performed RNA in situ hybridization (ISH) in E10.5 whole embryos and showed that *Playrr* expression continues to be asymmetric and complementary to that of *Pitx2* in the heart, specifically in the mesenchyme and mesothelium of the right lung bud, and splanchnopleure derived mesenchyme of the right pleuropericardial fold (**Fig. 2C**). Excitingly, I also discovered that *Playrr* is expressed in a highly restricted population of cells in the contralateral domain of *Pitx2* expression on the right side in the region of SAN development, suggesting that *Playrr* may play a role in SAN patterning (**Fig. 2D).**

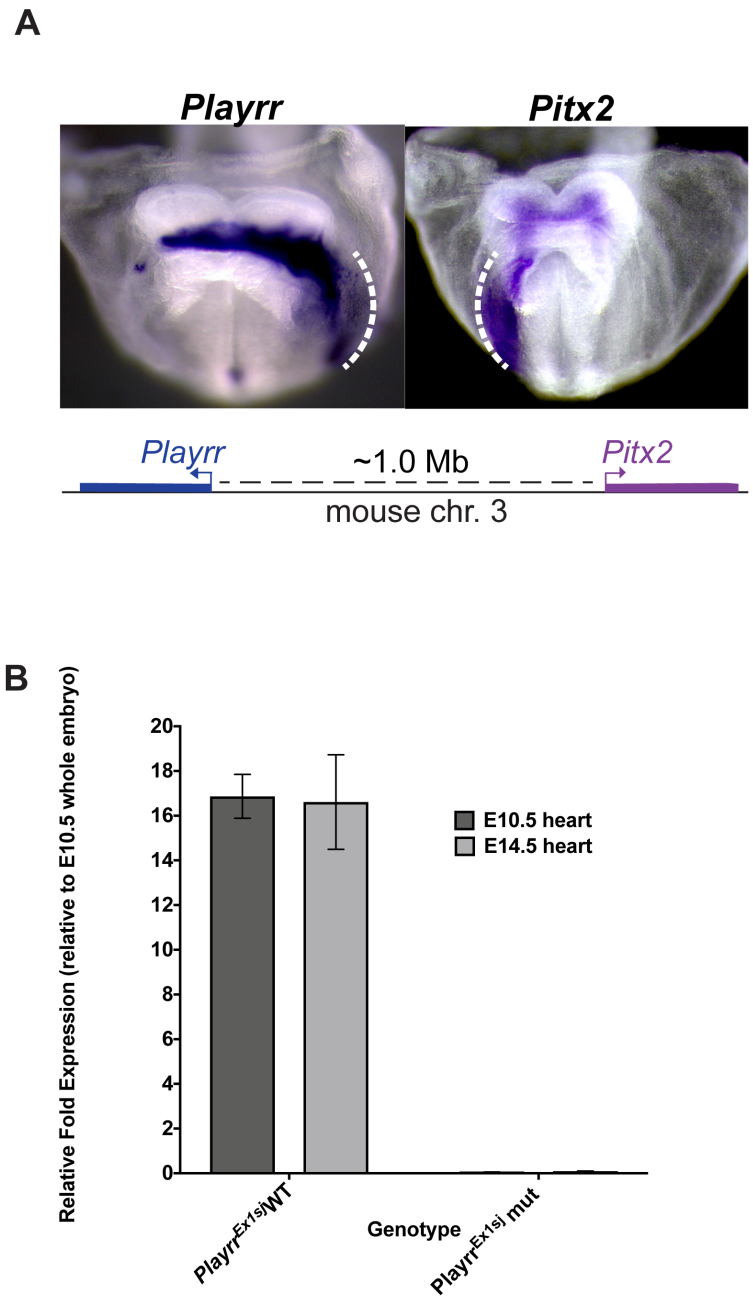


Fig. 3-2A-B: *Pitx2* and *Playrr* have asymmetric and complementary expression at E10.5 in the LR domain of asymmetrical SAN patterning
(A) *Pitx2* and *Playrr* physical linkage in the genome and asymmetric expression at E8.25. White dashed line denotes domain of expression in lateral plate mesoderm.
(B) qRT-PCR of *Playrr* expression in E10.5 and E14.5 hearts of WT and *Playrr*^{Ex1sj} mutants (to serve as negative control**).**

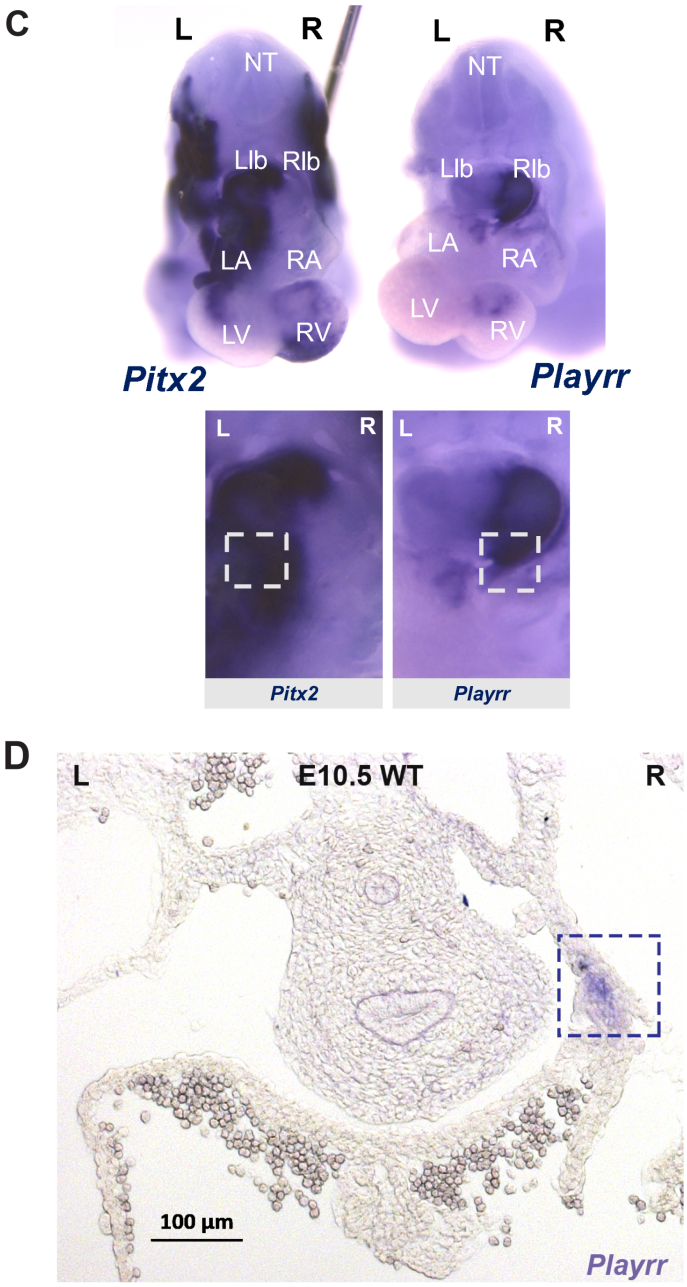


Fig. 2C-D: *Pitx2* and *Playrr* have asymmetric and complementary expression at E10.5 in the LR domain of asymmetrical SAN patterning

(C) Whole mount RNA in situ hybridization of asymmetric and complementary *Pitx2* and *Playrr* expression at E10.5 in the SAN domain. NT (neural tube); Llb (left lung bud); Rlb (right lung bud); LA (left atrium); RA (right atrium); LV (left ventricle); RV (right ventricle).

(D) Section of embryo from (**C**) showing *Playrr* expression in a discrete right-sided SAN at E10.5.

3.3.2 *Playrr*^{Ex1sj} mutant mice display stress induced cardiac arrhythmias

The asymmetric development of the SAN is essential for proper cardiac function and generation of the normal sinus rhythm (Mommersteeg et al., 2007). *Pitx2* plays key roles in LR patterning of the pro-pacemaking myocardial cells of sinus venosus preventing the formation of the SAN/pacemaker genetic program on the left, thereby ensuring the normal development of a single right-sided SAN (Ammirabile et al., 2012). Additionally, perturbed expression of *Pitx2* in mice and humans has been linked with abnormal SAN development and SND. Given that *Playrr* is expressed in the early SAN domain and that previous work suggests *Playrr* regulates *Pitx2* transcription in organ viscera (Welsh et al., 2015), we sought to investigate how loss of *Playrr* may affect electrical activity in the heart. By screening adult (12 week old) *Playrr*^{Ex1sj} mice for cardiac arrhythmias with an AliveCor ECG device, we discovered that awake, restrained *Playrr*^{Ex1sj} mutant mice have significantly reduced heart rates (bradycardia): (wild type and heterozygote controls (n=8) mean HR: 647.5±66.2 bpm; mutant (n=6) mean HR: 384.2±132 bpm). We also observed irregular duration between heartbeats (irregular R-R intervals) in 5/6 mutant mice vs 0/6 age matched heterozygote (het) controls, suggestive of SND. (**Fig. 3A-3C**).

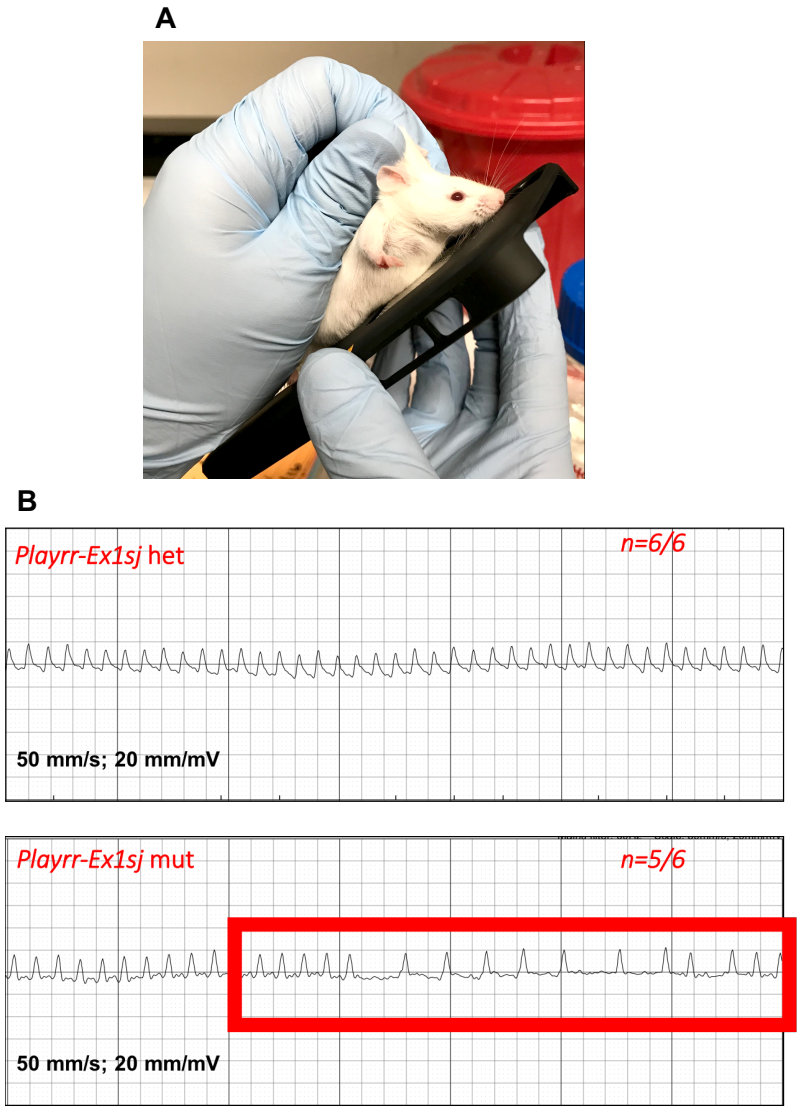


Fig. 3-3A-B: AliveCor surface ECG screening in awake, restrained *Playrr*^{Ex1sj} mice reveals bradycardia and irregular R-R intervals, evidence of SND
(A) Photograph of surface ECG device and screening of awake, restrained mice.
(B) ECG tracing from AliveCor screening in representative *Playrr*^{Ex1sj} het control (top) and mutant (bottom)**.**

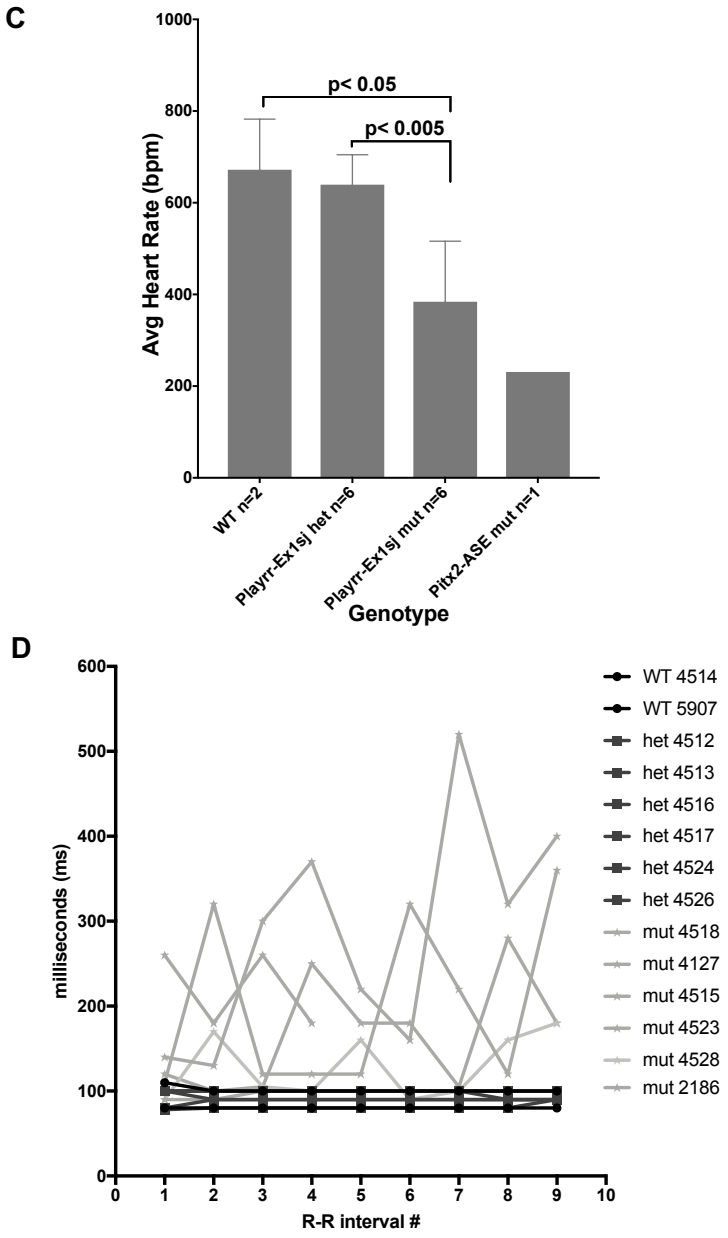


Fig. 3C-D: AliveCor surface ECG screening in awake, restrained *Playrr*^{Ex1sj} mice reveals bradycardia and irregular R-R intervals, evidence of SND
(C) Calculated mean heart rate (bpm) in *Playrr*^{Ex1sj} mice**.**
(D) Variation in beat to beat (R-R) interval in *Playrr*^{Ex1sj} mutants (grey lines) vs aged and age matched littermate WT and het controls (darker lines).

