GENETIC AND FUNCTIONAL STUDIES OF LOCI ASSOCIATED WITH ATRIAL FIBRILLATION

by

SHAMONE ROBINETTE GORE PANTER

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Dissertation Advisor: Jonathan D. Smith, Ph.D.

Department of Molecular Medicine

CASE WESTERN RESERVE UNIVERSITY

May, 2014
We hereby approve the thesis/dissertation of

Shamone Robinette Gore Panter

candidate for the Doctor of Philosophy degree *.

(signed) Dr. Peter Scacheri, Ph.D.
(chair of the committee)

Dr. Jonathan D. Smith, Ph.D.

Dr. Mina K. Chung, M.D.

Dr. Michcalal Aldred, Ph.D.

Dr. Christine S. Moravee, Ph.D.

Dr. David R. Van Wagoner, Ph.D.

(date) February 18, 2014

*We also certify that written approval has been obtained for any proprietary material contained therein.
Dedication

I dedicate this to Jackson

Thank you for allowing your mom to follow her dreams!

Since you have been alive I have been in school. You allowed me to study, write, and go away to conferences and you did not complain (much 😊). Without your sacrifices, this wouldn’t have been possible. I appreciate you being such a great kid!

To Jackson, Kyle, and Rosiland

I challenge all of you to dream big and work hard to take those dreams to greatness!

I love and believe in all of you!

To James

I love you my superman!
# Table of Contents

Acknowledgements............................................................................................................. 1
Abstract................................................................................................................................. 3

I. Introduction .......................................................................................................................... 5
   I.1 Atrial Fibrillation (AF) ..................................................................................................... 5
      I.1.1 Background of AF
      I.1.2 Stroke and AF
      I.1.3 Symptoms and treatment of AF
      I.1.4 Genetics of AF
   I.2 Genome wide association studies ..................................................................................... 10
      I.2.1 Background
      I.2.2 GWAS and AF
   I.3 PITX2c .......................................................................................................................... 12
      I.3.1 Background of PITX2c
      I.3.2 Role of PITX2 during development
      I.3.3 Potential role of PITX2c in AF
   I.4 Long noncoding RNAs and long intergenic noncoding RNAs .............................................. 15
      I.4.1 Background of lncRNAs and lincRNAs
      I.4.2 Functions of lncRNAs and lincRNAs
      I.4.3 Potential role of lincRNAs in AF

II. Atrial fibrillation associated chromosome4q25 variants are not associated with PITX2c expression in human adult left atria appendages ................................................................. 19
   II.1 Introduction .................................................................................................................... 19
   II.2 Material and Methods ...................................................................................................... 21
      II.2.1 Ethics statement
      II.2.2 Genome wide association studies and identification of independent AF associated SNPs at chromosome 4q25.
II.2.3 Human left atrial tissue processing
II.2.4 Genomic DNA isolation and SNP microarray
II.2.5 RNA isolation
II.2.6 cDNA preparation
II.2.7 Quantitative reverse transcriptase-polymerase chain reaction
II.2.8 PITX2c expression analysis
II.2.9 PITX2 eQTL analysis

II.3 Results ............................................................ 26
II.3.1 Patient characteristics of 239 adult left atrial tissue
II.3.2 Four independent SNPs at chromosome 4q25 associated with AF
II.3.3 Covariates affecting PITX2c expression levels
II.3.4 PITX2c cis-eQTL
II.3.5 PITX2c expression is inversely correlated with SHOX2 expression

II.4 Discussion .......................................................... 37

III. Characterization of a left atrial specific long intergenic noncoding RNA adjacent to PITX2c................................................................. 42

III.1 Introduction ........................................................ 42

III.2 Material and Methods............................................ 44
III.2.1 Human left atrial appendages
III.2.2 RNA isolation
III.2.3 cDNA preparation
III.2.4 Quantitative reverse transcriptase polymerase chain reaction
III.2.5 lincRNA expression analysis
III.2.6 PANCe and PITX2c tissue panel expression assay
III.2.7 Overexpression of PANCe by transfection into HEK293 cells and RNA immunoprecipitation
III.2.8 Differentiation of H9 cells to cardiomyocytes
III.2.9 siRNA knockdown of lincRNA and PITX2
III.2.10 RNAseq and analysis

III.3 Results

III.3.1 Discovery of a lincRNA adjacent to the PITX2 gene in human left atria
III.3.2 PANC and PITX2c expression in human tissues
III.3.3 Identification of eQTLs for PANC in human left atria
III.3.4 PANC binds to SUZ12, a polycomb repressive complex 2 protein
III.3.5 PANC and PITX2c coordinately expressed during cardiomyocyte differentiation of human H9 embryonic stem cells
III.3.6 PANC and PITX2c knockdowns in H9 derived cardiomyocytes
III.3.7 PANC and PITX2c KO induce significant global gene expression changes.

III.4 Discussion

IV. Unraveling the genetics of Atrial Fibrillation

IV.1 Conclusions and future directions

IV.1.1 AF SNPs in the 4q25 region and there possible effects on PITX2c expression
IV.1.2 PANC and its regulation of gene expression
IV.1.3 Final thoughts

V. Appendix

Permission of use forms

VI. Bibliography
List of Tables

2.1 PITX2c TaqMan primer and probe set ................................................................. 24
2.2 Left atrial appendage surgical and donor patient characteristics ......................... 27
2.3 4 q25 SNPs independently associated with AF ..................................................... 30
2.4 Linkage disequilibrium (r²) of 4 risk/independently associated AF SNPs in 4q25 region ...................................................................................................................... 30
2.5 Multivariate model for PITX2c expression ............................................................. 32
2.6 Chr. 4 AF risk SNPs not associated with PITX2c expression normalized to ACTC1 .... 33
2.7 Chr. 4 AF risk SNPs not associated with PITX2c normalized to PPIA expression ....... 33
2.8 Power analysis for PITX2c expression .................................................................... 34
2.9 P-values for AF SNPs association with PITX2c expression in individual AF history/rhythm groups ................................................................................................. 34
2.10 Linkage disequilibrium (r²) between significant PITX2ceQTL SNPs located in the ENPEP gene ............................................................................................................... 36
3.1 LincRNA Taqman primer and probe ........................................................................ 45
3.2 Custom sequences for lincRNA siRNA ..................................................................... 48
3.3 Left atrial appendage surgical patient characteristics .............................................. 53
3.4 AF SNPs in the chromosome 4q25 region not associated with expression of PANCR . 54
3.5 PANCR KD top 20 differentially expressed genes ..................................................... 62
3.6 PITX2 KD top 20 differentially expressed genes ...................................................... 63
3.7 PANCR and PITX2 simultaneous KD top 20 differentially expressed genes .......... 64
4.1 Patient characteristics of PV-LA cuff trimmings ....................................................... 65
4.2 AF SNPs associations to PITX2c and PANCR normalized to ACTC1 ..................... 65
List of Figures

1.1 Cis-acting quantitative trait loci ................................................................. 11
1.2 Schematic of PITX2 isoforms ..................................................................... 13
1.3 Functions of lnc/lincRNAs in the human cell .............................................. 17
2.1 Adjusted and unadjusted expression of PITX2c in human left atrial appendages in AF controls .............................................................................................................. 27
2.2 Identification of four SNPs independently associated with AF at the 4q25 locus in the Cleveland Clinic Lone AF GWAS .................................................................................... 29
2.3 PITX2 expression was associated with AF history and rhythm status .......... 31
2.4 QQ plot of PITX2c eQTLs in the 4q25 region ............................................... 35
2.5 PITX2c eQTLs for SNPs in the 4q25 region .................................................. 35
2.6 Correlation analysis of PITX2c with SHOX2 expression in human left atrial appendages. .......................................................................................................................... 36
3.1 Overexpression of lincRNA in HEK293 cells .............................................. 47
3.2 Left atrial specific uncharacterized lincRNA discovered by RNAseq ............ 49
3.3 PANCR shares left/right expression pattern in human atria ...................... 50
3.4 Tissue specific expression of PANCR ......................................................... 51
3.5 Tissue specific expression of PITX2c ......................................................... 52
3.6 PANCR and PITX2c are positively correlated in human adult left atrial appendages ... 53
3.7 RNA immunoprecipitation of PANCR transfected HEK293 cells .............. 55
3.8 PANCR and PITX2c are coordinately induced during differentiation .......... 56
3.9 Knock down of PANCR, PITX2c and the combined siRNA in differentiated cardiomyocytes ............................................................................................................ 57
3.10 Venn diagram of significant altered overlapping genes ............................ 58
4.1 Effect size of minor allele on PITX2c expression in PVTs ......................... 66
4.2 PITX2c expression in PVTs is significantly higher than AF/SR .................. 66
ACKNOWLEDGEMENTS

First and foremost I have to give honor to God for all of the blessings that have been bestowed upon me and for giving me the strength to persevere through the many trials that have come my way. I wish to thank my advisor Dr. Jonathan Smith for allowing me the opportunity to join his lab, work on this interesting and challenging project and for his constant support and guidance through this journey, I am forever grateful. I must thank the members of my thesis committee and the AF genetics group; their constructive criticism, support, time and expertise helped me to successfully move through this journey. Dr. Moravec, thank you for giving me my first opportunity in translational research and continuing to mentor me over the years, I appreciate all you have done to help me to get to where I am today. Dr. Van Wagoner and, Van Wagoner lab and the Moravec Lab, thanks for all of your support and your help with obtaining tissue samples and RNA. Dr. Krontiris-Litowitz and Dr. Lovelace-Cameron, thank you all for your support throughout my academic career, I truly appreciate it. Dr. Reizes, thanks for the advice when it came to choosing graduate school and the opportunity I had to help start your lab, the experience was invaluable. Dr. Clemencia Colmenares, thank you for all that you did to help me really understand gene expression and all of the wonderful advice and support that you have given me, I truly appreciate it. I am blessed to have wonderful labmates in the Smith Lab; I thank all of you for your camaraderie and support. I especially want to thank Jeff Hsu, I appreciate your awesomeness in analyzing data, teaching me R studio (even if I still can’t do it without your help©), helping me move....twice, and providing comic relief, you
rock! Stela, thank you so much for your wonderful advice and support over the years, you are not only a colleague but a great friend! I want to thank the department of Cellular & Molecular Medicine and its front office; all of you have been so helpful, kind and supportive over the years. I especially want to thank Raquel, you are one of my very best friends and as close to me as a sister, I love you and thank you for not letting me give up! Thank you to the Molecular Medicine PhD program and the research education office. This program challenged me to grow as a scientist and has provided me great support and assistance throughout my time as a graduate student.