Importantly, we recapitulated this finding with 24 hour telemetry ECG and analyzed heart rate parameters to show that *Playrr*^{Ex1sj} mutant mice (n=4) have on average statistically significant decreased heart rate, increased R-R interval duration, and increased variation in R-R intervals compared to WT controls (n=3) (**Fig. 4A-4C**). Additionally, telemetry ECG during an acute restraint stress protocol (Zimprich et al., 2014) dramatically exacerbated the bradycardia and irregular R-R intervals (**Fig. 5**). Collectively, these findings strongly suggest that *Playrr*^{Ex1sj} mutant mice have SAN dysfunction, which mirrors ECG findings in loss-of-function *Pitx2* mutants and is a common comorbidity in human AF patients (Wang et al., 2010; Tao et al., 2010; Ammirabile et al., 2011; John and Kumar, 2016).

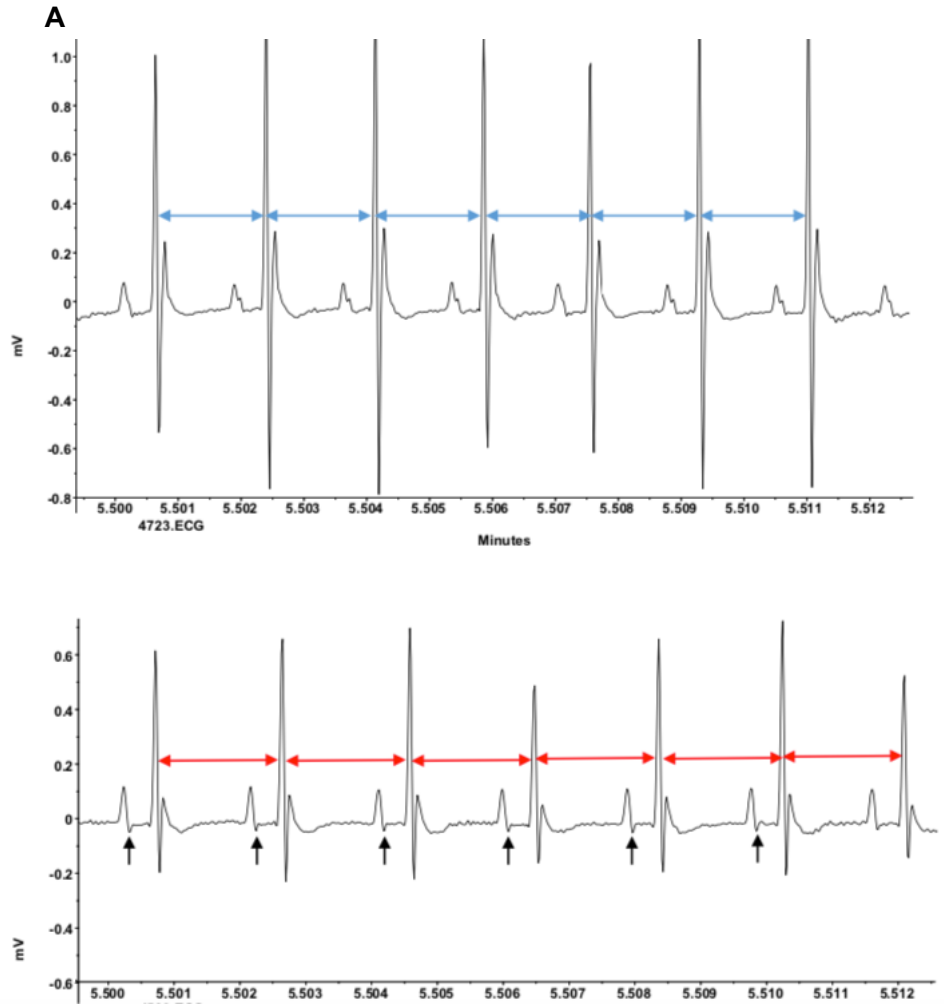


Fig. 3-4A: Heart Rate and Rhythm in 24 hour telemetry ECG recording in unstressed adult *Playrr*^{Ex1sj} mutant mice

(A) Representative ECG tracing in WT (top, n = 3) and mutant (bottom, n = 4) animals; double arrowheads mark R-R intervals. Black arrowheads point to biphasic P waves.

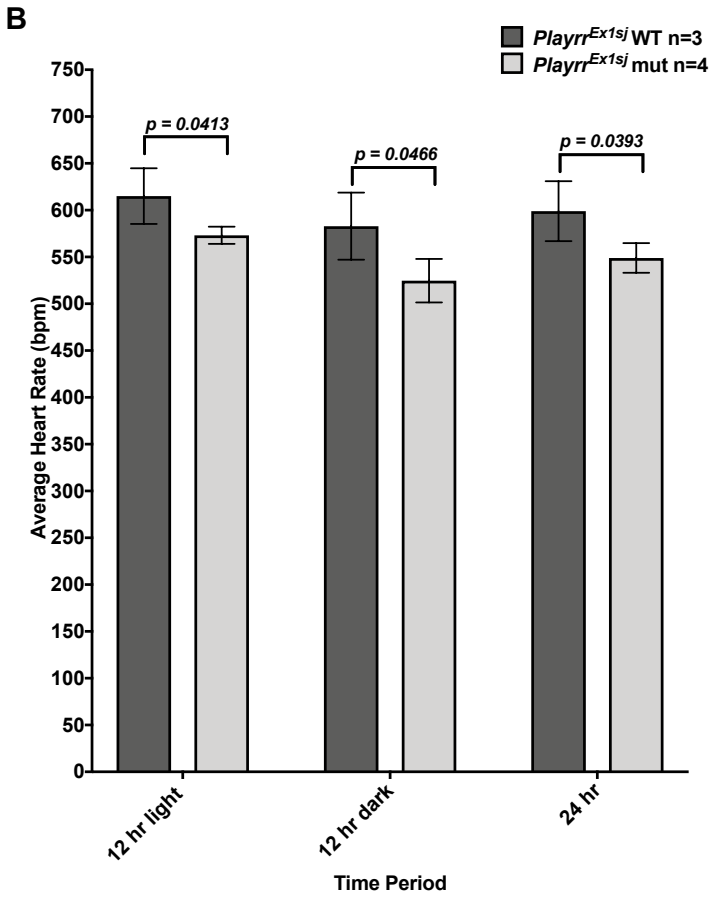


Fig. 3-4B: Heart Rate and Rhythm in 24 hour telemetry ECG recording in unstressed adult *Playrr*^{Ex1sj} mutant mice
(B) Mean heart rate (bpm) by 12 hour photoperiod and 24 hour average in WT vs mutant mice.

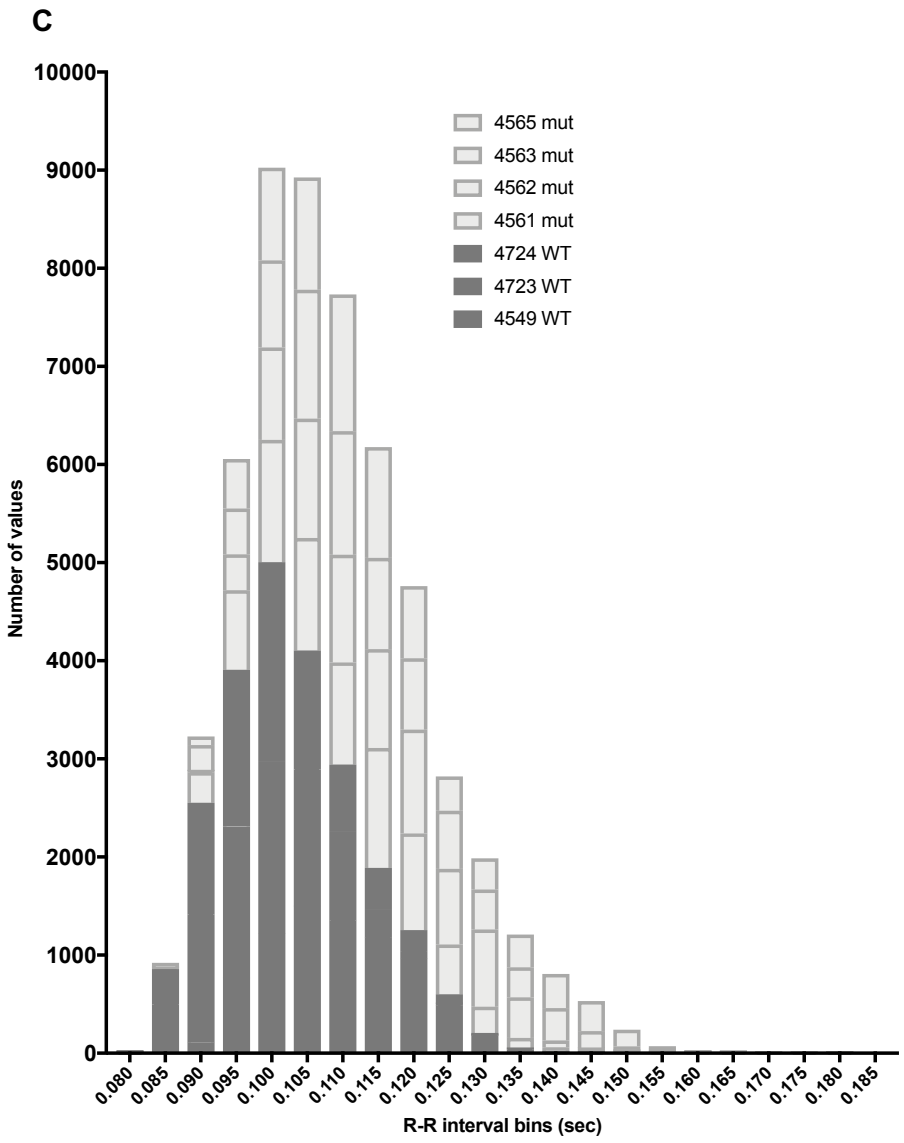


Fig. 3-4C: Heart Rate and Rhythm in 24 hour telemetry ECG recording in unstressed adult *Playrr*^{Ex1sj} mutant mice
(C) Frequency histogram of R-R interval duration in WT (n=3) vs mutant (n=4) animals.

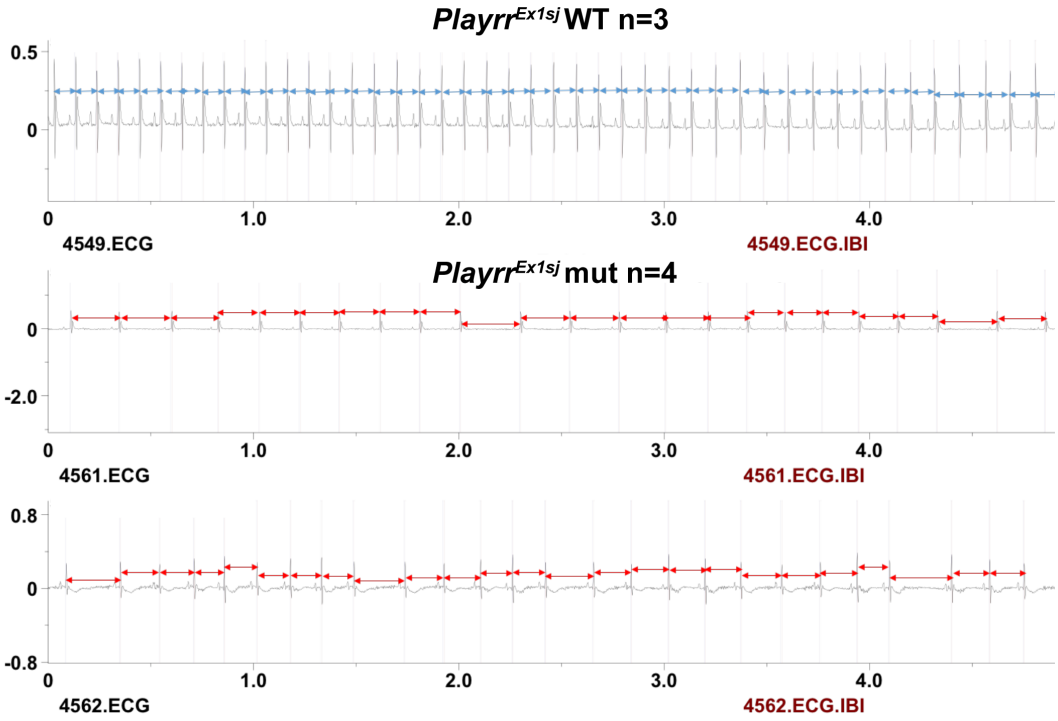


Fig. 3-5: Evidence of bradycardia and irregular R-R intervals in *Playrr*^{Ex1sj} mutant mice compared to WT controls when subjected to standard stress protocol. Representative telemetry ECG tracings of one WT (top) and two mutant mice (bottom) during 30 second restraint stress protocol.