To my amazing husband James, thank you for joining me on this journey. I am so blessed to be building a life with you, I love you and I appreciate all that you do for our family! Jackson, Kyle and Lindi, I love you guys with all of my heart, and pray that all of you grow up to do more than me and be better than me! To the Gore’s, Young’s Anderson’s, Gilbert’s, Lisbon’s, Linden’s, Panter’s, Lindsey’s Landes’s, Assal’s Vaughn /Clark’s, Miller/Harris’s, Krasienko’s and Davis’s, you ALL are my family and I appreciate the love and support you all have shown me. I am truly blessed to have such a wonderful family, you all are the wind beneath these wings, and I love you all. Last but certainly not least, I have to thank my Mom Jacqueline Gore, my grandmother Mattie Lisbon, and my Aunt Juanita Young, my angels in heaven. From day one, you all believed that I could and would do great things, I love and miss you all every day and pray that I will live up to the examples of strong black women that you all set for me, because of you I am able to soar.
Genetic and Functional Studies of Loci

Associated with Atrial Fibrillation

Abstract

by

SHAMONE ROBINETTE GORE PANTER

Atrial Fibrillation (AF) is the most common sustained arrhythmia, and is associated with an increased risk of mortality, morbidity and stroke. Genome wide association studies (GWAS) have identified that the single nucleotide polymorphisms (SNPs) most strongly associated with AF are located on chromosome 4q25 in an intergenic region that is closest to the PITX2 gene. The PITX2c isoform expressed specifically in the left atria plays a role in cardiac left/right asymmetry during development and hemizygous knockout mice are susceptible to pacing induced arrhythmia. The Cleveland Clinic Lone AF GWAS identified four independent AF risk SNPs at the chromosome 4q25 locus, with the most significant SNP identified located ~171kb distal to PITX2c. In addition to PITX2c, our group identified an uncharacterized PITX2 Adjacent long intergenic Noncoding RNA (PANCR), ~2 kb proximal to PITX2. We hypothesized that the AF associated SNPs located on chromosome 4q25 might directly affect expression of PITX2c and/or PANCR. We determined that PITX2c and PANCR levels were positively correlated with each other in 223 left atrial appendages. Expression of both genes was examined in a panel of 33 human tissues, and both were highly expressed in left atria and eye. Using left atrial
appendages surgically obtained from subjects of European ancestry, we obtained SNP genotypes via microarray and measured PITX2c and PANCR expression via quantitative RT-PCR (qRT-PCR). We found that the AF risk SNPs were not associated with PITX2c or PANCR expression. During differentiation of H9 human embryonic stem cells into cardiomyocytes, both transcripts were induced early prior to expression of cardiac troponin, suggesting they are coordinately expressed. Knock-down of PANCR in differentiated H9 cardiomyocytes led to decreased expression of itself and of PITX2c. RNAseq after knock-down of PITX2 and PANCR independently and simultaneously revealed a large number of genes that were differentially altered by the different conditions. We conclude that expression of PITX2c and PANCR in human adult left atrial appendages is not associated with the 4q25 AF risk SNPs. We speculate that these risk SNPs may alter expression of these genes in a different location or during cardiac development, and we plan to study this using differentiating H9 cells.
CHAPTER 1

Introduction

1.1 Atrial Fibrillation

1.1.1 Background of Atrial Fibrillation (AF)

Atrial fibrillation (AF) is a complex disease and the most common cardiac arrhythmia encountered in clinical practice.\(^1,2\) It affects more than 3 million Americans with millions more being affected worldwide; the incidence of AF increases rapidly with age with 1 in 4 individuals likely to develop AF in their lifetime.\(^1,3\) There are several known clinical risk factors for AF such as valvular heart disease, congestive heart failure, hypertension, myocardial infarction and diabetes mellitus.\(^3,4\) Additionally, structural abnormalities such as left atrial enlargement and left ventricular hypertrophy may also increase the incidence of AF.\(^3\) It is important to note that some patients develop AF without any other known risk factors which is known as lone AF.\(^4\) AF is associated with a significantly increased risk for morbidity and mortality and a 4- to 5-fold increase risk for stroke, accounting for one third of all strokes in patients over the age of 65.\(^5\)

Atrial Fibrillation is described as an uncoordinated contraction of the atria propagating to the ventricles instead of a coordinated contraction from the sinoatrial (SA) node to the atroventricle (AV) node, down the purkinje bundles through to the ventricles. The underlying mechanisms of AF are complex; however multiple re-entrant circuits or multiple rapidly discharging foci have long been accepted as theories to explain AF.\(^4,6\)
Stability of the multi re-entry theory is dependent on the wavelets in the atria. During a contraction, a wavelength is the distance traveled during 1 refractory period and this is the base unit of a traveling wavelet. Thus shorter wavelengths result in more wavelets in the atria and more sustained AF. Conversely, atrial ectopic foci, which originate in groups of cells not located in the SA node, are spontaneously firing regions that may derive from increased automaticity, micro-re-entry and after-depolarization; and, these triggers may initiate re-entrant circuits in the atria when conduction blocks in the atria are present.

Atrial ectopic foci that originate in the pulmonary vein or in the region between the left and right atria may trigger AF. These foci are the site of several treatments such as radiofrequency ablation which will be described in a later section.

Patients typically present with paroxysmal AF, meaning AF from time to time that self terminates, probably because of sporadic atrial ectopic episodes. When these repeated episodes become more consistent, the atria undergo electrical and structural remodeling resulting in the maintenance of AF (possibly multiple re-entry) eventually leading to persistent AF, which is AF that last ≥ 7 days. Additionally, AF can progress to long-standing persistent AF, which means the AF symptoms have been continuous for greater than a year and permanent AF in which the patient is left in AF due to failure of rhythm control efforts.

I.1.2 Stroke and Atrial Fibrillation

Atrial fibrillation is a major risk factor for stroke. Stroke is a medical condition that can be classified into two major categories, ischemic or hemorrhagic; ischemic stroke occurs as a result of a blockage in blood flow to the brain as a result of a clot/embolism being lodged
in a main vein to the brain. Hemorrhagic stroke is a condition of essentially bleeding of
the brain either into the skull, intracranial hemorrhaging, or within the brain tissue itself,
intracerebral hemorrhaging. In AF, clots form as a result of incomplete emptying of the
upper chambers of the heart (atria), forming pools of blood in these chambers, and more
specifically in the atrial appendages. These clots can then break from the main location
and travel to the brain causing ischemic stroke. The risk of stroke in patients with AF
increase significantly with age, 5% in patients over 65 and 9% in patients over 80.59
Additionally, in patients over the age of 75, the single most important cause of ischemic
stroke is AF.10

I.1.3 Symptoms and treatments of AF

Patients with AF may be asymptomatic without careful diagnoses; those who do have
symptoms may experience variable symptoms and severity such as palpitations, decreased
blood pressure, dyspnea, dizziness, weakness, lightheadedness, confusion, or chest pain.11
There are two approaches used when managing a patient with AF. One involves
controlling ventricular rate and anticoagulation while allowing atrial fibrillation to persist
and the second involves restoring and maintaining sinus rhythm.12,13 Ventricular rate and
anticoagulation are usually controlled by medication. Digoxin, beta-blockers and calcium
channel blockers are used to control ventricular rate. Warfarin, a vitamin K antagonist
that alters the levels of coagulation factors that are vitamin K dependent, was the gold
standard of anti-coagulation drugs; however, the potential side effects and slow rate of
action have made it widely under used. More recently, two new classes of drugs that
directly inhibit coagulation factors thrombin or factor Xa, have been created. Dabigatran,
a thrombin inhibitor, and rivaroxaban and apixaban, factor Xa inhibitors, provide safer, more convenient oral anticoagulation therapy, which all reduce the likelihood of stroke. In contrast, cardioversion (first performed in 1962), both electrical and pharmacologic, are used to restore an abnormal or fast heartbeat to a normal rhythm. Electrical cardioversion involves giving the heart electrical pulses to restore its rhythm, it should be noted the patient is sedated prior to the pulses being applied. Pharmacologic cardioversion involves using a variety of anti-arrhythmic drugs (AAD) as a way to restore and maintain sinus rhythm (SR). In the AFFIRM study, they concluded that there was no benefit of managing AF by rate control, which is defined as controlling the ventricular response rate of AF as compared to rhythm control, which is defined as the maintenance of sinus rhythm; however, the use of anticoagulants significantly reduced the chances of stroke. The MAZE procedure, first performed by J.L. Cox M.D. in 1987, is said to be the most effective surgical treatment for AF. There have been 4 iterations of the MAZE procedure to primarily improve sinus tachycardia in response to maximal exercise as well as reduce left atrial dysfunction. Physicians are moving away from surgical ablation methods to catheter ablation methods, thus the MAZE surgery is performed less often than in the past, despite its high success rate. Catheter ablations were developed after specific trigger locations for initiation of AF were found. It is a technique that uses radio frequency energy to ablate locations of ectopic foci. If there is pro-arrhythmic activity located between the left atrium (LA) and pulmonary vein (PV), ablation of this region, PV isolation, can be performed. PV isolation is usually successful in patients with paroxysmal AF; however, in patients with persistent AF, LA linear ablation is necessary.
Ablation is a procedure in which linear lesions are placed at the roof and mitral isthmus and these are used to prevent large atrial re-entrant circuits.\textsuperscript{15,19} There is no one treatment that is effective for all patients so combinations of the above described treatments can be used.

### 1.1.4 Genetics of AF

The genetic basis of AF is an exciting, yet enigmatic, field of cardiovascular genetics. In 2004, AF heritability was described in the large Framingham Study cohort. They found that in offspring with at least one parent diagnosed with AF, the risk for AF increased (odds ratio (OR) =1.85; 95% confidence interval (CI) 1.12-3.06; p= 0.02).\textsuperscript{20} However, this level of increased risk was for subjects that also had other risk factors for AF, for patients with a history for lone AF, the genetic risk is thought to be even greater. With the knowledge that AF is in fact partly heritable, determining the causal genetic variants and how these variants increase AF susceptibility is critical to understanding the genetics of AF.

There are two major forms of AF, familial AF, which is rare and occurs as a result of Mendalian transmission, or common/non-familial AF. Mendalian families with clear and highly penetrant inheritance of a single causal gene, like those described below, have provided a great deal of information on AF pathophysiology allowing researchers to locate several causative mutations in familial AF.\textsuperscript{21} Numerous chromosomal locations have been identified by linkage studies in familial AF. Chromosome 10q22-24 was identified in 3 families with autosomal dominant transmission; however, a causative gene for these families was not identified.\textsuperscript{22} Studies in specific individuals of Asian descent discovered an AF locus on chromosome 11p15.5 coding for the KCNQ1 gene, which is the α sub-unit of
the delayed rectifier potassium channel.\textsuperscript{23,24} In the same cohort used in the previous study, an additional locus at chromosome 21q22 was identified encoding for \textit{KCNE2}, another potassium channel.\textsuperscript{25} Xia et al. and Hong et al. identified mutations in the \textit{KCNJ2} gene, which encodes for the Kir2.1 channel that mediates an inward rectifier potassium current in the heart, and in the \textit{KCNH2} gene, which encodes for the HERG protein the $\alpha$-subunit of the cardiac IKr channel, respectively.\textsuperscript{26,27} In a subsequent study to identify similar mutation in individuals of European ancestry, Ellinor et al. screened for mutations in \textit{KCNJ2} and \textit{KCNE1-5} genes in 96 subjects and no mutations were identified, indicating that potassium channel mutations may be dependent on ethnicity.\textsuperscript{28} More recently, rare variants in the human \textit{HCN4} gene that affect protein expression have been shown in family cases (Ellinor et al., unpublished). Familial AF accounts for the minority of all AF subjects, thus heritability in most AF subjects is due to common susceptibility alleles.\textsuperscript{20,21,29}

Several genome wide association studies (described in section I.2), which are used to study common non-familial AF cases, have identified a strong locus for AF disease susceptibility on chromosome 4q25. The odds ratio (OR) for the minor allele ranges from $\sim 1.4$ to 2.03 in individuals of European and Asian descent.\textsuperscript{30-32} Among the single nucleotide polymorphisms (SNPs) in this region, several independent risk variants have been identified with the strongest being rs2200733 with an odds ratio of $\sim 1.9$.\textsuperscript{33} Currently no genes have been found to be causative in this region; however, we and others have investigated whether \textit{PITX2c} (described in section I.3.1 and chapter II), which is the closest gene located $\sim 171$ kb proximal to the rs2200733 SNP, may be controlled by this and the other SNPs found in this region.\textsuperscript{32}
I.2. Genome Wide Association Studies

1.2.1 Background of Genome wide association studies

Genome wide association studies (GWAS) examine thousands of common genetic variations or single nucleotide polymorphisms (SNPs) in hundreds to thousands of people to see if there are any variants associated with a disease or complex traits.\(^{34}\) It has been shown that common SNPs in non-coding regions may affect complex traits or disease by regulating expression of nearby genes, otherwise known as cis-acting quantitative trait loci (Figure 1.1). GWAS have uncovered many SNPs that confer disease susceptibility,\(^ {35}\) with the first successful GWAS to show association to disease published in 2005. In that study, two polymorphisms in the complement factor H gene were found to be strongly associated with age-related macular degeneration, which is a major cause of blindness in the elderly.\(^ {36}\) Since that time numerous studies have identified SNPs that are significantly associated with >150 distinct diseases and traits.\(^ {34}\)

1.2.2 GWAS and AF

The first study to show association of AF to a specific region was performed by Gudbjartsson et al. in 2007. In this study, Icelandic cases from patients diagnosed with AF
and/or Atrial flutter (AFL) from 1987-2005 were used to reveal a strong association of AF with SNPs on chromosome 4q25. Two SNPs were considered to be associated with AF based on OR and p-value; rs2200733 (OR=1.75; p=1.6 x 10^{-10}) and rs10033464 (OR=1.42; p=.0024). Since those original studies, several additional GWAS have confirmed the above mentioned SNPs in the 4q25 region and uncovered additional independent SNPs that are significantly associated with AF. Ellinor et al. identified 10 loci that are significantly associated with AF. The top 3 loci, which are located on chromosomal regions: 4q25 near PITX2c, 16q22 in the ZFHX3 gene and 1q21 in the KCNN3 gene were identified in earlier studies. The additional 7 loci are located on chromosomal regions: 1q24 in the PRRX1 gene, 7q31 in CAV1, 14q23 in the intron of SYNE2, 9q22 in an open reading frame (ORF) of chromosome 9 (C9orf3), 15q24 in the first intron of HCN4, 10q22 just upstream of SYNPO2L and 5q31 in the intron of WNT8A; although, the last gene was not replicated in additional AF cohorts. Identification of these SNPs is important because they may allow for identification of causal genes for AF thus leading to potential targets for therapies and preventions.

1.3 **PITX2**

1.3.1 **Background of PITX2**

PITX2, a member of the bicoid class of homeodomain proteins, is a transcription factor that is expressed in many organs throughout the body, including the heart and brain. PITX2, which was then called the RIEG gene, was first identified as mutated in 1996 from
patients with Axenfeld-Rieger syndrome, a congenital disease with malformations of the face, teeth, and skeletal system. There are three major isoforms of PITX2 that have been identified in various vertebrates; PITX2a, PITX2b, and PITX2c, and 1 minor isoform identified exclusively in humans, PITX2d. PITX2a and PITX2b occur as a result of alternative splicing, while PITX2c uses an alternative promoter located upstream of the 4th exon. PITX2d is generated using the PITX2c alternative promoter and alternative splicing (Figure 1.2). The N-terminus of all the isoforms are different while the C-terminus and homeodomain are identical.

I.3.2 Role of PITX2c during development

PITX2 has been shown to be important in left/right signaling of vertebrates during development. Pitx2−/− knock-out mice develop severe cardiac malformations and are thus embryonic lethal, while mice that are Pitx2+/− show defects in left/right asymmetry, but are viable, further confirming that Pitx2 is an important gene during cardiac development. Additionally, several groups have demonstrated that specific isoforms may play a larger role in cardiac asymmetry during development. Mouse knockout experiments have shown that the Pitx2c isoform is: 1) required for left-right asymmetry during cardiac
development, 2) critical for left atrium development, and 3) behaves in a dose-dependent manner during organ morphogenesis, suggesting that only a small amount is necessary during atrium formation. Mommersteeg et al. found that a complete loss of Pitx2c in mice resulted in sinoatrial nodes that essentially became indistinguishable between the left and right atrium in fetal mice, mainly due to a failure to suppress a default program for sinoatrial node formation in the left atria. All of these studies provide solid evidence that Pitx2 is critically important during development and that the Pitx2c isoform is necessary for cardiac development and left-right asymmetry.