3.3.3 Conserved LR Pitx2 expression in Playrr-Ex1sj mutant hearts

Because *Pitx2* is implicated in delimiting SAN development on the left (Wang et al., 2010; Ammirabile et al., 2012) and perturbation of *Pitx2* dosage during development leads to multiple structural and arrhythmogenic pathology in the heart (Liu et al., 2001; Franco et al., 2016), we asked if loss of *Playrr* leads to perturbed *Pitx2* expression during patterning of the SAN.

We first used ISH to qualitatively investigate *Pitx2* expression in the L and R sinoatrial patterning domains at E10.5. In *Playrr*^{Ex1sj} mutant embryos (n=5), normal left sided *Pitx2* expression was seen in the left SAN domain, left sinus venosus, and left atrial myocardium. No *Pitx2* expression was detected in any of the right sided corresponding domains (**Fig. 6A**). Using qRT-PCR we assayed individual *Playrr*^{Ex1sj} mutant and WT hearts (n=9) at E10.5 in *Playrr*^{Ex1sj} and did not detect any differences in global *Pitx2* expression at these stages, consistent with the lack of grossly visible structural abnormalities in *Playrr*^{Ex1sj} mutants **(Fig. 6B).** These data indicate that *Playrr*^{Ex1sj} mutant mice have normal LR asymmetric patterning of the SAN.

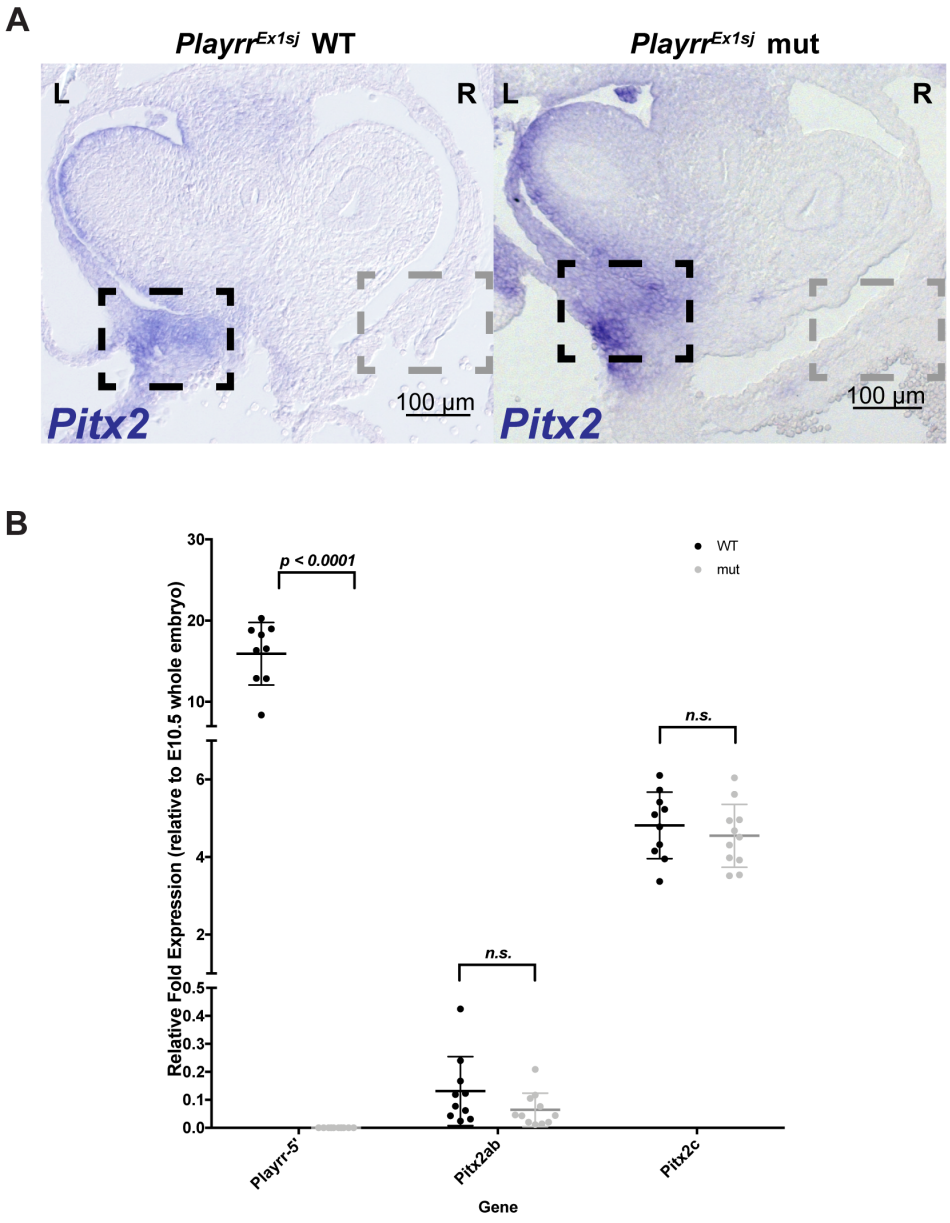


Fig. 6: *Pitx2* expression is conserved during LR patterning of the SAN at E10.5

(A) Sectioned whole mount RNA in situ hybridization for *Pitx2* in *Playrr*^{Ex1sj} WT (n=3) and mutant (n=5) embryos. Black dashed box demarcates L SAN domain; grey dashed box demarcates R SAN domain.

(B) *Pitx2* isoform expression assayed by qRT-PCR in microdissected E10.5 whole heart tissue of individual *Playrr*^{Ex1sj} WT vs mutant (mut) embryos (n=9).

3.5.4 Hcn4-LacZ expression in *Playrr*^{Ex1sj} mutant mice and WT controls

A key developmental and functional marker of the SAN is *Hcn4*, hyperpolarization-activated cyclic nucleotide-gated channel 4, responsible for mediating the pacemaker current (Steiber et al., 2003). Loss-of-function mutations of *Hcn4* cause varying manifestations of SND, ranging from sinus pauses and arrest to fatal bradycardia (Bucchi et al., 2013; Herrmann et al., 2013). Importantly, *Pitx2* has been shown to be an upstream transcriptional regulator of *Hcn4* during development (Wang et al., 2010; Amirabile et al., 2012). To determine if *Playrr*^{Ex1sj} mice have perturbed *Hcn4* expression during development, we crossed *Playrr*^{Ex1sj} mutant mice with those carrying the transgenic reporter allele, *Hcn4-CatCH*-IRES-*LacZ* (Cornell Heart Lung and Blood Resource for Optogenetic Signaling, CHROMus) and assayed for LacZ expression in E10.5 embryos. In *Playrr*^{Ex1sj} WT and *Playrr*^{Ex1sj} mutant *Hcn4-CatCH*-IRES-*LacZ* +transgenic embryos, patterns of LacZ staining are similarly found in the primary heart tube (left ventricle, first heart field), sinus venosus, LR venous myocardium, and in the right sided SAN, with no corresponding expression or structure on the left (n=3) **(Fig. 7A).** These data are in concordance with the normal, uninterrupted *Pitx2* expression in *Playrr*^{Ex1s} mutant mice.

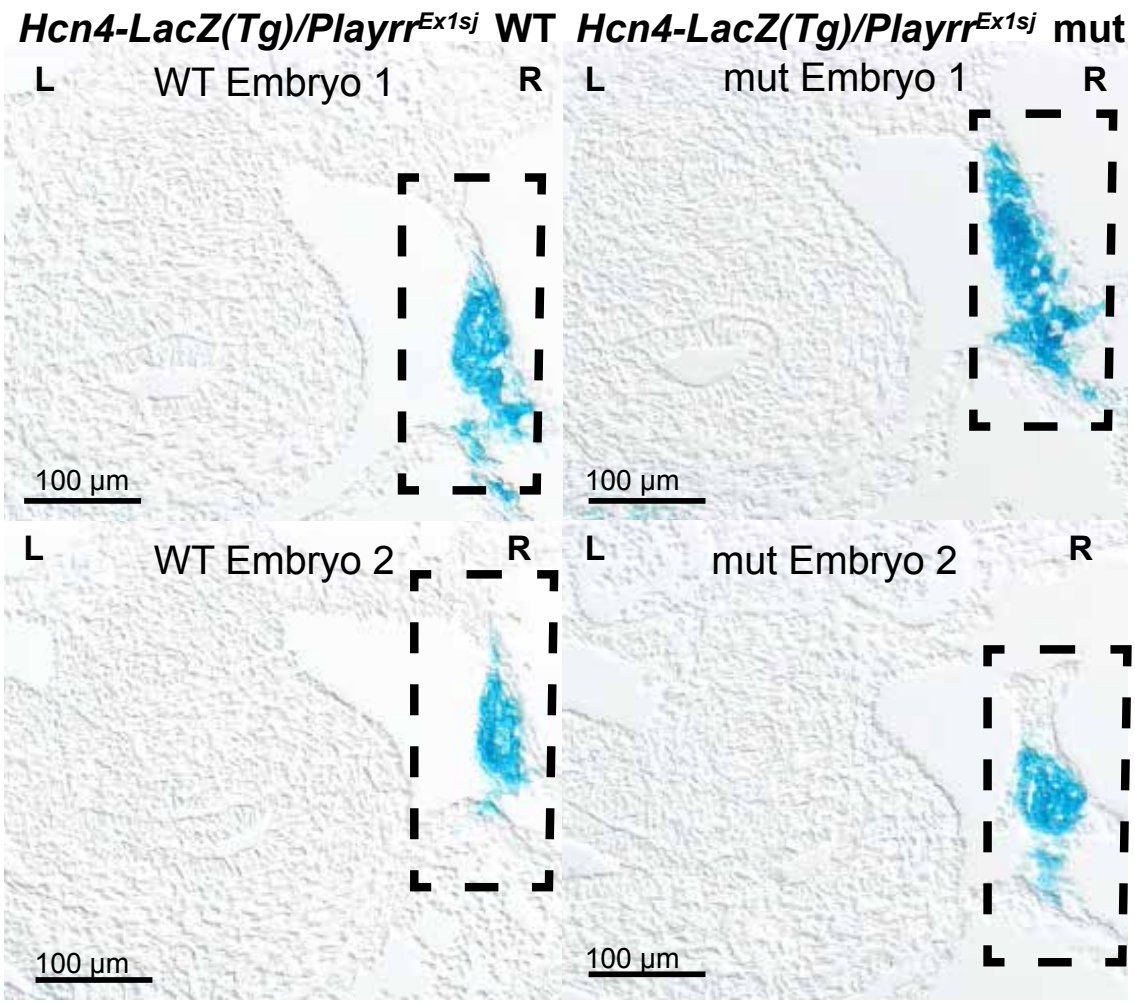


Fig. 3-7: X-gal staining of *Hcn4-LacZ transgenic (Tg)/ Playrr*^{Ex1sj} WT vs *Hcn4-LacZ transgenic (Tg)/ Playrr*^{Ex1sj} mutant E10.5 embryos. Black dashed box highlights R sided SAN cells expressing *Hcn4-LacZ*. No ectopic L SAN was detected based on *Hcn4-LacZ* staining (n=3). Two representative sections from two different embryos of each genotype are shown.

3.5.5 Pitx2 mice are predisposed to pacing induced AF

Several genome wide association studies (GWAS) have reproducibly implicated noncoding genetic variants in the intergenic desert at the human 4q25 locus just upstream of the *Pitx2* gene promoter in significant association with AF (Gudbjartsson, et al., 2007; Sinner et al., 2011, Ellinor et al., 2012; Christopherson et al., 2017). However, the mechanism by which these genetic variants confer AF risk is still unknown. Although *Playrr* is located in the more proximal region of the desert and linearly farther upstream of *Pitx2* and these AF risk loci by approximately 900 kilobases, our previous work investigating chromatin topology at the *Pitx2* locus demonstrates conserved long range looping that brings the *Playrr* and *Pitx2* locus to a mere distance of approximately 20 nm in 3 dimensional space (Welsh et al., 2015). We hypothesized that *Playrr* may provide a missing link for noncoding regulation of *Pitx2* in arrhythmias. In collaboration with the Martin lab (Baylor, TX) we sought to determine if $Playrr^{Ex1sj}$ would similarly phenocopy haploinsufficient *Pitx2* mutant mice and be more susceptible to pacing-induced AF (Wang et al., 2010; Tao et al., 2014). Indeed, we found that compared to WT controls where 0% of mice subjected to programmed stimulation develop AF, approximately 45% of $Playrr^{Ex1sj}$ mutant mice are susceptible to pacing-induced AF (**Fig. 8**).

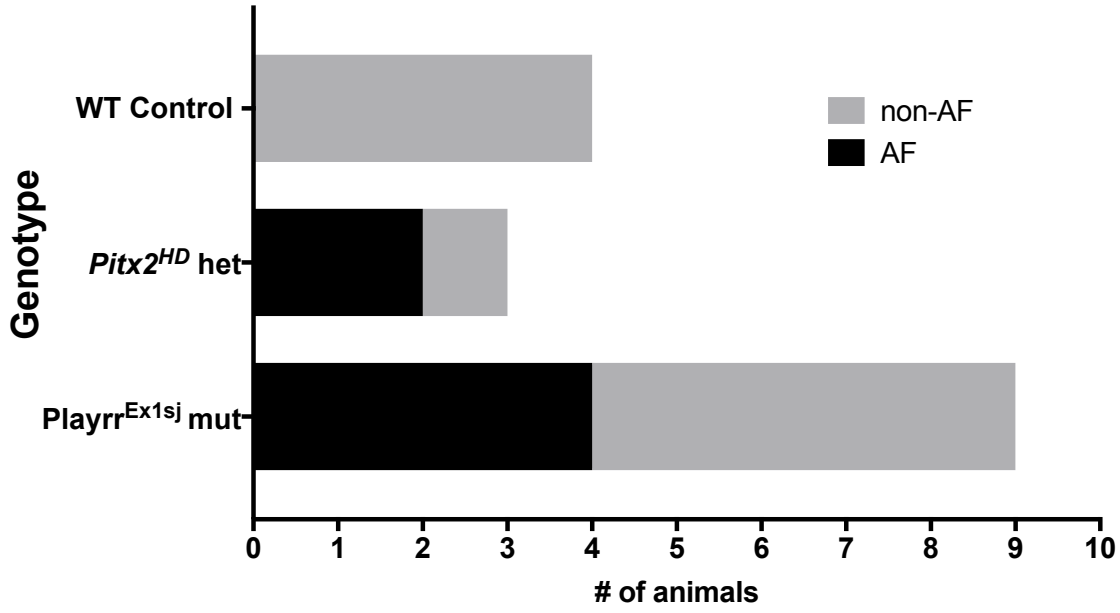


Fig. 3-8: Programmed stimulation and susceptibility to pacing induced AF of *Playrr*^{Ex1sj} mutant mice. *Pitx2*^{HD} het (n=2/5), positive control for pacing induced AF (Wang et al., 2010).