I.3.3 Potential role of PITX2c in AF

As mentioned in the above sections, multiple GWAS have found SNPs in the 4q25 region that are strongly associated with AF susceptibility. PITX2 and more specifically PITX2c, which is the closest gene to this region, has been implicated as a possible gene target for these SNPs. Prior to 2010, no studies provided functional evidence for PITX2c’s role in AF. In 2010, Wang et al. published a manuscript implicating Pitx2 and the Pitx2 mediated signaling pathway in of AF and other arrhythmias. Utilizing LacZ Pitx2 knock-in mice, Pitx2 expression in postnatal day three was found only in the left atrium and the pulmonary vein, which are both regions implicated in AF. Expression of Pitx2 was much lower in the left atrium at one year of age. Then using Pitx2c+/− LacZ knock-in mice, they showed that expression of Pitx2c was primarily expressed in the left atrium and confirmed that Pitx2c was the primary isoform expressed in the left atrium. The most crucial experiment used hemizygous Pitx2+/− mice. They were able to induce an AF-like phenotype.
in these mice by atrial pacing.\textsuperscript{50} Using microarray and qRT-PCR, they were able to show that genes such as Shox2 and Tbx3, which are important in sinoatrial node function and pacemaker activity, were up-regulated when Pitx2c was knocked down indicating that Pitx2c inhibits pacemaker activity thus playing a role in AF and arrhythmias susceptibility.\textsuperscript{50}

I.4 Long noncoding RNAs (IncRNAs) and long intergenic noncoding RNAs (lincRNAs)

I.4.1 Background of IncRNAs and lincRNAs

Long noncoding RNAs (IncRNA) are generally described as mammalian transcripts longer than 200 nucleotides that are 5’capped and have a polyadenylated tail similar to mRNA yet they do not code for a functional protein. There have been thousands of IncRNAs discovered although they tend to be weakly conserved between species they are decently conserved among mammals.\textsuperscript{51,52} There have been several categories of IncRNAs described in the literature and they tend to be defined by their location in relation to protein-coding genes.\textsuperscript{51} Intronic IncRNAs are located in the introns of protein-coding genes and can go in either direction. Antisense IncRNAs begin in the 3’ or within a protein-coding gene but are transcribed in the opposite direction of the gene. Sense IncRNAs are similar to antisense IncRNAs but are transcribed in the same direction of the gene. Bidirectional IncRNAs are initiated in a divergent manner from the promoter of the nearest protein-coding gene. Lastly intergenic IncRNAs also known as long intergenic noncoding RNAs (lincRNAs) are transcribed in the genomic regions between genes and typically have independent transcriptional units from their neighboring genes.\textsuperscript{51,53} Our group has
identified an uncharacterized lincRNA and it will be discussed further in section IVc and chapter 3.

**I.4.2 Functions of lncRNAs**

Thousands of lncRNAs have been discovered in the mammalian genome, and of those greater than 3500 have been classified as lincRNAs. While a significant portion of the genome is understood to encode lncRNAs, only a few of these have been well-characterized. Those include XIST and TSIX which are important in X-chromosome inactivation, and H19, AIR, and HOTAIR, which are responsible for imprinting. Over the last 5-10 years, a great deal of progress has been made in understanding the functions of lncRNAs. In 2009, Khalil et al. and others have shown that ~20% of lincRNAs that are expressed, such as ANRIL, are bound to the polycomb repressive complex (PRC) 2, while some are bound to other chromatin modifying complexes. This is important because PRC2, which is involved in numerous biological processes, helps to maintain gene-expression patterns through epigenetic modification of histones, and if disrupted can lead to down or up-regulation of genes. There are additional classes of lncRNA that control gene expression. Rinn et al. showed that the lncRNA HOTAIR, which is transcribed within the HOXC gene locus on chromosome 12, effects gene expression of EZH2, a member of the PRC2, by modulating its chromatin state in trans. Ørom et al. found that several lncRNAs displayed an enhancer-like function in human cell lines. For example, they showed that ncRNA-a7 was a regulator of Snai1, which is a member of the Snail zinc finger family, important in cell adhesion and migration. LncRNAs have also been shown to function as miRNA sponges by binding to and reducing the effects of certain miRNAs.
on their miRNA targets.\textsuperscript{61} Splicing, cell signaling, and apoptosis have also been shown to be regulated by lncRNAs.

MALAT1, a lncRNA that is upregulated in small cell cancer, was shown to regulate alternative splicing by interacting with the nuclear phosphoprotein family of proteins which are involved in the splicing.\textsuperscript{56} Tripathi et al. found that MALAT1 modulates the expression of cell cycle genes and is critical for cell division.\textsuperscript{62}

Additionally, another lncRNA, growth-arrest-specific 5 (Gas5), was shown to sensitize cells to apoptosis by limiting the action of glucocorticoids on the glucocorticoid receptor during nutrient starvation.\textsuperscript{61,56} Translation has also been shown to be regulated by lncRNAs.

Faghihi et al. showed that in Alzheimer’s disease, an antisense lncRNA BACE1-AS, may be responsible for setting into motion a feed-forward cascade that causes an increase in BACE1 that ultimately results in the toxic accumulation of Aβ plaques in the brain (Figure 1.3).\textsuperscript{56,64}

### I.4.3 Potential role of lincRNAs in AF

Our lab has identified an uncharacterized lincRNA in the 4q25 AF risk region using RNA sequencing (RNAseq) of 4 human left/right atrium pairs.\textsuperscript{65} This lincRNA is a 19 kb gene
with 2 exons, 446 nucleotide transcript and a potential 35 amino acid long open reading frame. It is located 3’ to the PITX2 gene, and is differentially expressed in the left vs right atrium, sharing the same expression pattern as PITX2c. This finding has been confirmed using both RNAseq and qRT-PCR of left and right atrium. Because of its location relative to PITX2c, we have named it PANCR, *PITX2c* Adjacent long intergenic Noncoding RNA. PANCR is not conserved in mice expect for a 55bp region, and RNAseq analysis of mouse hearts has shown no expression of PANCR. However, this lincRNA is well conserved in all primates. We found that knockdown of PANCR in cardiomyocytes derived from H9 human embryonic stem (ES) cells leads to decreased expression of not only itself but also expression of *PITX2c* (see chapter III). However, the expression of PANCR is not down-regulated when *PITX2* is knocked-down in these cells. As stated in previous sections, *PITX2c* is the closest gene to the region most strongly associated with increased risk of AF, the chromosome 4q25 region. We have shown that in human adult left atrial appendage tissue, the SNPs in the 4q25 region are not associated with *PITX2c* expression (chapter II). Additionally, we have shown that in these same tissues, PANCR expression is not associated with the SNPs in the 4q25 region. As mentioned with *PITX2c*, one of the limitations to our studies is that we may be looking during the wrong time/wrong location to find an effect of the AF SNPs on PANCR expression (discussed further in chapter III).
CHAPTER II:

Atrial fibrillation associated chromosome 4q25 variants are not associated with \textit{PITX2c} expression in human adult left atria appendages


II.1 Introduction

Atrial Fibrillation (AF), a complex cardiac arrhythmia, is the most common sustained arrhythmia encountered in clinical practice and affects over 2.3 million Americans and millions more worldwide.\textsuperscript{1} AF is associated with a 2-fold increase in mortality and 4- to 5-fold increased risk for stroke, resulting in significant cost to the healthcare system.\textsuperscript{1,66} Previous genome wide association studies (GWAS) have found that the strongest single nucleotide polymorphisms (SNPs) associated with AF are located on chromosome 4q25, with the AF odds ratio for the minor allele ranging from \textasciitilde1.4 to 2.0.\textsuperscript{30-32} Among the AF-associated SNPs on chromosome 4q25, several independent risk variants have been identified.\textsuperscript{38} These SNPs are located in an intergenic region of chromosome 4q25 with the
closest gene, PITX2, located ~171 kb proximal to the most significant SNP identified in the earliest GWAS, rs2200733.32

PITX2, a member of the bicoid class of homeobox genes, is expressed in many organs throughout the body, including the heart and brain.41 There are 3 major human isoforms, PITX2a, PITX2b, and PITX2c. PITX2c is the only isoform expressed in the left atrium, and it has been shown to play a critical role in left/right asymmetry during development in the heart.46 Using RNAseq, we have previously shown that PITX2c is expressed in the human left atrial appendage and not in the right atrial appendage.65 Although homozygous Pitx2c deficiency in mice results in embryonic lethality,67 hemizygous Pitx2c+/− mice are viable and susceptible to pacing-induced atrial arrhythmia, making it an attractive candidate as an AF-causal gene.50 Many SNPs identified in GWAS for complex traits such as anthropomorphic measures and common diseases are located in intergenic regions and thus it is challenging to determine the mechanism for the observed associations. The overriding hypothesis is that these SNPs affect complex traits by regulating the expression of nearby genes, thus these SNPs may be classified as cis-acting expression quantitative trait loci (eQTLs). It is also challenging to determine which are the causal SNPs, as the index SNPs may not be causal but in linkage disequilibrium (LD) with causal SNPs. Recent findings from the ENCODE project show that most of the GWAS SNPs, or SNPs in strong LD with the GWAS SNPs, are within regulatory regions. These regions were identified due to their location within a DNaseI hypersensitive region or within a site in which histone modification or transcription factor binding indicates regulatory activity, as determined by chromatin-immunoprecipitation and next generation sequencing.68 We
hypothesized that the four SNPs independently associated with AF in the chromosome 4q25 region would be associated with PITX2c expression. To study this, we used SNP arrays to obtain genotypes and qRT-PCR to measure PITX2c expression in left atrial appendages obtained from 239 subjects of European ancestry, including 40 samples from subjects with no history of AF. We found that these AF-associated SNPs were not associated with PITX2c mRNA expression in adult left atrial appendages in all subjects combined, or in the subgroup of subjects with no history of AF. However, we identified several SNPs in introns of the ENPEP gene, on the opposite side of the PITX2 gene relative to the location of the AF-associated SNPs, which were modestly associated with PITX2c mRNA expression levels. Thus, the mechanism of the AF-associated SNPs on 4q25 remains unknown.

II.2 Materials and Methods

II.2.1 Ethics Statement

All patients provided informed consent for use of discarded atrial tissue. Prior to 2008 verbal consent was obtained and documented in the medical records in a process approved by the Cleveland Clinic Institutional Review Board (IRB). From 2008 onward and for donor tissues, patients provided separate IRB-approved written informed consent. The IRB approved the studies included in this report. Subjects were not consented for genetic information sharing, thus the SNP data cannot be loaded to dbGAP.
II.2.2 Genome wide association studies (GWAS) and identification of independent AF associated SNPs at chromosome 4q25

Descriptions of the cohorts used in the Cleveland Clinic lone AF (LAF) GWAS, genotyping, and quality control and filtering of samples and SNPs were previously described. Here we include an additional 111 Caucasian LAF cases from the Cleveland Clinic Lone Atrial Genebank, all typed on the Illumina Hap610 chip, giving a total of 607 LAF cases and 2956 population controls for GWAS and fine mapping analysis. Maximum likelihood logistic regression was used to estimate the association between odds of LAF and each of the 479,618 filtered genotyped SNPs. SNPs were coded as allele dosages. The logistic model included sex and four principal components of genetic sharing, all of which were associated with LAF. LAF-SNP association p-values were adjusted for any residual population stratification using the genomic control method. Odds ratios of LAF were estimated for each SNP.

Results from our LAF GWAS were used to find independent LAF-SNP associations in the 4q25 region. Starting with SNP rs2200733, which was previously reported as the strongest AF-associated SNP in a GWAS and had the largest odds ratio in our LAF GWAS, we searched for additional SNPs in the vicinity of rs2200733 (150 SNPs over 1 Mb) that were independently associated with LAF using a forward step-wise approach. SNPs were added until the conditional p-value, adjusting for all SNPs chosen in the step-wise search, was less than 0.01 with Bonferroni correction for 150 SNPs (p<6.67x10^-5).

II.2.3 Human Left Atrial Tissue Processing
Human left atrial appendage tissues obtained from elective surgery were snap frozen in liquid nitrogen and stored at -80°C until RNA extraction. AF history, type of AF, structural heart disease, demographics, and other clinical data were collected in a research database and a prospectively collected database of all cardiac surgeries (The Cardiovascular Information Registry) maintained by the Department of Cardiothoracic Surgery. Subjects were categorized as “lone AF” if they had a history of AF and did not have coronary artery disease or valvular disease. AF rhythm status was determined by review of electrocardiograms obtained prior to surgery. 16 left atrial tissue specimens were obtained from non-failing donor hearts not used for transplant. These hearts were perfused with cardioplegia prior to explant and processed in the same manner as hearts used for organ transplant. As with the surgical specimens, donor tissue samples were snap frozen in liquid nitrogen and kept at -80°C until RNA extraction.

II.2.4 Genomic DNA isolation and SNP microarray

25-50 mg of left atrial appendage tissue was used to extract DNA. The tissue, in one mL of DNAzol® (Invitrogen,), was homogenized (PowerGen700, Fisher Scientific) with sterile Omni Tip Disposable Generator Probes (Omni International,). DNA was isolated from the homogenate following the manufacturer’s protocol. The DNA pellet was resuspended in 20 µl of 10 mM Tris buffer (pH 7.4) and the DNA concentration was measure with a NanoDrop ND-1000 Spectrophotometer (Thermo Fisher Scientific Inc.), diluted up to 100 ng/µl and stored at -20°C until use. The DNA was genotyped using Illumina Hap550v3 and Hap610-quad SNP microarrays. Only directly genotyped SNPs were used in this study.
II.2.5 RNA isolation

50-100 mg of left atrial appendage tissue was used to extract RNA. The tissue, in one ml of TRIzol® (Invitrogen), was homogenized with a sterile Omni Tip Disposable Generator Probes. RNA was isolated from the homogenate following the manufacturer’s protocol. The RNA pellet was dried and resuspended in 80 µl of DEPC water and the concentration was measured with the NanoDrop ND-1000 Spectrophotometer and stored at -80°C.

II.2.6 cDNA preparation

1 µg of purified RNA was added to 4 µl of Superscript® Vilo™ mastermix (Invitrogen) and water added to bring the reaction volume to 20 µl. The reaction was run in an ABI thermocycler at 25°C for 10 min, 42°C for 120 min and 85°C for 5 min with a 4°C hold temp. After completion, 5 µl of the newly synthesized cDNA was diluted with 90 µl of nuclease free water and stored at -20°C until further use.

II.2.7 Quantitative reverse transcriptase-polymerase chain reaction (qRT-PCR)

An Eppendorf Epmotion 5070 robotic pipettor was used to prepare the working and

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>PITX2c Forward Primer</td>
<td>5'-GCG GTT CCT CTG GAA AGT GG-3'</td>
</tr>
<tr>
<td>PITX2c Reverse Primer</td>
<td>5'-GCA CAC CAT CTC CGA CAC CT-3'</td>
</tr>
<tr>
<td>Probe*</td>
<td>5'/56-FAM/CCC GGA GGC /ZEN/CGC AGA GAA AGA TAA /3IABkFQ/-3'</td>
</tr>
</tbody>
</table>

* FAM fluorophore with internal ZEN and 3' IOWA BLACK FQ quencher modification
reaction plates. To prepare the master mix for each sample, 12.5 µl of the TaqMan® gene expression master mix (Applied Biosystems) was used along with 1.25 µl of the custom designed PITX2c primer/probe set (Table 2.1, obtained from IDT) or SHOX2 primer/probe set (assay number Hs00243203_m1 from Applied Biosystems) and the primer limited cardiac actin (ACTC1) primer/probe mix (assay number Hs00606316_m1 from Applied Biosystems). PITX2c expression was also normalized to primer limited cyclophilin A (PPIA) primer/probe mix (assay number Hs04194521_s1 from Applied Biosystems). This 15 µl mix was pipetted into individual wells of a 96-well working plate. Using the robot, 10 µl of the diluted cDNA was added. 5 µl of the total mixture from the working plate was pipetted in triplicate to a 384-well assay plate. PCR was performed in a Bio-RAD CRX qRT-PCR machine that had been calibrated for our FAM and VIC fluorescent probes. Thermal cycling was performed with a hot-start at 95°C for 10 minutes, followed by 40 cycles of 95°C for 15 seconds and 60°C for 60 seconds. Delta C(t) values for PITX2c and SHOX2 expression levels were calculated relative to ACTC1 expression, and the ΔΔCT method was used to compare expression among samples\textsuperscript{20}, yielding log\textsubscript{2} based expression values.

II.2.8 PITX2c expression analysis

Relative log\textsubscript{2} gene expression levels were corrected for plate and batch effects using three standardized atrial RNA samples on each plate. Relative expression levels were fit to an additive linear model including age, gender, donor/surgical sample, atrial fibrillation history and pre-operative rhythm status, using the R statistical program. Differences in PITX2c expression among the rhythm groups was determined by non-parametric ANOVA.
II.2.9 *PITX2* eQTL analysis

For the four AF susceptibility SNPs on chromosome 4q25, relative *PITX2c* expression levels were fit to an additive linear model including age, gender, donor/surgical sample, AF history, pre-operative rhythm status, and genotype using the R statistical program. Analysis was performed on all 239 samples. For regional eQTL analysis in the chromosome 4q25 locus, 169 assayed SNPs from the Illumina SNP microarray +/- 500 Kb from the *PITX2* gene were tested for association with *PITX2c* expression levels using R. This analysis was confined to 223 samples, excluding the 16 donors, where all clinical information was known. Relative *PITX2c* expression levels were fit to an additive linear model including age, gender, history of coronary artery disease (CAD), history of mitral valve disease (MVD), history of hypertension, body mass index (BMI), atrial fibrillation history, pre-operative rhythm status, and genotype using the R statistical program. Significance for the regional eQTL p-values was determined by deviation from the expected values using a quantile-quantile (QQ) plot. Additional analysis and plotting were performed with GraphPad Prism software. Power analyses for eQTL studies were performed in R using a linear model test at $f^2$ values of 0.02, 0.15, and 0.35.

II.3 Results

II.3.1 Patient Characteristics of 239 adult left atria tissue
223 left atrial appendages were obtained during cardiac surgery from subjects of European ancestry. 16 additional were obtained from transplant donors that were not used for transplantation. Samples were divided into three groups based on their history of AF and their preoperative rhythm status: no history of AF (No AF, N=40); history of AF in sinus rhythm at time of sample collection (AF/SR, N=78); and history of AF in AF rhythm at time of sample collection (AF/AF, N=121). 24 of the 40 No AF subjects were in surgery to treat other cardiac conditions, while the remaining 16 donor samples were assumed to have no history of AF. There was no significant difference in \( \text{PITX2c} \) expression between the 24 No AF surgical samples and the 16 No AF donor samples before or after correction for sex and age (Figure 2.1). However, there was a trend for lower \( \text{PITX2c} \) expression in the donor samples. Thus, in subsequent analyses we combined these 40 subjects into one No AF group, and corrected expression for donor status. We examined if the AF/rhythm status groups were associated with sex, age, BMI, and history of hypertension, CAD, and MVD; although, for the No AF group we had to exclude the donor samples for association.