3.6 Discussion

Pitx2 has critical roles in LR cardiac morphogenesis, development of the SAN, and in preventing susceptibility to atrial arrhythmias. The presence of the most significantly associated AF risk variants from multiple GWAS in the noncoding gene desert just upstream of *Pitx2* suggests that noncoding regulatory content provides crucial cis-regulatory inputs at the locus to mediate precise spatiotemporal expression of *Pitx2* so critical in development and disease. Remarkably, here we show that loss of transcription of the lncRNA *Playrr* in mice leads to cardiac arrhythmias mirroring those found in loss-of-function *Pitx2* mutants, including SND and predisposition to pacing-induced AF. What remains to be elucidated are the genomic mechanisms by which the $Playrr^{Ex1sj}$ mutant allele affects *Pitx2* transcription (-in cis and/or -in trans) and/or the dysregulation of other downstream genes that may mediate atrial substrate, trigger, and cardiac rhythm.

3.6.1 Relationship between lncRNA transcription and function

Here we demonstrate specific spatiotemporal expression of *Playrr* in the heart, notably in a small restricted mesenchymal region of sinoatrial patterning in the right pleuropericardial fold mesenchyme at the junction of the right atrium and the right common cardinal vein at E10.5. This is in contrast to *Pitx2,* which in addition to being expressed in the corresponding left SAN domain, demonstrates much broader expression dorsal-ventrally in all left splanchnopleure derived tissues, from the dorsal mesocardium/dorsal mesentery to the left atrium of the heart (**Fig 1B**). Importantly, computational modeling approaches integrating genome wide expression data,

chromatin marks, and gene ontology analyses have been developed to infer biological function from the highly tissue restricted expression of lncRNAs (Li et al., 2015; Perron et al., 2017). Our findings that a) *Playrr* is expressed in the developmental domain of SAN patterning and that b) $Playrr^{Ex1sj}$ mutant mice show evidence of SND is consistent with the idea that the expression pattern of a lncRNA, like that of protein coding genes, can be used to predict its role in biological/physiological processes at the organismal level (Perron et al., 2017).

Of course, ultimately expression does not equate or even necessarily indicate function, and a long standing debate in the lncRNA field is the question of how many of these transcripts represent functional gene regulatory RNAs as opposed to transcriptional by-products resulting from regulatory processes acting on nearby biologically critical gene products. As was utilized here, CRISPR/Cas9 and other genome editing approaches have made it much more feasible to investigate lncRNA function in vivo with reverse genetics approaches. However, as was outlined previously in Chapter 1, separating the role of the RNA transcript from the DNA sequence element as well as interpreting the phenotypic in vivo findings remain challenging (Kopp and Mendell, 2018), and our $Playrr^{Ex1sj}$ allele does not separate the RNA transcript from the act of transcription itself. Future studies will bridge the genomic mechanisms to our discovery of *Playrr* biological function by performing transcriptional profiling and assaying chromatin structure (3C based methods, DNA FISH) within the tissue specific context of the cardiac conduction system.

3.6.2 Stress induced cardiac arrhythmias in *Playrr*^{Ex1sj} mice: a lncRNA model for investigating common genetic pathways and genomic mechanisms underlying atrial substrate and autonomic modulation in SND and AF

Our results from AliveCor ECG screening on awake, restrained mice and subsequent validation with 24 hour telemetry ECG in non-stressed mice under physiological conditions clearly demonstrate that *Playrr*^{Ex1sj} mutants have bradycardia, increased heart rate variability, and increased variation in the duration of beat to beat intervals. Importantly, these signs represent diagnostic criteria for SND, an important risk factor for AF. Furthermore, our observation that more severe bradyarrhythmia occurred both during our AliveCor restrained screening as well as during our pilot restraint stress suggest that *Playrr* mutant mice have SND that is exacerbated by stress. By performing additional telemetry studies using exercise-induced stress and greater sample sizes, we may be able to leverage *Playrr*^{Ex1sj} mice as a novel arrhythmia disease mouse model for investigating the clinically relevant interaction of autonomic modulation and atrial substrate in the development of cardiac arrhythmia.

3.6.3 Dose dependent effects of *Pitx2* on modulating atrial rhythm

We previously showed that *Playrr* transcriptionally inhibits *Pitx2* expression; *Playrr*^{Ex1sj} mutant mice have approximately ~1.5 fold upregulation of *Pitx2* in E14.5 pooled visceral organs (heart, lungs, intestines). Subsequent qRT-PCR analyses assaying for both *Pitx2ab* (symmetric isoforms with unclear roles in morphogenesis) and *Pitx2c* (asymmetric isoform responsible for LR patterning) did not detect differences in pooled heart and lungs samples ($n \geq 4$) or individual heart samples at

E14.5 ($n \geq 5$) (**Appendix A**). In this study, neither our qRT-PCR nor our qualitative ISH analyses were able to detect any differences in *Pitx2* expression in heart or dorsal mesentery/gut primordia at E10.5, when tissue is far less heterogeneous than at E14.5. Indeed, qRT-PCR of individual embryonic heart tissue at E10.5 (n=9) did not detect any differences in *Pitx2ab* or *Pitx2c* isoform expression when the entire heart was assayed (**Fig. 6B)**. This is likely due to lowered specificity and sensitivity (dilution effect) of heterogeneous cell types constituting any body of tissue, especially during embryonic development when gene expression is particularly dynamic across space and time. Although qRT-PCR assays show overall *Playrr* expression is enriched in the heart (**Fig. 1B**), subsequent ISH showed that the critical domain of *Playrr* expression in the SAN is restricted to a highly tissue specific region in the pleuropericardial fold mesenchyme (**Fig. 1C)**. Microdissection and qRT-PCR methods used may not have been sensitive enough to detect differences between WT and mutant within this small, discrete domain. Simultaneously, ISH of *Pitx2* expression in $Playrr^{Ex1sj}$ mutant embryos at E10.5 demonstrated conserved L-R *Pitx2* expression in SAN patterning domains, suggesting that LR asymmetric SAN patterning is conserved in $Playrr^{Ex1sj}$ mutants.

In summary, the studies done here were limited in their ability to detect subtle changes in *Pitx2* or *Hcn4* expression without more information detailing precise domains and developmental dynamics of *Playrr* expression, including during postnatal and adult stages. However, detecting *Playrr* expression by traditional ISH technologies as well as differential methods of probe synthesis have yielded limited results. Future analysis will utilize RNASCOPE advanced ISH technologies to fully

characterize *Playrr* expression and provide much needed resolution to identifying additional *Playrr* spatiotemporal expression domains. These populations and lineages of cells can then be more specifically targeted to achieve the sensitivity required to uncover subtle but critical modulation of *Pitx2* gene dosage leading to susceptibility to cardiac arrhythmias.

3.6.4 *Playrr*^{Ex1sj} mutant mice are predisposed to pacing induced AF

As the most common arrhythmia in the world, AF is a disease of critical importance to study. AF has lifetime risks of 1 in 5 for men and women 55 years of age and older, even in the absence of prior history of congestive heart failure or myocardial infarction, with lifetime risks rising to greater than 1 in 3 (37.8%) for men and women with at least one elevated risk factor (Lloyd-Jones et al., 2004; Staerk et al., 2018). In contrast to rare Mendelian diseases, AF is a case example of a common human disease with complex etiological factors. Because of the ability to leverage large sample sizes, GWAS have proven a powerful tool to demonstrate that AF has a strong genetic component, with the 4q25/*Pitx2* locus region being the most robustly associated in multiple GWAS covering European, African-American, and Japanese ancestry groups (Christophersson, 2017). Yet the causal genetic mechanisms for AF are far from being elucidated.

In collaboration with the group of Dr. James F. Martin we demonstrate here that *Playrr*^{Ex1sj} mutant mice are predisposed to pacing induced AF, comparable to frequency of AF found in positive controls (Pitx2-heterozygous mice, Wang et al., 2010). The work of several key groups have now implicated levels of *Pitx2* dosage in

regulating not only developmental genes required for proper LR development of the SAN (*Shox2, Hcn4, Tbx3)* and atria but also in direct modulation of calcium handling genes (*Atp2a2*, *Casq2*, *Plb*) and novel GWAS AF-associated genes (*Zfhx3*, *Kcnn3*, *Wnt8a*) (Wang et al., 2010; Wang et al., 2014; Tao et al., 2014; Amirabile et al., 2012; Lozano-Velasco et al., 2016). Future studies investigating the status of these critical downstream transcriptional targets of Pitx2 in *Playrr*^{Ex1sj} mutant mice, both during development and in the adult, will lend invaluable insight into the role of this functional lncRNA within gene regulatory networks relevant to cardiac disease.

3.7 Materials and Methods

3.7.1 Animals

Mouse embryos were collected after setting up timed matings with the morning of the plug defined as E0.5 and staged at the time of isolation according to the EMAP eMouse Atlas Project, which extends original Theiler and Kaufman staging criteria (Theiler, 1989; Kaufman, 1992; EMAP eMouse Atlas Project, http://www.emouseatlas.org, 2014). The following mouse strains and alleles were used for analyses: for WT and backcrossing: FVB/NJ (JAX® Laboratories (Stock No: 001800); for *Pitx2* mutants: Pitx2^{hd} on 129S4/SvJaeSor background (Pitx2^{tm1Jfm}; MGI: 2136268); Pitx2^{deltaASE} on C57BL/6J (Pitx2^{tm1Hmd}; MGI: 3767234)); *Playrr*^{Ex1sj} on FVB/NJ background; *Playrr*$^{\Delta 926}$ on FVB/NJ background; *Hcn4-CatCH-IRES-LacZ*. All experiments adhered to guidelines of the Institutional Animal Care and Use

Committee of Cornell University, under the Animal Welfare Assurance on file with the Office Laboratory Animal Welfare.

3.7.2 RNA in situ hybridization

Whole embryos were sliced transversely perpendicular to the gut tube into 250 μm thick embryo slices with a McIlwain tissue chopper (Campden Instruments), fixed in 4% PFA/PBS O/N, dehydrated, and stored in 100% methanol prior to processing. Whole mount ISH was performed on both whole embryos, microdissected embryonic hearts, and 250 μm tissue chopped slices followed standard protocols as previously described (Welsh et al., 2013). Antisense in situ probes were synthesized as previously described (Welsh et al., 2013) corresponding to 3' UTR regions of: *Pitx2-pan* (nucleotides 1150-1801 on NM_011098) and *Playrr* (nucleotides 652-1995 on AK050844).

3.7.3 Quantitative reverse-transcriptase PCR (QRT-PCR)

Embryonic tissue was isolated in cold PBS, and stored in RNAlater until RNA was extracted with a Qiagen RNeasy miniprep kit. 2 mg of RNA was reverse transcribed using the ABI high-capacity cDNA archive kit and diluted to 20ng/μl.

For microdissected tissue (heart, limb buds, gut/DM) from individual embryos at E10.5, TaqMan Cells to Ct Kit was adapted for low input tissues as previously described (Morrison et al., 2015) $Pitx2^{deltaASE}$ and $Pitx2^{hd}$ mutant and heterozygote tissue was used to validate the sensitivity (reduction of Pitx2 expression) and specificity (differential isoform expression) of this technique.

The following TaqMan gene expression assays were used for relative quantification using an ABI7500 realtime PCR system: Actb, Mm00607939_s1; 5730508B09Rik (mC4orf32), Mm02375228_s1; D030025E07Rik (Playrr), Mm03937997_m1; Enpep, Mm00468278_m1; Elovl6, Mm00851223_s1; GAPDH, Mm99999915_g1; Pitx2ab, Mm00660192_g1; Pitx2c, Mm00440826_m1; pan- Pitx2, Mm01316994_m1. Statistical analysis was performed according to MIQE guidelines, using two tailed students' t-tests in JMP® statistical software (SAS).

3.7.4 AliveCor ECG

3-6 month old mice were humanely restrained according to standard IACUC approved procedures and prepped for 30 second to 2 minute non-invasive ECG screening with AliveCor (Kardia). A minimal amount of 70% EtOH was applied to a kimwipe and used to clean and expose the skin of the ventrum of each mouse. Approximately 0.5 mL of ultrasound gel was applied to the thoracic region at the level of the sternum and each mouse to establish contact with the device. AliveCor ECGs were analyzed manually for calculation of HR and R-R intervals.

3.7.5 Telemetry ECG on unstressed mice and mice subjected to space restraint stress

6 month old mice were implanted with the telemetry transmitter TA11PA-C10 (Data Sciences International, P/N: 270-0135-001) following the protocol of the manufacturer. Transmitters were placed subcutaneously and ECG leads were placed in a lead II configuration with subcutaneous fixation of leads within the right and left axial areas. After a postoperative recovery period of ten days, telemetric ECG data

was collected by DSI Telemetric Physiological Monitor System RPC-1 and processed by Dataquest A.R.T 4.31. ECG data was collected on mice for a 24 hour period, including both a 12 hour light; 12 hour dark cycle (24 hours) without stress from handling or anesthesia. For the space restraint stress test, mice were placed in perforated/aerated 50 mL conical tubes 48 hours after the initial 24 hours recording, as previously described (Zimprich et al, 2014) and telemetric ECG data was collected as above for a minimum of 30 seconds and maximum of 5 minutes.