Figure 2.1: Adjusted and unadjusted expression of \( \text{PITX2c} \) in human left atrial appendages in AF controls. Log2 \( \text{PITX2c} \) expression, normalized to \( \text{ACTC1} \), in the 16 donor and 24 surgical No AF samples uncorrected (A), or after correction for age and sex (B). There was no significant difference in \( \text{PITX2c} \) expression between these groups by non-parametric Mann-Whitney test. Individual values are shown along with median and interquartile range.
with BMI, hypertension, CAD, and MVD, since this data was not available for these samples (Table 2.2). Females constituted 23% of the cohort with no statistically significant differences in sex among the rhythm groups (p=0.11). Age (range 16-86 years old) was associated with rhythm status with the AF/AF group being the oldest and the AF/SR group being the youngest (p-value = 0.040). The four subjects < 31 years old were all in the No AF group. BMI for our cohort ranged from 17.8 to 46.9, with a trend for the highest BMI in the AF/AF group and the lowest in the No AF group (p-value = 0.064). History of hypertension was present in 51% of the subjects, with a trend towards a higher frequency in the No AF group (p=0.085). History of CAD was present in 34% of the subjects and trended higher in those with No AF (50%, p=0.17). History of MVD was present in 49% of the subjects and trended higher in the No AF group (70%, p=0.076). Among the 199
subjects with a history of AF, 35 had lone AF. These subjects were equally represented in
the AF/SR and AF/AF groups. However, the lone AF subjects were significantly younger
with a median age of 56 (45 - 61 interquartile range) than the other AF subjects with a
median age of 63 (interquartile range 44-70, p< 0.0001 by Mann Whitney two tailed t-test).

II.3.2 Four independent SNPs at chromosome 4q25 associated with AF

Prior GWAS and meta-analyses have shown that the 4q25 locus has the strongest

association with AF. We performed a GWAS using 607 Cleveland Clinic lone AF cases
and 2956 Illumina population controls, which confirmed that the strongest locus
associated with AF resides at chromosome 4q25. Figure 2.2A shows the fine map of this
AF association at 4q25. We performed a conditional analysis to identify independent
SNPs associated with AF in this region. Four SNPs were found to be independently
associated with AF at the locus-wide significance threshold of p< 6.67x10⁻⁵ (Figure 2.2B

Figure 2.2: Identification of four SNPs independently associated with AF at the 4q25 locus in the Cleveland Clinic Lone AF GWAS. A. AF associations of all genotyped SNPs in the 4q25 locus using the marginal model described in Table 3. The locations of the PITX2c and ENPEP genes are shown above. The dashed line is the Bonferroni corrected level of significance for the 150 SNPs tested at p=0.01. Since many of the SNPs in this region are in LD with each other, the AF associations were recalculated after adjustment for the remaining significant SNPs using the full model described in Table 3. The four independently associated SNPs are in the center of the circles in both panels.
and Table 2.3). Two of these SNPs, rs2200733 and rs3853445, had previously been found to be independently associated with AF. The minor alleles of three of these SNPs are associated with increased risk, while the minor allele of rs385445 was associated with decreased risk for AF. The most highly AF-associated SNP, rs2200733, had an odds ratio of 2.47. These four SNPs are in weak linkage disequilibrium with each other (Table 2.4), showing that four separate haplotype blocks in this locus are associated with AF.

<table>
<thead>
<tr>
<th>SNP</th>
<th>Position on Chr 4</th>
<th>MAF</th>
<th>OR</th>
<th>P-Value</th>
<th>LOG10 P-Value</th>
<th>OR</th>
<th>P-Value</th>
<th>LOG10 P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2200733</td>
<td>111929618</td>
<td>0.17</td>
<td>2.47</td>
<td>1.80E-25</td>
<td>24.75</td>
<td>2.46</td>
<td>1.80E-23</td>
<td>22.75</td>
</tr>
<tr>
<td>rs3853445</td>
<td>111980936</td>
<td>0.25</td>
<td>0.59</td>
<td>1.14E-09</td>
<td>8.94</td>
<td>0.59</td>
<td>3.65E-09</td>
<td>8.44</td>
</tr>
<tr>
<td>rs10033464</td>
<td>111940210</td>
<td>0.10</td>
<td>1.32</td>
<td>8.34E-03</td>
<td>2.08</td>
<td>1.99</td>
<td>2.26E-09</td>
<td>8.65</td>
</tr>
<tr>
<td>rs1448818</td>
<td>111789672</td>
<td>0.25</td>
<td>1.42</td>
<td>3.87E-06</td>
<td>5.41</td>
<td>1.44</td>
<td>7.66E-06</td>
<td>5.12</td>
</tr>
</tbody>
</table>

Results from logistic regression fits of the following 2 models. Complete case analysis used for both models.

Marginal Model: HxLAfib ~ Sex + EV1 + EV3 + EV6 + EV8 + SNP

Full Model: HxLAfib ~ Sex + EV1 + EV3 + EV6 + EV8 + rs2200733 + rs3853445 + rs10033464 + rs1448818

II.3.3 Covariates affecting *PITX2c* expression levels
Expression of PITX2c normalized to ACTC1 was measured by qRT-PCR in RNA derived from the left atrial appendages. Log₂ PITX2c expression levels were examined in the different AF history/rhythm groups (Figure 2.3). Surprisingly, there was a U-shaped relationship with increased AF disease status, such that PITX2c expression was higher in subjects with no history of AF, lower in AF/SR subjects, and higher again in AF/AF subjects, with this difference highly significant (p=2x10⁻⁴ by non-parametric Kruskal-Wallis ANOVA) (Figure 2.3A). Dunn’s ANOVA post-test indicated that the only significant difference among the three groups was between the AF/SR and AF/AF groups (p<0.001), although the effect size was moderate with 17% higher median PITX2c expression levels (antilog2 transformed) in the AF/AF vs. AF/SR groups. After adjusting for sex and age, the relationship between PITX2c expression and AF history/rhythm was maintained (p<1x10⁻⁴ overall), with 24% higher median PITX2c expression (antilog2 transformed) in the AF/AF vs. AF/SR groups (p<0.001, Figure 2.3B).
We then looked at each covariate individually for an association with PITX2c expression. Age, history of hypertension, and history of MVD were all significantly associated with PITX2c expression at \( p < 0.05 \); and, history of CAD had a trend with \( p = 0.07 \). The strongest covariate associated with PITX2c expression was rhythm status in those with a history of AF, such that PITX2c expression in the AF/SR group was significantly different from those in the AF/AF group (\( p = 9.14 \times 10^{-4} \), Table 2.5). However in a multivariate model adjusting for all covariates, the only covariates significantly associated with PITX2c expression were age (\( p = 5.79 \times 10^{-3} \), Table 2.5) and rhythm status in those with a history of AF, such that PITX2c expression in the AF/SR group was significantly different from those in the AF/AF group (\( p = 4.59 \times 10^{-3} \), Table 2.5). In the multivariate analysis the \( \log_2 \) effect size and direction on PITX2c expression for each covariate is indicated by the \( \beta \) coefficient.

Table 2.5: Multivariate Model for PITX2c expression

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Unadjusted p-value</th>
<th>Adjusted p-value$^a$</th>
<th>( \beta ) for adjusted p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>0.20</td>
<td>0.22</td>
<td>0.10</td>
</tr>
<tr>
<td>Age</td>
<td>1.13E-03</td>
<td>5.79E-03</td>
<td>0.01</td>
</tr>
<tr>
<td>BMI$^a$</td>
<td>0.18</td>
<td>0.79</td>
<td>1.57E-03</td>
</tr>
<tr>
<td>Hypertension$^a$</td>
<td>0.01</td>
<td>0.24</td>
<td>0.08</td>
</tr>
<tr>
<td>CAD$^a$</td>
<td>0.07</td>
<td>0.82</td>
<td>0.02</td>
</tr>
<tr>
<td>MVD$^a$</td>
<td>0.03</td>
<td>0.16</td>
<td>-0.10</td>
</tr>
<tr>
<td>No history of AF/donors</td>
<td>0.25*</td>
<td>0.19$^b$</td>
<td>-0.15</td>
</tr>
<tr>
<td>Donors only</td>
<td>0.185*</td>
<td>0.94$^b$</td>
<td>-0.013</td>
</tr>
<tr>
<td>AF/SR</td>
<td>9.14E-04*</td>
<td>4.59E-03</td>
<td>-0.21</td>
</tr>
</tbody>
</table>

$^a$, vs. AF/AF group

$^b$, adjusted for all covariates vs. AF/AF group, unless marked otherwise

$^a$, Not including donors, for which this information not available

$^b$, adjusted only for age and sex vs. AF/AF group
II.3.4  \textit{PITX2c} cis-eQTLs

We assessed the four independent AF SNPs on chromosome 4q25 for association with the expression of \textit{PITX2c}, thus probing whether these SNPs serve as cis-eQTLs for \textit{PITX2c}.