3.7.7 Statistical analysis

24 hour telemetry data and qRT-PCR expression data were displayed and analyzed using Prism 7 for Mac OS X (GraphPad Software). Sample means were compared using unpaired two-tailed t tests. Statistical significance determined using the Holm-Sidak method, with alpha = 0.05, without assuming a consistent SD.

3.7.8 LacZ staining and histological analysis

Whole E10.5 embryos were isolated in ice cold PBS and fixed in 4% PFA for 1 hour at 4 °C. Embryonic tissue was used for genomic DNA preps and genotyping. Whole embryos were washed with LacZ rinse buffer (100 mM sodium phosphate, pH 7.3; 2 mM MgCl2; 0.01% sodium deoxycholate; 0.02% NP-40) 3x30 min at 4 °C, rocking. Embryos were detected for 18 hours in LacZ staining buffer (5 mM potassium ferricyanide, 5 mM potassium ferrocyanide, 1 mg/mL X-gal; in LacZ rinse buffer). LacZ stained embryos were washed in PBS and transferred to 30% sucrose/PBS solution before being cryoembedded in OCT and cryosectioned at 10 um for imaging.

3.8 Acknowledgements

We thank Dr. Frank Lee and Dr. Michael Kotlikoff of CHROMus/Kotlikoff lab for providing *Hcn4-CatCH-IRES-LacZ* transgenic mice; Dr. Marcos Simoes-Costa for guidance and reference for adapting TaqMan Cells-to-Ct kit for qRT-PCR on low input, microdissected organs from single embryos; Rachael Labbitt and Scott Butler for help with telemetry implant surgeries; Becky Dubowitz and Sienna Perry for technical assistance. We are grateful to all members of the Kurpios lab for feedback on the manuscript.

REFERENCES

Ai, D., Liu, W., Ma, L., Dong, F., Lu, M.-F., Wang, D., … Martin, J. F. (2006). Pitx2 regulates cardiac left-right asymmetry by patterning second cardiac lineage-derived myocardium. *Developmental Biology*, *296*(2), 437–449. https://doi.org/10.1016/j.ydbio.2006.06.009

Ammirabile, G., Tessari, A., Pignataro, V., Szumska, D., Sutera Sardo, F., Benes, J., … Campione, M. (2012). Pitx2 confers left morphological, molecular, and functional identity to the sinus venosus myocardium. *Cardiovascular Research*, *93*(2), 291–301. https://doi.org/10.1093/cvr/cvr314

Benjamin, E. J., Chen, P.-S., Bild, D. E., Mascette, A. M., Albert, C. M., Alonso, A., … Wyse, D. G. (2009). Prevention of atrial fibrillation: report from a national heart, lung, and blood institute workshop. *Circulation*, *119*(4), 606–618. https://doi.org/10.1161/CIRCULATIONAHA.108.825380

Bucchi, A., Barbuti, A., DiFrancesco, D., & Baruscotti, M. (2012). Funny Current and Cardiac Rhythm: Insights from HCN Knockout and Transgenic Mouse Models. *Frontiers in Physiology*, *3*, 240. https://doi.org/10.3389/fphys.2012.00240

Cabili, M. N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., & Rinn, J. L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes & Development*, *25*(18), 1915–1927. https://doi.org/10.1101/gad.17446611

Chang, H.-Y., Lin, Y.-J., Lo, L.-W., Chang, S.-L., Hu, Y.-F., Li, C.-H., … Chen, S.-A. (2013). Sinus node dysfunction in atrial fibrillation patients: the evidence of regional atrial substrate remodelling. *EP Europace*, *15*(2), 205–211. https://doi.org/10.1093/europace/eus219

Chinchilla, A., Daimi, H., Lozano-Velasco, E., Dominguez, J. N., Caballero, R., Delpon, E., … Franco, D. (2011). PITX2 Insufficiency Leads to Atrial Electrical and Structural Remodeling Linked to Arrhythmogenesis. *Circulation: Cardiovascular Genetics*, *4*(3), 269–279. https://doi.org/10.1161/CIRCGENETICS.110.958116

Christophersen, I. E., Rienstra, M., Roselli, C., Yin, X., Geelhoed, B., Barnard, J., … Ellinor, P. T. (2017). Large-scale analyses of common and rare variants identify 12 new loci associated with atrial fibrillation. *Nature Genetics*. https://doi.org/10.1038/ng.3843

Ellinor, P. T. P. T., Lunetta, K. L. K. L., Albert, C. M. C. M. C. M., Glazer, N. L., Ritchie, M. D., Smith, A. V, … Kääb, S. (2012). Meta-analysis identifies six new susceptibility loci for atrial fibrillation. *Nature Genetics*. https://doi.org/10.1038/ng.2261.Meta-analysis

John, R. M., & Kumar, S. (2016). Sinus node and atrial arrhythmias. *Circulation*. https://doi.org/10.1161/CIRCULATIONAHA.116.018011

Kirchhof, P., Kahr, P. C., Kaese, S., Piccini, I., Vokshi, I., Scheld, H.-H., … Brown, N. A. (2011). PITX2c Is Expressed in the Adult Left Atrium, and Reducing Pitx2c Expression Promotes Atrial Fibrillation Inducibility and Complex Changes in Gene Expression. *Circulation: Cardiovascular Genetics*, *4*(2), 123–133. https://doi.org/10.1161/CIRCGENETICS.110.958058

Kopp, F., & Mendell, J. T. (2018). Functional Classification and Experimental Dissection of Long Noncoding RNAs. *Cell*. https://doi.org/10.1016/j.cell.2018.01.011

Kozasa, Y., Nakashima, N., Ito, M., Ishikawa, T., Kimoto, H., Ushijima, K., … Takano, M. (2018). HCN4 pacemaker channels attenuate the parasympathetic response and stabilize the spontaneous firing of the sinoatrial node. *The Journal of Physiology*, *596*(5), 809–825. https://doi.org/10.1113/JP275303

Lappalainen, T. (2015). Functional genomics bridges the gap between quantitative genetics and molecular biology. *Genome Research*, *25*(10), 1427–1431. https://doi.org/10.1101/gr.190983.115

Liang, X., Evans, S. M., & Sun, Y. (2017). Development of the cardiac pacemaker. *Cellular and Molecular Life Sciences : CMLS*, *74*(7), 1247–1259. https://doi.org/10.1007/s00018-016-2400-1

Lloyd-Jones, D. M., Wang, T. J., Leip, E. P., Larson, M. G., Levy, D., Vasan, R. S., … Benjamin, E. J. (2004). Lifetime risk for development of atrial fibrillation: The framingham heart study. *Circulation*. https://doi.org/10.1161/01.CIR.0000140263.20897.42

Logan, M., Pagán-Westphal, S. M., Smith, D. M., Paganessi, L., & Tabin, C. J. (1998). The transcription factor pitx2 mediates situs-specific morphogenesis in response to left-right asymmetric signals. *Cell*, *94*(3), 307–317. https://doi.org/10.1016/S0092-8674(00)81474-9

Mommersteeg, M. T. M., Hoogaars, W. M. H., Prall, O. W. J., de Gier-de Vries, C., Wiese, C., Clout, D. E. W., … Christoffels, V. M. (2007). Molecular Pathway for the Localized Formation of the Sinoatrial Node. *Circulation Research*, *100*(3), 354–362. https://doi.org/10.1161/01.RES.0000258019.74591.b3

Pérez-Hernández, M., Matamoros, M., Barana, A., Amorós, I., Gómez, R., Núñez, M., … Caballero, R. (2016). Pitx2c increases in atrial myocytes from chronic atrial fibrillation patients enhancing I_{Ks} and decreasing $I_{Ca,L}$. *Cardiovascular Research*, *109*(3), 431–441. https://doi.org/10.1093/cvr/cvv280

Rev, O. (2015). Development and Function of the Cardiac Conduction System in Health and Disease David. *Journal of Cardiovascular Development and Disease*, *14*(11), 871–882. https://doi.org/10.1111/obr.12065.Variation

Roberts-Thomson, K. C., Sanders, P., & Kalman, J. M. (2007). Sinus Node Disease: An Idiopathic Right Atrial Myopathy. *Trends in Cardiovascular Medicine*, *17*(6), 211–214. https://doi.org/10.1016/J.TCM.2007.06.002

Ryan, A. K., Blumberg, B., Rodriguez-Esteban, C., Yonei-Tamura, S., Tamura, K., Tsukui, T., … Belmonte, J. C. I. (1998). Pitx2 determines left-right asymmetry of internal organs in vertebrates. *Nature*, *394*(6693), 545–551. https://doi.org/10.1038/29004

Soares, R. J., Maglieri, G., Gutschner, T., Diederichs, S., Lund, A. H., Nielsen, B. S., & Holmstrøm, K. (2017). Evaluation of fluorescence in situ hybridization techniques to study long non-coding RNA expression in cultured cells. *Nucleic Acids Research*, *46*(1), e4. https://doi.org/10.1093/nar/gkx946

Sonawane, A. R., Platig, J., Fagny, M., Chen, C.-Y., Paulson, J. N., Lopes-Ramos, C. M., … Kuijjer, M. L. (2017). Understanding Tissue-Specific Gene Regulation. *Cell Reports*, *21*(4), 1077–1088. https://doi.org/10.1016/j.celrep.2017.10.001

Staerk, L., Wang, B., Preis, S. R., Larson, M. G., Lubitz, S. A., Ellinor, P. T., … Trinquart, L. (2018). Lifetime risk of atrial fibrillation according to optimal, borderline, or elevated levels of risk factors: cohort study based on longitudinal data from the Framingham Heart Study. *BMJ (Clinical Research Ed.)*, *361*, k1453. https://doi.org/10.1136/BMJ.K1453

Tao, Y., Zhang, M., Li, L., Bai, Y., Zhou, Y., Moon, A. M., … Martin, J. F. (2014). Pitx2, an atrial fibrillation predisposition gene, directly regulates ion transport and intercalated disc genes. *Circulation. Cardiovascular Genetics*, *7*(1), 23–32. https://doi.org/10.1161/CIRCGENETICS.113.000259

Wang, J., Klysik, E., Sood, S., Johnson, R. L., Wehrens, X. H. T., & Martin, J. F. (2010). Pitx2 prevents susceptibility to atrial arrhythmias by inhibiting left-sided pacemaker specification. *Proceedings of the National Academy of Sciences*, *107*(21), 9753–9758. https://doi.org/10.1073/pnas.0912585107

Zimprich, A., Garrett, L., Deussing, J. M., Wotjak, C. T., Fuchs, H., Gailus-Durner, V. Halter, S. M. (2014). A robust and reliable non-invasive test for stress responsivity in mice. Frontiers in Behavioral Neuroscience, 8, 125. https://doi.org/10.3389/fnbeh.2014.00125

CHAPTER 4- SUMMARY AND FUTURE DIRECTIONS

4.1 Introduction

The noncoding genome has been shown to play essential roles in transcriptional regulation and chromatin organization during development as well as in many physiological and disease states. Transcriptome studies in the past decade have also demonstrated that this noncoding sequence is 'pervasively' transcribed into thousands of noncoding RNA products, offering a promising new avenue for investigating another level of genomic regulation of gene expression and discovery of molecular markers and targets for disease treatment. However, despite extensive GWAS implicating that the vast majority of genetic disease associations occurs within non-coding genome regions, the functional roles of these sequence elements and their transcribed products remains to be elucidated. In Chapter 2, I have highlighted our discovery of the lncRNA *Playrr* as a promising candidate cis-regulatory element in modulating Pitx2 in the context of LR asymmetric organogenesis, and in Chapter 3, I have investigated a putative role for *Playrr* in regulation of *Pitx2* LR asymmetric expression in the developing heart and discovered a biological role for *Playrr* in susceptibility to cardiac arrhythmias. This final chapter will provide a summary of the key findings of preceding chapters, discussion of additional follow up work, and highlight important directions for future studies.

4.2 A putative functional role for *Playrr* in intestinal morphogenesis

Our work, detailed in Chapter 2, has shown that the LR asymmetric gut dorsal mesentery (DM) at the *Pitx2* locus is characterized by LR compartment-specific asymmetric chromatin architecture that correlates with conserved left specific transcription of *Pitx2* essential for proper gut laterality (Welsh et al, 2015). Additionally, we discovered the lncRNA *Playrr* conserved between mouse and chicken arising from the DNA enhancer element e926 from within the intergenic desert of the *Pitx2* locus in the right DM. By generating multiple CRISPR/Cas9 genome edited alleles at the *Pitx2* locus, we demonstrated a role for *e926/Playrr* in repression of *Pitx2* transcription (Welsh et al., 2015).

In order to follow up on this published work and to investigate the function of *Playrr* in regulating *Pitx2* expression in the context of intestinal morphogenesis, I subsequently performed organ specific gene expression analysis using qRT-PCR in $Playrr^{Ex1sj}$ mutant mice (**Appendix A**). I first recapitulated our original finding (Welsh et al., 2015) of increased *Pitx2* expression in E14.5 pooled visceral organs (heart, lungs, intestines (esophagus to colon)) of both $Playrr^{Ex1sj}$ and $Playrr^{\Delta 926}$ mutants compared to WT (n=5). These experiments also demonstrated an approximately 1.5 to 2 fold upregulation of both *Pitx2ab* and *Pitx2c* isoforms in pooled visceral organs **(Appendix A1a-A1b)**. I then focused my analyses on $Playrr^{Ex1sj}$ to investigate the effect of loss of *Playrr* transcript/transcription as opposed to a deletion of its genomic locus, which would not be able to separate out the functions of the DNA cis regulatory element e926 and *Playrr*. Subsequent qRT-PCR in E14.5 $Playrr^{Ex1sj}$ pooled heart and

lungs did not detect any significant differences in *Pitx2ab* or *Pitx2c* isoform expression between mutant and WT controls (n=4) **(Appendix A1c)**. *qRT*-PCR experiments on microdissected E14.5 small intestinal samples (duodenum, jejunum, and ileum; n=4) demonstrated a similar level (1.5 fold) of upregulation, specifically of *Pitx2ab* but not *Pitx2c* (**Appendix A1d)**. However, a replication of this study in *Playrr*^{Ex1sj} E14.5 small intestinal samples with a greater number of biological replicates (n=7) failed to recapitulate this upregulation (data not shown). Additionally, qRT-PCR experiments at E10.5 and E13.5 also failed to detect any significant differences in *Pitx2* expression. These inconsistencies most likely reflect the technical limitations of qRT-PCR and the inherent challenges of detecting subtle effects of lncRNAs *in vivo* in developing tissues, including the heterogeneity of cell types assayed and the dynamic nature of *Pitx2* expression levels during these developmental time points assayed. The initial upregulation detected in the pooled visceral organ analysis could be due to small, discrete population of specific cell types within each organ and therefore assaying individual organs may have led to a relative increased dilution of affected cells beyond or just within the range of the sensitivity of our qRT-PCR methods, leading to inconsistent expression results.

We decideded to address if loss of *Playrr* and a potential upregulation of *Pitx2* expression in these mice would functionally perturb intestinal morphogenesis. Using intestinal length and looping morphometric analyses in *Playrr*^{Ex1sj} mutant intestines and WT and Pitx2 loss-of-function mutant controls, I demonstrated that these parameters describing overall gut tilting and looping morphogenesis are not affected in *Playrr*^{Ex1sj} mutant mice **(Appendix A2a-d)**. These findings suggest that *Playrr* is not

necessary for regulating the onset and maintenance of LR asymmetric Pitx2c expression and intestinal looping morphogenesis.

One plausible hypothesis following from this work is that *Playrr* instead provides fine-tuning of Pitx2 isoform-specific expression and function of specific and small populations of intestinal cell types. Many previous attempts to identify biological functions of lncRNAs have proven difficult due to technical challenges of detecting lncRNA expression, which are expressed at much lower levels and in more highly tissue specific/cell type domains when compared to protein coding mRNAs (Yan et al., 2012; Soares et al., 2017). Additionally, attempts to uncover phenotypes under physiological conditions often fail, most likely due to redundancy and/or functional compensation at multiple biological levels, from cis-regulatory sequence in the genome to molecular pathways governing highly conserved developmental processes (Li et al., 2014). However, future experiments using advanced RNA ISH techniques, such as RNASCOPE, will help to localize the cell types in the intestines expressing *Playrr*. Finally, current work in the lab characterizing the physiological roles of *Pitx2* in lymphatic development and lipid transport will likely prove useful as technically validated assays in which to stress and assay gut functionality in $Playrr^{Ex1sj}$ mutant mice, thereby uncovering cell-type specific physiological roles for *Playrr* in the digestive tract.

4.3 The *Playrr*^{Ex1sj} mutant mouse model: *Pitx2* independent and dependent mechanisms underlying susceptibility to atrial arrhythmias

My data demonstrate that *Playrr*^{Ex1sj} have SND and predisposition to AF, yet the LR asymmetric Pitx2 expression and patterning of SAN during development are unchanged in the absence of *Playrr*. Studies in various conditional loss-of-function *Pitx2* mouse models have demonstrated distinct roles for *Pitx2* during development, to delimit SAN development on the left side as well as postnatally, via maintenance of expression in adult left atrium and direct regulation of critical ion channel and intercalated disc genes (Wang et al., 2010; Ammirabile et al., 2011; Tao et al., 2014; Wang et al., 2014). Thus, the critical question that remains is whether *Playrr* is acting independently or more likely, through modulation and/or maintenance of *Pitx2* dosage to provide protective effects against atrial arrhythmia during adulthood. Additionally, it remains to be seen whether *Playrr*, like *Pitx2*, has distinct roles during development as well as postnatally (Tao et al., 2014).

In order to validate the *Playrr*^{Ex1sj} mutant as a complementary and useful model for understanding mechanisms of *Pitx2* mediated arrhythmias and to address the pathological mechanisms induced by loss of *Playrr*, I suggest the following experiments to be pursued. Our current data indicate that *Playrr* is expressed in SAN precursor and progenitor cells at E8.25 and E10.5. Similar to expression analyses to be performed in the intestines, RNASCOPE can be utilized to narrow down the spatiotemporal domains of *Playrr* expression to more effectively target specific

populations of cells for changes in isoform specific *Pitx2* expression and increase the sensitivity from assaying less heterogeneous populations of cells. In particular, it will be important to assay for *Playrr* expression in the adult sinoatrial and surrounding atrial and pulmonary myocardial domains to determine separate developmental vs adult physiological roles.

Once the critical spatiotemporal domain of *Playrr* expression in the heart and SAN has been identified, I propose to pursue a single cell transcriptomics analysis via DROP-Seq (in collaboration with Charles Danko) to not only profile the heterogeneous genetic populations of the SAN but also to reveal downstream targets and upstream genetic pathways perturbed as a result of loss of *Playrr*. Leveraging this powerful single cell approach will allow us to both amplify and prioritize testable hypotheses that will direct future experimental work and generate an enhanced mechanistic understanding bridging lncRNA roles in the genome with pathophysiological relevance.

Finally, additional functional assays can be pursued to further characterize the role of *Playrr* in SND and AF via optical mapping and through optogenetic approaches. This would involve crossing *Playrr*^{Ex1sj} mutant mice with *Hcn4-CatCH-IRES-LacZ* mice (Kotlikoff lab/CHROMus). Optical mapping has been previously used to demonstrate altered electrical substrate in the heart at specific developmental stages (Ammirabile et al., 2012; Franco et al., 2017). Optogenetic tools are an exciting emergent technology that enables optical modulation of cardiac electrical activity (Boyle et al., 2018). The *Hcn4-CatCH-IRES-LacZ* transgene expresses the rhodopsin

effector (*CatCH)* under the *Hcn4* (SAN marker) promoter. In collaboration with Nozomi Nishimura, we intend to leverage these genetic tools and collaborative expertise to optogenetically stimulate and perturb the SAN in *Playrr*^{Ex1sj} mice, providing additional insight into if, when, and how SAN function during development and postnatally is affected by the loss of *Playrr*.

4.4 Redundancy and combinatorial action of cis-regulatory elements: using CRISPR/Cas9 for deletion of the entire noncoding gene desert and a potential novel ARS disease model (APPENDIX B)

Ultimately it is to be expected that perturbing individual cis-regulatory elements is unlikely to have dramatic effects on the highly stereotypical process of morphogenesis given the evidence pointing to redundant, additive, and combinatorial action of cis-regulation of transcription to ensure robust phenotypes (Frankel et al., 2010; Spitz et al., 2012; Cannavo et al., 2016). However, our results together indicate that perturbing a single noncoding regulatory element, in this case the transcription of the lncRNA, *Playrr*, can cause a subtle but crucial functional defect in discrete localized cellular domains that have important pathophysiological relevance for human disease. Indeed, my results finding no differences in intestinal looping morphogenesis nor in gross heart structure and LR patterning of *Playrr*^{Ex1sj} mutant mice suggest that a single cis-regulatory perturbation is unlikely to make gross disruptions to tightly controlled gene regulatory networks underlying highly conserved, stereotypical processes of morphogenesis, but instead demonstrate that subtle modulatory perturbations in certain key functional cell types either during development and/or during postnatal life

are sufficient to cause physiological defects at the level of organ function. What remains to be determined is the exact temporal domains in which *Playrr* and perturbation of *Playrr* affects cardiac and/or SAN function, specifically whether it mirrors *Pitx2* in having both developmental and postnatal functions.

Nonetheless, to investigate the potential of functional redundancy or combinatorial action of noncoding regulatory elements at the *Pitx2* locus playing crucial biological roles in *Pitx2* related development and disease, I have utilized CRISPR/Cas9 genome editing to create additional functional perturbations and mutant mouse alleles (**Appendix B-1a).** Complementary to our work using a candidate approach to evaluate the function of the single cis-regulatory element *Playrr*/e926, I also successfully designed a CRISPR/Cas9 strategy utilizing two guide RNAs and a bridging oligonucleotide as previously described (Boroviak et al., 2016) to generate a novel mutant allele, $Pitx2^{\Delta desert}$, which removes essentially the entirety of the almost over 1 megabase of the noncoding gene desert and all its putative regulatory elements in between the 5' *c4orf32* and 3' *Pitx2* genes in mice (**Appendix B-1a**). Preliminary analysis of Mendelian ratios in liveborn and embryonic litters in these mice suggests embryonic lethality of $Pitx2^{\Delta desert}$ homozygotes, most likely due to perturbations of heart morphogenesis (**Appendix B-1b)**. Future studies will investigate whether the loss of both copies of the *Pitx2* locus gene desert in $Pitx2^{\Delta desert}$ homozygotes is due to *Pitx2* related insufficiency and/or pathology. Importantly, this work may help to directly validate the biological role of the noncoding domain at the *Pitx2* locus, provide a genomic model for assaying changes to chromatin structure and transcription in the absence of regulatory information, and/or provide a novel disease

model for *Pitx2*-related disease by recapitulating ARS patients with upstream desert deletions but intact *Pitx2* coding sequence (Volkmann et al, 2011; Reis et al., 2012).

REFERENCES

Ammirabile, G., Tessari, A., Pignataro, V., Szumska, D., Sutera Sardo, F., Benes, J., … Campione, M. (2012). Pitx2 confers left morphological, molecular, and functional identity to the sinus venosus myocardium. *Cardiovascular Research*, *93*(2), 291–301. https://doi.org/10.1093/cvr/cvr314

Boroviak, K., Doe, B., Banerjee, R., Yang, F., & Bradley, A. (2016). Chromosome engineering in zygotes with CRISPR/Cas9. *Genesis*, *54*(2), 78–85. https://doi.org/10.1002/dvg.22915

Cannavò, E., Khoueiry, P., Garfield, D. A., Geeleher, P., Zichner, T., Gustafson, E. H., … Furlong, E. E. M. (2016). Shadow Enhancers Are Pervasive Features of Developmental Regulatory Networks. *Current Biology*, *26*(1), 38–51. https://doi.org/10.1016/j.cub.2015.11.034

Frankel, N., Davis, G. K., Vargas, D., Wang, S., Payre, F., & Stern, D. L. (2010). Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature*, *466*(7305), 490–493. https://doi.org/10.1038/nature09158

Reis, L. M., Tyler, R. C., Volkmann Kloss, B. A., Schilter, K. F., Levin, A. V, Lowry, R. B., … Semina, E. V. (2012, December 9). PITX2 and FOXC1 spectrum of mutations in ocular syndromes. *European Journal of Human Genetics*. https://doi.org/10.1038/ejhg.2012.80

Soares, R. J., Maglieri, G., Gutschner, T., Diederichs, S., Lund, A. H., Nielsen, B. S., & Holmstrøm, K. (2017). Evaluation of fluorescence in situ hybridization techniques to study long non-coding RNA expression in cultured cells. *Nucleic Acids Research*, *46*(1), e4. https://doi.org/10.1093/nar/gkx946

Spitz, F., & Furlong, E. E. M. (2012). Transcription factors: from enhancer binding to developmental control. *Nature Reviews Genetics*, *13*(9), 613–626. https://doi.org/10.1038/nrg3207

Tao, Y., Zhang, M., Li, L., Bai, Y., Zhou, Y., Moon, A. M., … Martin, J. F. (2014). Pitx2, an atrial fibrillation predisposition gene, directly regulates ion transport and intercalated disc genes. *Circulation. Cardiovascular Genetics*, *7*(1), 23–32. https://doi.org/10.1161/CIRCGENETICS.113.000259

Wang, J., Klysik, E., Sood, S., Johnson, R. L., Wehrens, X. H. T., & Martin, J. F. (2010). Pitx2 prevents susceptibility to atrial arrhythmias by inhibiting left-sided pacemaker specification. *Proceedings of the National Academy of Sciences*, *107*(21), 9753–9758. https://doi.org/10.1073/pnas.0912585107

Welsh, I. C., Kwak, H., Chen, F. L., Werner, M., Shopland, L. S., Danko, C. G., … Kurpios, N. A. (2015). Chromatin Architecture of the Pitx2 Locus Requires CTCF- and Pitx2-Dependent Asymmetry that Mirrors Embryonic Gut Laterality. *Cell Reports*, *13*(2), 337–349. https://doi.org/10.1016/j.celrep.2015.08.075

Volkmann, B. A., Zinkevich, N. S., Mustonen, A., Schilter, K. F., Bosenko, D. V, Reis, L. M., … Semina, E. V. (2011). Potential novel mechanism for Axenfeld-Rieger syndrome: Deletion of a distant region containing regulatory elements of PITX2. *Investigative Ophthalmology and Visual Science*, *52*(3), 1450–1459. https://doi.org/10.1167/iovs.10-6060

Yan, B., Wang, Z. H., & Guo, J. T. (2012, February 1). The research strategies for probing the function of long noncoding RNAs. *Genomics*. Academic Press. https://doi.org/10.1016/j.ygeno.2011.12.002

APPENDIX

APPENDIX A

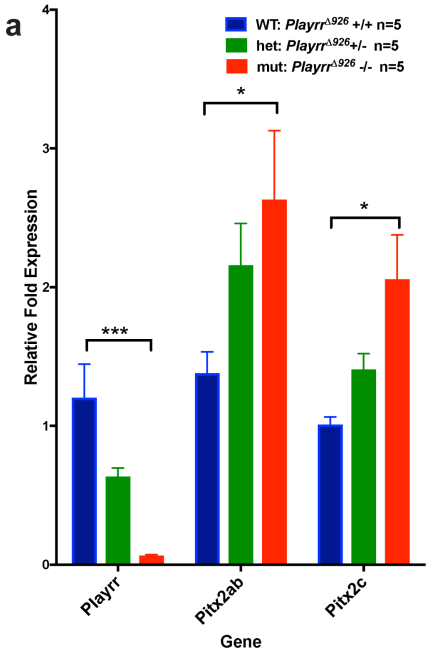
THE L-R ASYMMETRIC CIS-ELEMENT e926/PLAYRR AT THE PITX2 LOCUS DIFFERENTIALLY REGULATES PITX2 ISOFORM EXPRESSION IN THE DEVELOPING GUT

This appendix represents my work in follow up to our publication in *Cell Reports* (Welsh et al., 2015) detailed in Chapter 2. This body of work used gene expression analyses and intestinal morphometric analyses in *Playrr*^{Ex1sj} and *Playrr*$^{\Delta 926}$ mutant mouse embryos to investigate the transcriptional and biological roles of the cis-regulatory element/long noncoding RNA module, e926/*Playrr*.