None of these SNPs were associated with \textit{PITX2c} levels when normalized to ACTC1, a cardiomyocyte-specific reference gene (p>0.15, Table 2.6) or when normalized to PPIA, a reference gene expressed in all cell types (p>0.25, Table 2.7). After adjusting \textit{PITX2c} expression for sex, age, donor status, and AF history/rhythm there still was no association of these four SNPs with expression of \textit{PITX2c} (Tables 2.6, 2.7). We were well powered to detect SNPs effects on the expression of \textit{PITX2c} in the 239 human left atrial appendages, with 99% power to observe a 15\% log2 fold change in \textit{PITX2c} expression (Table 2.8). In order to determine if we could uncover any cryptic associations in any rhythm subgroup, we evaluated each of the three AF history/rhythm groups to determine if there were any

\begin{table}[h]
\centering
\begin{tabular}{cccc}
\hline
SNP & Location & \textit{PITX2c} p-value uncorrected & \textbf{R of \textit{PITX2c} uncorrected and 95\%CI} & \textit{PITX2c} p-value phenotype corrected* \\
\hline
rs2200733 & 111929618 & 0.90 & 0.0084 [-0.12 - 0.14] & 0.72 \\
rs3853445 & 111980936 & 0.68 & 0.027 [-0.10 - 0.15] & 0.96 \\
rs1448818 & 111789672 & 0.40 & 0.0548 [-0.073 - 0.18] & 0.37 \\
rs10033464 & 111940210 & 0.15 & -0.036 [-0.23 - 0.034] & 0.09 \\
\hline
\end{tabular}
\caption{Chr. 4 AF risk SNPs not associated with \textit{PITX2c} expression normalized to ACTC1}
\end{table}

\begin{table}[h]
\centering
\begin{tabular}{cccc}
\hline
SNP & Location & \textit{PITX2c} p-value uncorrected & \textit{PITX2c} p-value phenotype corrected* \\
\hline
rs2200733 & 111929618 & 0.25 & 0.39 \\
rs3853445 & 111980936 & 0.63 & 0.97 \\
rs1448818 & 111789672 & 0.40 & 0.33 \\
rs10033464 & 111940210 & 0.81 & 0.51 \\
\hline
\end{tabular}
\caption{Chromosome 4.25 AF risk SNPs not associated with \textit{PITX2c} expression normalized to PPIA expression}
\end{table}

*, corrected for sex, age, donor status, and AF history/rhythm

CI, confidence interval
significant associations of these four SNPs with unadjusted or adjusted PITX2c expression.

No significant associations were discovered at p<0.05 in any of the subgroups (Table 2.9).

Table 2.8: Power Analysis for PITX2c expression

<table>
<thead>
<tr>
<th><em>PITX2c</em> log2 effect</th>
<th>Power in No AF cohort (N=40)*</th>
<th>Power in full cohort (N=239)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.02</td>
<td>8%</td>
<td>30%</td>
</tr>
<tr>
<td>0.15</td>
<td>38%</td>
<td>99%</td>
</tr>
<tr>
<td>0.35</td>
<td>78%</td>
<td>99.99%</td>
</tr>
</tbody>
</table>

*, Covariates were age, sex, donor status, and SNP yielding 3 degrees of freedom

#, Covariates were age, sex, donor status, 3 AF history/rhythm at surgery groups, and SNP yielding 6 degrees of freedom.

Among the 40 No AF subjects, we had moderate power to detect SNP effects on PITX2c expression, with 38% and 78% power to observe a 15% and 35% log2 fold change in PITX2c expression, respectively (Table 2.8). Upon analysis of genotyped SNPs over the region +/- 500 kb from PITX2c, there were seven SNPs outside of the expected p-value

Table 2.9: p-values for AF SNPs association with PITX2c expression in individual AF history/rhythm groups

<table>
<thead>
<tr>
<th>SNP</th>
<th>No AF (N=40) Uncorrected</th>
<th>No AF (N=40) Corrected*</th>
<th>AF/SR (N=78) Uncorrected</th>
<th>AF/SR (N=78) Corrected*</th>
<th>AF/AF (N=121) Uncorrected</th>
<th>AF/AF (N=121) Corrected*</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2200733</td>
<td>0.267</td>
<td>0.513</td>
<td>0.189</td>
<td>0.110</td>
<td>0.978</td>
<td>0.886</td>
</tr>
<tr>
<td>rs3853445</td>
<td>0.522</td>
<td>0.451</td>
<td>0.863</td>
<td>0.998</td>
<td>0.374</td>
<td>0.448</td>
</tr>
<tr>
<td>rs1448818</td>
<td>0.832</td>
<td>0.917</td>
<td>0.157</td>
<td>0.119</td>
<td>0.789</td>
<td>0.652</td>
</tr>
<tr>
<td>rs10033464</td>
<td>0.422</td>
<td>0.122</td>
<td>0.237</td>
<td>0.302</td>
<td>0.518</td>
<td>0.592</td>
</tr>
</tbody>
</table>

*, corrected for age, sex, and donor status.

#, corrected for age, sex, BMI, and history of hypertension, CAD, and MVD.
range in QQ plots that were associated with the adjusted $PITX2c$ expression levels (Figure 2.4). These SNPs were associated with $PITX2c$ expression at $p<0.01$, with the top SNP at $p=3.0\times 10^{-4}$ (Figure 2.5, Table 2.9). These SNPs are located within introns of the ENPEP gene, which is proximal on chromosome 4 to the $PITX2$ gene; while the AF associated SNPs are distal to $PITX2$. These seven SNPs are all in LD with each other and thus represent one haplotype block ($r^2$ from 0.292 to 1, Table 2.10).

II.3.5 $PITX2c$ expression is inversely correlated with $SHOX2$ expression
SHOX2 is a transcription factor that plays a role during embryonic development promoting the formation of the sinoatrial node.\textsuperscript{71} In mice, Shox2 is repressed by Pitx2, as demonstrated by higher Shox2 mRNA levels in Pitx2\textsuperscript{-/-} vs. Pitx2\textsuperscript{+/-} mouse hearts.\textsuperscript{50}

Bioinformatic analysis identified a conserved Pitx2c recognition element in the 2nd intron of the Shox2 gene.\textsuperscript{50}

Reported gene transfections confirmed that Pitx2c directly represses gene expression via this recognition element in the Shox2 gene.\textsuperscript{50} We measured SHOX2 expression by qRT-

\begin{table}
\centering
\begin{tabular}{|c|c|c|c|c|c|c|c|c|}
\hline
SNP name & rs11731078 & rs2348427 & rs2881913 & rs639194 & rs16997154 & rs1448808 & rs6533524 & \multicolumn{2}{c|}{PITX2 pvalue} & Chr \# bp & Intron \\
\hline
rs11731078 & 1 & & & & & & & 3.04E-04 & 111645790 & 3rd \\
rs2348427 & 0.571 & 1 & & & & & & 3.07E-04 & 111633848 & 3rd \\
rs2881913 & 0.571 & 1 & 1 & & & & & 3.07E-04 & 111635731 & 3rd \\
rs639194 & 0.915 & 0.505 & 0.505 & 1 & & & & 2.06E-03 & 111669772 & 10th \\
rs16997154 & 0.298 & 0.522 & 0.522 & 0.245 & 1 & & & 2.56E-03 & 111638646 & 3rd \\
rs1448808 & 0.44 & 0.802 & 0.802 & 0.485 & 0.501 & 1 & & 2.67E-03 & 111665036 & 10th \\
rs6533524 & 0.341 & 0.68 & 0.68 & 0.292 & 0.55 & 0.519 & 1 & 4.42E-03 & 111626900 & 1st \\
\hline
\end{tabular}
\caption{Linkage disequilibrium ($r^2$) between significant PITX2c eQTL SNPs located in the ENPEP Gene}
\end{table}

\begin{figure}
\centering
\begin{minipage}[b]{0.45\textwidth}
\centering
\includegraphics[width=\textwidth]{SHOX2_expression.png}
\caption{A. SHOX2 expression was inversely correlated with PITX2c expression in all 239 subjects ($r = -0.20$, p=0.0021). B. Among the three AF status/rhythm groups, the inverse correlation between SHOX2 and PITX2c was only found in the No AF subgroup ($r = -0.47$, p=0.0023).}
\end{minipage}
\begin{minipage}[b]{0.45\textwidth}
\centering
\includegraphics[width=\textwidth]{PITX2c_expression.png}
\caption{B. PITX2c expression was inversely correlated with SHOX2 expression in all 239 subjects ($r = -0.47$, p=0.0023).}
\end{minipage}
\end{figure}
PCR to examine whether there was a correlation between \textit{PITX2c} and \textit{SHOX2} expression in the human left atrial appendages. We found an inverse correlation between \textit{PITX2c} and \textit{SHOX2} expression in these samples ($r=-0.20$, $p=0.0021$, Figure 2.6a).

Additionally we further analyzed the results based on AF history/rhythm status and found an even stronger inverse correlation within the subgroup with no history of AF ($r=-0.47$, $p=0.0023$, Figure 2.6b), while this correlation was not significant in the AF/SR and AF/AF groups.