Tissue specific upregulation of Pitx2 isoform expression in e926 and Playrr-Ex1sj mutant mice **(Fig. A-1)**

Previously we showed using qRT-PCR that *Pitx2* expression was found to be significantly upregulated in visceral organ primordia (heart, lungs, & gastrointestinal tract) of E14.5 in both *Playrr*$^{\Delta 926}$ and *Playrr*^{Ex1sj} compared to WT littermates. Using qRT-PCR probes specific for symmetric isoforms *Pitx2ab* and the asymmetric isoform *Pitx2c*, I demonstrate here that loss of *Playrr* transcription leads to an approximately ~1.5-2 fold upregulation of overall dosage of *Pitx2* isoforms *Pitx2ab* and *Pitx2c* in both *Playrr*$^{\Delta 926}$ and *Playrr*^{Ex1sj} mice (**Fig. A-1a; A1-b)**. Upon further investigation, qRT-PCR of individual E14.5 organs (intestines, heart, and lungs) recapitulated similar levels of upregulation for *Pitx2ab* but not *Pitx2c* in E14.5 intestines/gut while no change was found in pooled heart and lungs (**Fig. A-1c)**. Further microdissection

localized modest (approximately 1.5 fold) *Pitx2ab* upregulation in E14.5 small intestines/midgut samples (dudoenum, jejunum, and ileum) (**Fig. A-1d)**.



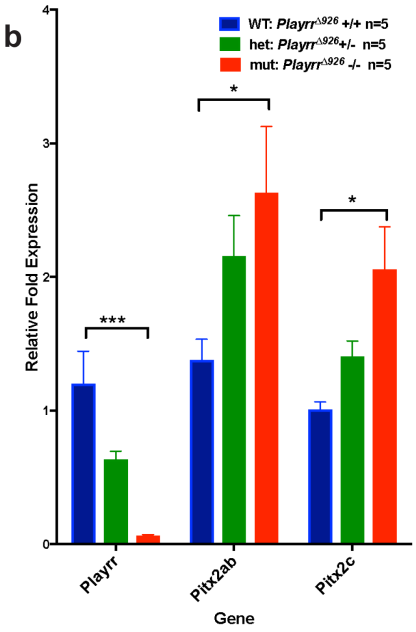
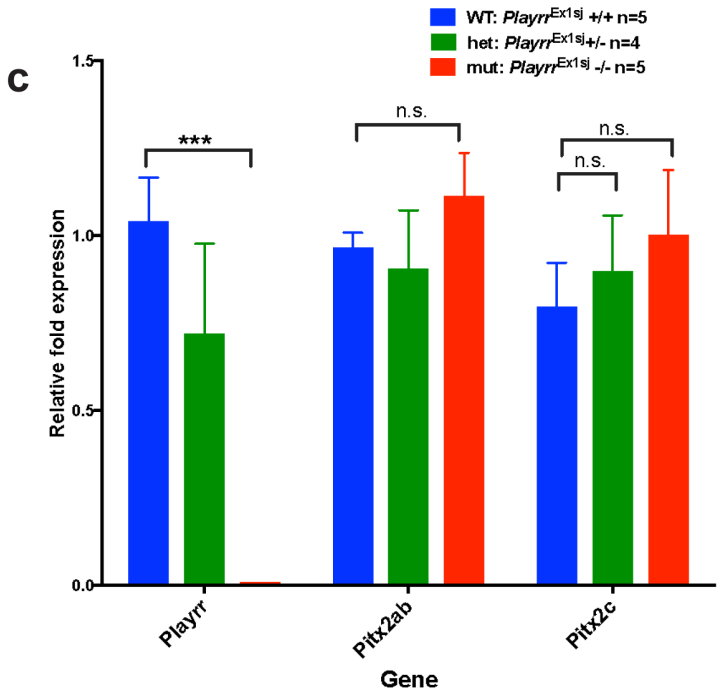


Fig A1: Upregulation of *Pitx2* isoforms *Pitx2ab* and *Pitx2c* in pooled E14.5 visceral organs (heart, lungs, intestines)

(1a) qRT-PCR graphs for *Playrr*, *Pitx2ab,* and *Pitx2c* expression in E14.5 pooled visceral organs of *Playrr*$^{\Delta 926}$ wild type (WT), heterozygote (het), and mutant (mut) embryos. *p<0.05; ***p<0.001

(1b) qRT-PCR graphs for *Playrr*, *Pitx2ab,* and *Pitx2c* expression in E14.5 pooled visceral organs of *Playrr* ^{Ex1sj} wild type WT, het, and mut embryos. *p<0.05; ***p<0.001



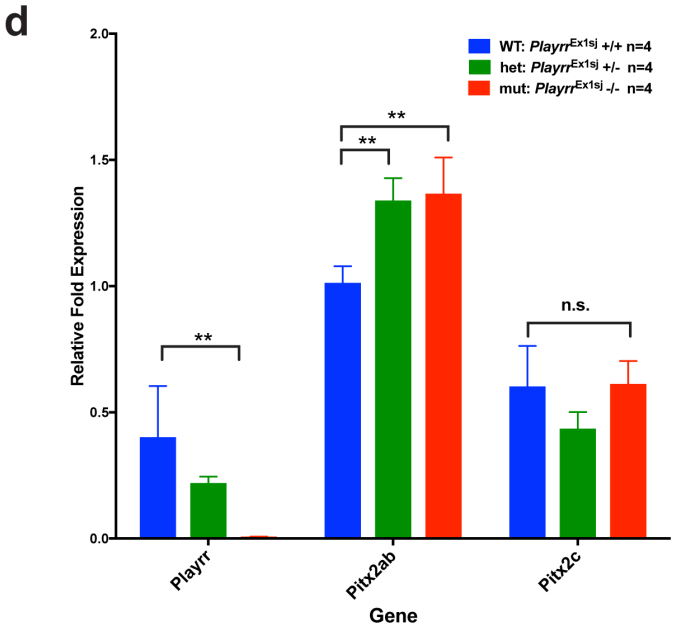


Fig 1 cont'd: Tissue and isoform specific upregulation of *Pitx2* isoforms *Pitx2ab* and *Pitx2c* in pooled E14.5 visceral organs (heart, lungs, intestines)

(1c) qRT-PCR graphs for *Playrr*, *Pitx2ab,* and *Pitx2c* expression in E14.5 heart and lungs of *Playrr*^{Ex1sj} wild type (WT), heterozygote (het), and mutant (mut) embryos. *p<0.05; ***p<0.001

(1d) qRT-PCR graphs for *Playrr*, *Pitx2ab,* and *Pitx2c* expression in E14.5 small intestinal (duodenum, jejunum, ileum) of *Playrr*^{Ex1sj} wild type WT, het, and mut embryos. *p<0.05; **p<0.005; ***p<0.001

Stereotypical looping morphogenesis is conserved in e926/Playrr-Ex1sj mutant mice (**Fig. A-2)**

Pitx2 dosage has tissue specific effects (proliferation, cell fate specification, and differentiation) on organ morphogenesis and different organs have different sensitivities to Pitx2 gene dosage (Gage et al 1999; Liu et al 2001; Shiratori et al 2006). Decreasing levels of Pitx2 can partially or fully alter asymmetrical morphogenesis of the heart and lungs and alter the directionality of duodenal rotation (Gage et al 1999; Liu et al 2001). Recognizing that dosage changes of *Pitx2* can lead consequences on morphogenesis, we asked if a modest ~1.5 fold upregulation of *Pitx2ab* in the small intestine would alter intestinal lengths or the directionality of intestinal looping as has been previously reported for *Pitx2* loss of function muants (Liu et al., 2001). Using the distances between the pyloric-duodenal junction (PDJ), duodenal-jejunal junction (DJJ), and the ileocecal junction (ICJ) as landmarks as previously described (Sivakumar et al, manuscript accepted *2018*), we used these morphometric parameters to characterize the geometry of stereotypical intestinal coiling **(Fig. A-2a)**. We validated this method showing significant differences between PDJ-ICJ distance in Pitx2^{deltaASE} mice correlating with differences in *Pitx2* dosage (**Fig. A-2b)** but were unable to detect statistically significant differences using this method between *Playrr*$^{\Delta 926}$ mutant mice and WT littermates (**Fig. A-2c)** as well as between *Playrr*^{Ex1sj} mutant mice and WT littermates (**Fig. A-2d).** Additionally, no significant differences in small intestinal length (duodenum to ileum) at stages E13.5, E14.5, and E15.5 were observed between *Playrr*^{Ex1sj} mutant mice and WT littermates

(**Fig. A-2e).** These results suggest that *Playrr* transcription may not be required for the overall growth and conserved chirality of the small intestine during development.

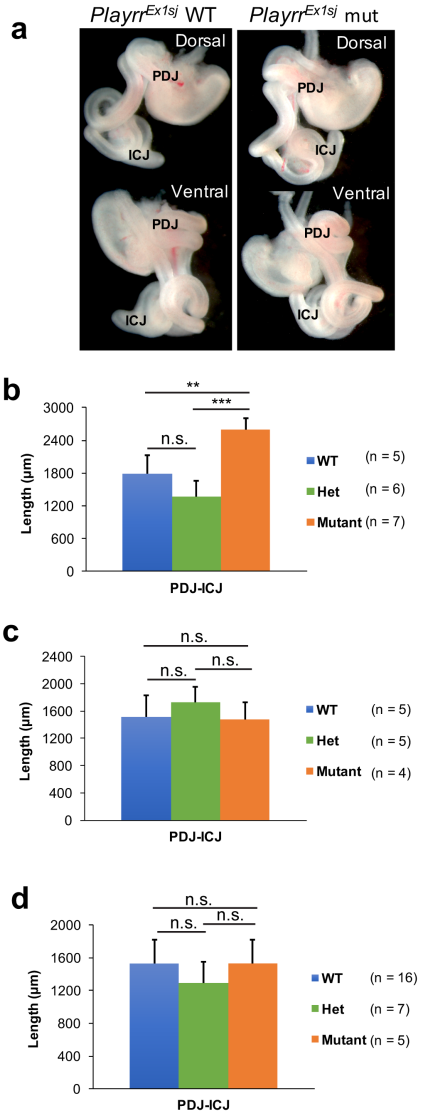


Fig A2: Intestinal looping morphometric analyses in *Playrr$^{\Delta 926}$*, *Playrr^{Ex1sj}*, and *Pitx2^{deltaASE}* embryos.

(1a) representative two-view images of dorsal (top) and ventral (bottom) aspect of E13.5 small intestinal tract in *Playrr^{Ex1sj}* WT and mut mice. PDJ= pyloric duodenal junction; ICJ = ileocecal junction

(1b) Avg length of measured distance between PDJ and ICJ from two views of looped intestinal tract in *Pitx2^{deltaASE}* WT, het, and mut embryos. **p<0.005; ***p<0.0005; n.s. = not significant

(1c) Avg length of measured distance between PDJ and ICJ from two views of looped intestinal tract in *Playrr$^{\Delta 926}$* WT, het, and mut embryos. n.s. = not significant

(1d) Avg length of measured distance between PDJ and ICJ from two views of looped intestinal tract in *Playrr^{Ex1sj}* WT, het, and mut embryos. n.s. = not significant

e

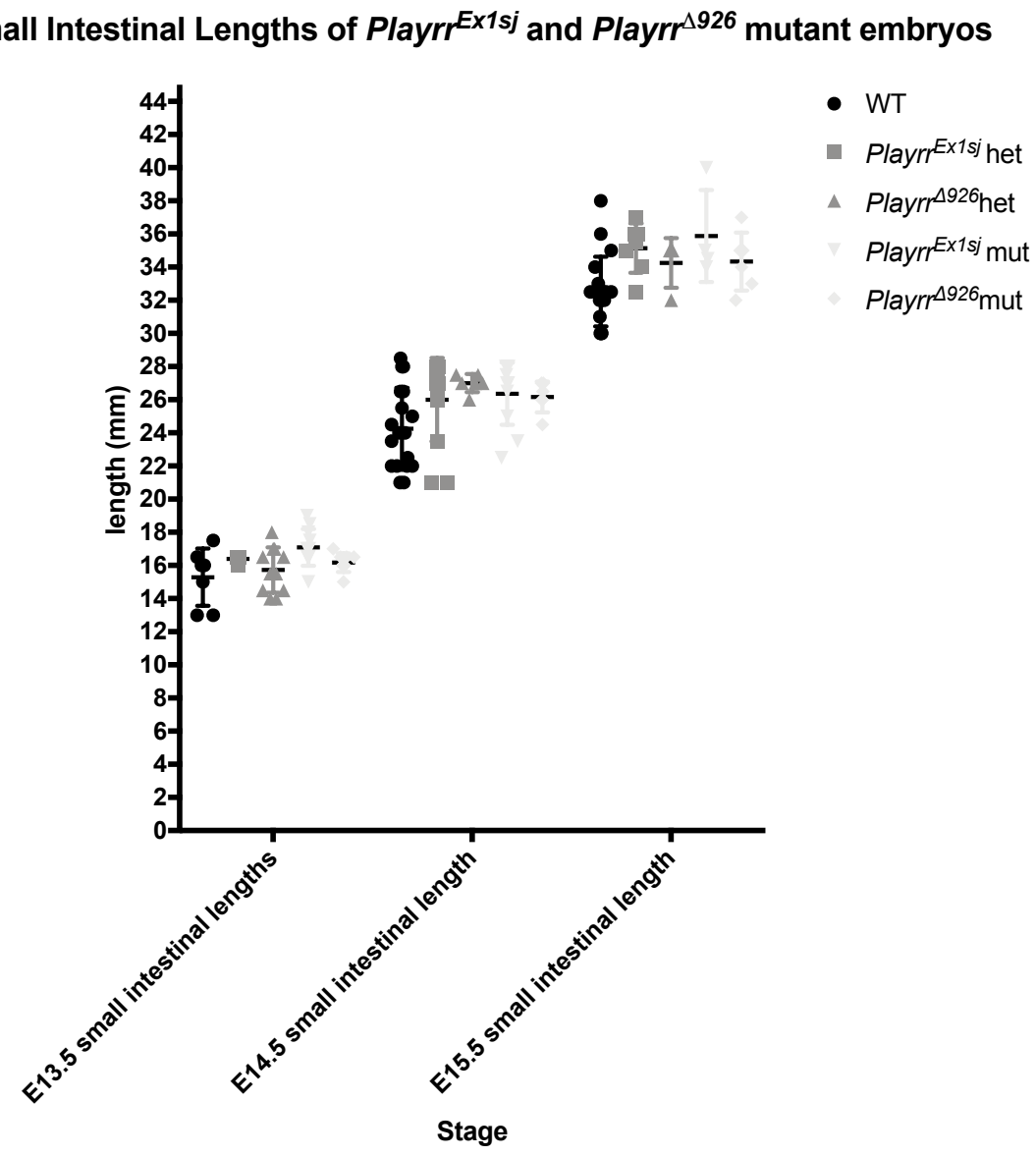


Fig A2: Intestinal looping morphometric analyses in *Playrr*$^{\Delta 926}$, *Playrr*^{Ex1sj}, and *Pitx2*^{deltaASE} embryos.

(2e) Small intestinal lengths measured from pyloric duodenal junction to ileocecal junction on linearized intestinal samples at E13.5, E14.5 and E15.5. No significant differences were observed between any genotypes. Biological replicates numbered at least 5 per biogroup/genotype.

REFERENCES

Gage, P. J., Suh, H., & Camper, S. A. (1999). Dosage requirement of Pitx2 for development of multiple organs. *Development (Cambridge, England)*, *126*(20), 4643–4651. Retrieved from http://dev.biologists.org/content/develop/126/20/4643.full.pdf

Liu, C., Liu, W., Palie, J., Lu, M. F., Brown, N. A., & Martin, J. F. (2002). Pitx2c patterns anterior myocardium and aortic arch vessels and is required for local cell movement into atrioventricular cushions. *Development (Cambridge, England)*, *129*(21), 5081–5091. https://doi.org/10.1242/dev.00173

Kitamura, K., Miura, H., Miyagawa-Tomita, S., Yanazawa, M., Katoh-Fukui, Y., Suzuki, R., … Yokoyama, M. (1999). Pitx2 deficiency and abnormal morphogenesis. *Development*, *126*, 5749–5758. Retrieved from http://dev.biologists.org/content/develop/126/24/5749.full.pdf

Shiratori, H. (2006). Conserved regulation and role of Pitx2 in situs-specific morphogenesis of visceral organs. *Development*, *133*(15), 3015–3025. https://doi.org/10.1242/dev.02470

APPENDIX B

This appendix represents my work in leveraging CRISPR/Cas9 genome editing techniques to efficiently generate mutant alleles containing a deletion of the entire noncoding regulatory gene desert to demonstrate in vivo the biological role of the entire noncoding gene desert and to recapitulate Axenfeld Reiger Syndrome (ARS) patients with similar large noncoding desert deletions (Protas et al., 2017; Ansari et al., 2016; Reis et al., 2012; Volkmann et al., 2011) upstream of *Pitx2* but with no mutations within the *Pitx2* gene itself. Here I show generation and germline transmission of the mutant desert deletion allele via generation of F1 heterozygotes from founder lines and subsequent Mendelian ratios on embryonic and neonatal litters **(Fig. B-1, Table B-1; Supplementary Methods)**. Applying Pearson's chi-square test for goodness of fit to these Mendelian ratios reveals a significant lack of live born mutant animals then to be expected (χ^2 (2, N = 22) =11, p <0.005). Together with similar but inconclusive results from isolation of het x het embryonic litters, these preliminary findings may suggest that *Pitx2* locus desert deletion homozygous mutants are embryonic lethal. If validated by further investigation, this would implicate the *Pitx2* noncoding gene desert as essential for viability. Finally, I show an additional CRISPR/Cas9 generated mutant allele containing a 24 bp deletion within the gene desert overlapping a constitutive CTCF binding site (validated by ChIP-seq in mES and MEF cells; Kagey et al., 2010) that may be used for future studies on investigating the *in vivo* role of CTCF in directing transcription and/or chromatin architecture at the *Pitx2* locus (**Fig. B-2**)

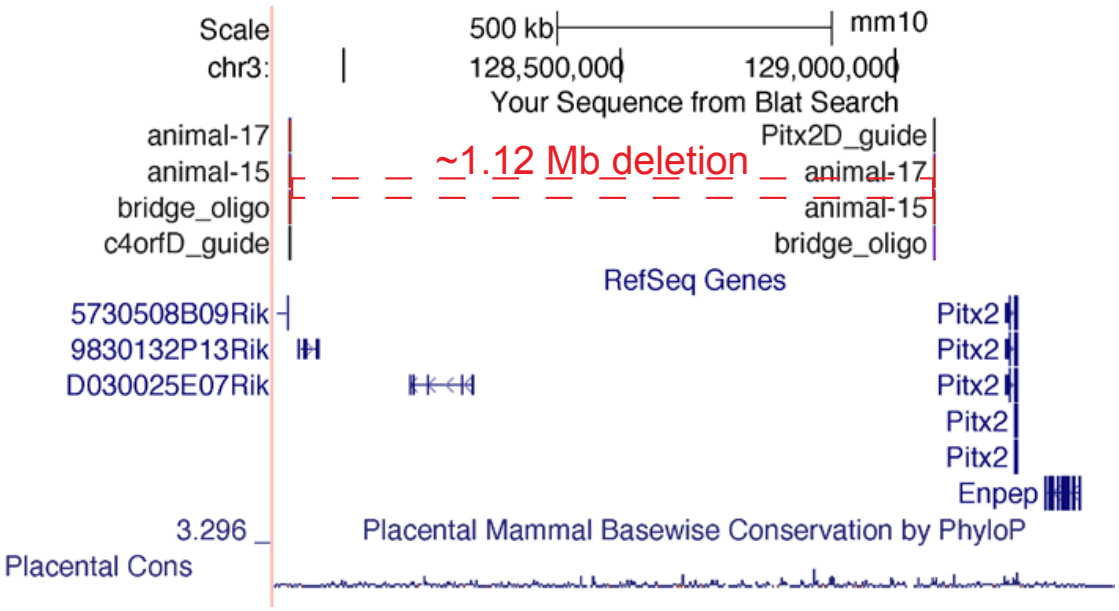


Fig B-1: Successful CRISPR/Cas9 generation of 1.12 Mb noncoding desert deletion at the *Pitx2* locus. Snapshot of UCSC genome browser with squences of: a) PCR products from founder animals (animal-17; animal-15) b) guide RNAs (c4orfD_guide, Pitx2D_guide) and c) bridge oligonucleotide (bridge_oligo) blatted to mm10.

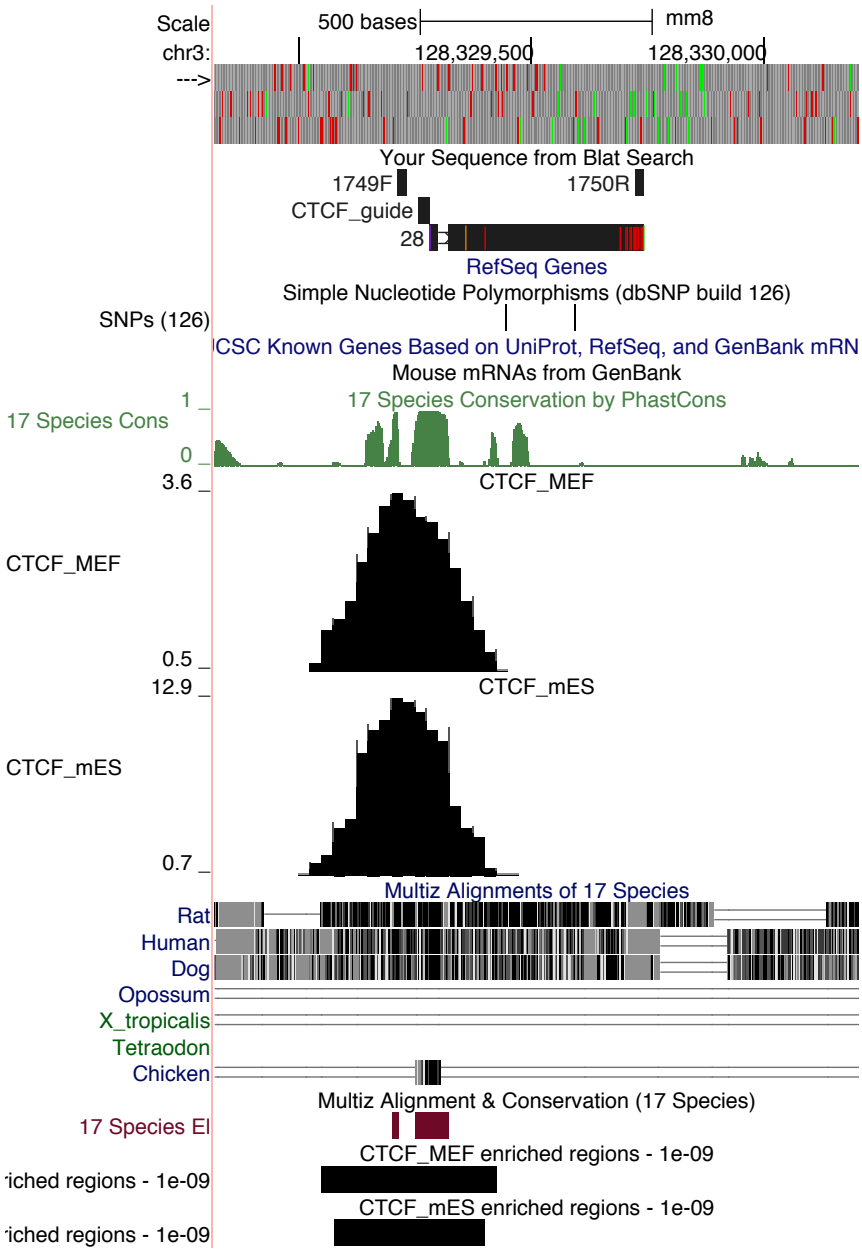


Fig B-2: CRISPR/Cas9 generation of 24 bp deletion within CTCF ChIP-seq binding peak (GEO Series GSE22562; Kagey et al., 2010) within *Pitx2* locus gene desert. Snapshot of UCSC genome browser with squences of: a) PCR products from founder animal (28) b) guide RNA (CTCF_guide) c) genotyping primers 1749F and 1750R blatted to mm8 d) CTCF binding peaks from public dataset (Kagey et al., 2010) in mES cells and MEF displayed to show overlap with CRISPR/Cas9-induced indel.

Strain; genotype cross	# of WT	# of het	# of mut	total #
DD-L17; het x het (live born)	11	11	0	22
DD-L17; het x WT (live born)	11	9	N/A	20
DD-L17; het x het (embryo)	4	9	0	13

Table B-1: Genotype ratios from liveborn and embryonic litters isolated from *Pitx2* locus desert deletion animals line 17 (DD-L17).

SUPPLEMENTARY METHODS

CRISPR/Cas9 targeting of *Pitx2* locus desert deletion and CTCF site perturbation

sgRNAs were designed using the CRISPR Design tool (crispr.mit.edu; Hsu et al., 2013) to target the 5' end of the gene desert flanking, the 3' end of the gene desert flanking Pitx2, and the targeting a previously described (Welsh et al., 2015) CTCF binding site within the gene desert. Primers to synthesize the DNA template for sgRNA transcription were commercially supplied by IDT with the following sequences: c4orf32_desert FP: GAAATTAATACG ACTCACTATAGGACACCGAGTCGACTCGACTTGTTTTAGAGCTAGAAATAG C; Pitx2_desert FP: GAA ATT AAT ACG ACT CAC TAT AGG CAAGCTTTGACCACCACCGGGTTTTAGAGCTAGAAATAGC; and CTCF_desert FP: GAA ATT AAT ACG ACT CAC TAT AGG GCT TCA TCTTAGCTCATGCTGTTTTAGAGCTAGAAATAGC; reverse primer: AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTA TTTTAACTTGCTATTTCTAGCTCTAAAAC. DNA template was synthesized using an overlap PCR strategy according to standard PCR protocols. Template DNA was purified using the Nucleospin PCR purification commercial kit (Machery-Nagel) and used for in vitro transcription (Ambion MEGAshortscriptT7 kit). sgRNA was purified (Ambion MEGAclear kit) and all three sgRNAs were injected (50 ng/ul) along with Cas9 mRNA (25 ng/ul; TriLink) and Cas9 protein (25 ng/ul; Protein PNA Bio) into F1

hybrid (C57Bl/6J x FVB/N) one-cell embryos. Injected embryos were cultured to the two cell stage prior to transfer to recipient. One line was established for each independent mutation.

REFERENCES

Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, Ebmeier CC, Goossens J, Rahl PB, Levine SS, et al. 2010. Mediator and cohesin connect gene expression and chromatin architecture. Nature 467: 430–435.

Protas, M. E., Weh, E., Footz, T., Kasberger, J., Baraban, S. C., Levin, A. V., … Gould, D. B. (2017). Mutations of conserved non-coding elements of PITX2 in patients with ocular dysgenesis and developmental glaucoma. *Human Molecular Genetics*, *26*(18), 3630–3638. https://doi.org/10.1093/hmg/ddx251

Reis, L. M., Tyler, R. C., Volkmann Kloss, B. A., Schilter, K. F., Levin, A. V, Lowry, R. B., … Semina, E. V. (2012, December 9). PITX2 and FOXC1 spectrum of mutations in ocular syndromes. *European Journal of Human Genetics*. https://doi.org/10.1038/ejhg.2012.80

Volkmann, B. A., Zinkevich, N. S., Mustonen, A., Schilter, K. F., Bosenko, D. V, Reis, L. M., … Semina, E. V. (2011). Potential novel mechanism for Axenfeld-Rieger syndrome: Deletion of a distant region containing regulatory elements of PITX2. *Investigative Ophthalmology and Visual Science*, *52*(3), 1450–1459. https://doi.org/10.1167/iovs.10-6060