\textbf{II.4 Discussion}

Analysis of our Cleveland Clinic Lone AF GWAS data revealed four independent AF associated SNPs in the chromosome 4q25 region, the strongest being the previously identified rs2200733 SNP, which had an odds ratio of 2.47 ($p=1.8\times10^{-25}$). This odds ratio is higher than those previously reported of 1.4 to 2,\textsuperscript{30,32} which we attribute to the use of a lone AF cohort. Genetic susceptibility in the lone AF cohort may play a larger role in AF pathogenesis than in a mixed cohort of AF subjects, in which many of the cases may be secondary to structural heart diseases such as CAD and MVD. We then utilized 239 human adult left atrial appendages to determine whether these four independent AF SNPs were associated with expression of \textit{PITX2c}, the gene closest to these SNPs, although rs2200733 is 150.6 kb distal to the \textit{PITX2} gene on chromosome 4. While these SNPs were not associated with expression of \textit{PITX2c} in adult human left atrial appendage tissues, we did find seven SNPs in introns of \textit{ENPEP}, proximal to the \textit{PITX2} gene, that were associated with \textit{PITX2c} expression.
PITX2c expression was associated with AF rhythm status, such that there was significantly increased expression of PITX2c in the AF/AF groups compared to the AF/SR group. This was surprising because we predicted that PITX2c expression might decrease with AF burden, since Pitx2c<sup>−/−</sup> mice are susceptible to pacing-induced atrial arrhythmia;<sup>50</sup> yet, the highest PITX2c expression we observed was in the AF/AF group. A prior study using surgically obtained atrial tissues from only 5 No AF and 5 matched AF subjects found that PITX2c expression was lower in 4/5 of the AF subjects than the No AF controls;<sup>72</sup> however, the current study has a much larger sample size, allowing a multivariate regression model to compare PITX2c expression levels among the various rhythm groups. It is known that AF, especially permanent AF, is associated with structural, contractile, and electrophysiological remodeling.<sup>73,74</sup> It has been suggested that AF induces a cellular adaptation through de-differentiation into a more fetal-like cell phenotype that promotes cell survival during stress.<sup>74,76</sup> Using a mouse model where lacZ was knocked into the PITX2 gene, Wang et al. demonstrated that Pitx2 expression is highest in the left atrium in 3-day old mice, moderate in the left atrium of 42-day old mice, and only expressed in a few residual cells of the left atrium in 1-year old mice.<sup>50</sup> Thus, we speculate that the fetal-like reprogramming associated with increased AF burden might lead to the reactivation of PITX2c expression in previously non-expressing cells and explain why we observed the highest PITX2c levels in the AF/AF group. Overall, we found that PITX2c expression in the subjects with no history of AF was not statistically different than its expression in the subjects with a history of AF; thus, PITX2c expression levels in adult left atrial appendages cannot be used to distinguish AF cases from controls.
We predicted that the AF-associated SNPs at chromosome 4q25 might regulate left atrial expression of PITX2c; however, our results were not consistent with this hypothesis in the adult left atrial appendages. This negative finding may have resulted from examining PITX2c expression at the wrong time and/or wrong place. For example, it might be possible to identify the effect of these SNPs on PITX2c expression in left atrial tissue from neonates, where PITX2c expression may be expressed highly in all left atrial cardiomyocytes. However, we speculate that the effects of these SNPs on expression may be lost in the residual cells that retain PITX2c expression in adult left atria, where epigenetic modifications may mask the SNP effects. Furthermore, Pitx2 expression in the lacZ knock-in mouse was very high in the pulmonary vein region of 3-day old mice; and the pulmonary vein region is the target of therapeutic ablation that often suppresses AF. Thus, it is possible that the AF-associated SNPs regulate PITX2c expression in the pulmonary vein, but not in the left atrial appendage. Another possibility is that the AF-associated SNPs actually control the expression of other adjacent protein coding or noncoding genes that are involved in AF pathogenesis, even though PITX2 is the closest gene.

One approach that might be used to identify functional activity of the regions containing the four AF-associated SNPs would be the identification of enhancer/silencer activity using reporter gene transfection studies, or using transgenic mice or zebrafish. However, a caveat to this strategy is that heart enhancers have been shown to be weakly conserved even between humans and mice, thus limiting the efficacy of animal models that could be used to investigate heart enhancer functional activity.
We identified several SNPs, located in introns one, three, and ten of the ENPEP gene that were associated with expression of PITX2c in human adult left atrial appendages. The ENPEP gene codes for glutamylaminopeptidase or aminopeptidase A (APA). APA is a homodimeric type II membrane-bound protease that converts angiotensin (Ang) II to Ang III by cleaving the N-terminal aspartic acid residue of Ang II. Ang III is a potent agonist of the angiotensin receptor type 1, and there is evidence that it may have a more important role than Ang II in the brain in sustaining hypertension in the spontaneous hypertensive rat. However, the Enpep knockout mouse displays hypertension, arguing that APA also plays a systemic role in the catabolism of angiotensin activity. A human GWAS has shown that a common SNP in ENPEP, rs6825911, is associated with hypertension in East Asians. This SNP is in weak LD with rs16997154 in Asians (r2 = 0.321 in the Asian populations, HapMap release 2.2), which we found to be an eQTL for PITX2c expression in subjects of European ancestry. However, the SNPs in ENPEP that we identified as PITX2c eQTLs in adult left atria were not associated with lone AF in our GWAS (Figure 2) or for AF in prior GWAS. Thus, we found no connection between PITX2c expression, its association with SNPs in the ENPEP gene, and susceptibility to AF.

In our cohort, we confirmed that PITX2c expression was inversely correlated with SHOX2 expression as seen in a mouse study, however when analyzing the data based on rhythm status, only the group with no history of AF remained significant. We found this surprising, because we would have expected to observe a significant inverse correlation in all subgroups. This suggest that tissue remodeling and/or epigenetic changes in disease conditions may override the direct effect of PITX2c on SHOX2 expression, thus rendering
the no AF history group better for analysis of transcription factor associations with their downstream targets.

Our study was limited by the restricted availability of atrial appendages from subjects with no history of AF. The No AF group consisted of 40 subjects, 24 subjects undergoing cardiac surgery for medical reasons plus 16 donor samples that were not used for transplantation. Within this No AF group, we still found no association of any of the four AF associated SNPs with PITX2c expression; however, we had only moderate power to find eQTLs in this subgroup (Table 8). In general, fewer surgical samples are available currently, as catheter ablation has become more common than surgical ablation for AF treatment, and because left atrial appendage resections have become less common with the introduction of left atrial appendage exclusion devices. In conclusion, we found no evidence that the AF risk SNPs at chromosome 4q25 are associated with PITX2c expression in adult left atrial appendages, and we suggest that these risk SNPs may be regulating PITX2c and/or other nearby genes during development or in the pulmonary vein region.
CHAPTER III:

Characterization of a left atrial specific long intergenic noncoding RNA adjacent to PITX2c

III. 1 Introduction

Atrial Fibrillation (AF) is the most common sustained arrhythmia encountered in clinical practice and is associated with a 2-fold increase in mortality and 4- to 5-fold increased risk for stroke worldwide.\textsuperscript{1,66} Prior genome wide association studies (GWAS) identified the strongest single nucleotide polymorphisms (SNPs) associated with AF located on chromosome 4q25, with the AF odds ratio for strongest SNP minor allele ranging from $\sim 1.4$ to 2.0.\textsuperscript{30,32} There are four independent risk SNPs associated with AF in this chromosome 4q25 locus.\textsuperscript{83} These SNPs are located in an intergenic region on chromosome 4q25 from 26 to 217 kb distal to the closest gene on the chromosome, PITX2c. However, we demonstrated that these four AF-risk SNPs are not associated with the expression of PITX2c in a large cohort of human adult left atrial appendages.\textsuperscript{83}

We performed RNAseq of 4 left/right human atrial appendage pairs and identified numerous transcripts with strong expression bias in the left or right atria, including PITX2c, which is expressed only in the left atria.\textsuperscript{65} We also found several novel spliced transcripts that were differentially expressed between the left and right atria.\textsuperscript{65} Here we report the expression of a long intergenic noncoding RNA (lincRNA) adjacent to PITX2c,
which is also expressed specifically in the left vs. right atria. LncRNAs are a subset of long noncoding RNAs (lncRNAs) that are transcribed in between genes and typically have their own transcriptional controlling elements.\textsuperscript{51,53} LncRNAs are characterized as transcripts >200 nucleotides in length that typically have 2-4 exons and are 5' capped and 3' polyadenylated similar to mRNA, yet they do not code for a functional protein, but they may contain short open reading frames.\textsuperscript{51} LncRNAs tend to be weakly conserved between species, with many examples of lncRNAs that are conserved among primates but not between humans and mice.\textsuperscript{51,52} Although many are weakly conserved, lncRNAs/lincRNAs may have important functional roles as they have been implicated in numerous biological processes such as epigenetic regulation, imprinting, cell-cycle control, cellular differentiation, splicing, nuclear and cytoplasmic trafficking, and regulation of transcription and translation.\textsuperscript{54-56,84,85} Additionally, mutations in protein binding partners of lncRNAs/lincRNAs, possibly resulting in defective ribonucleoprotein complexes have been shown in human diseases such as amyotrophic lateral sclerosis (ALS) and fragile X.\textsuperscript{56}

Our central hypothesis is that the AF associated SNPs located on chromosome 4q25 directly affect expression of \textit{PITX2c} and/or the lincRNA adjacent to \textit{PITX2}. We have already shown that in human adult left atrial appendages, the AF risk SNPs located on chromosome 4q25 did not control expression of \textit{PITX2c}.\textsuperscript{83} Before we determined if expression of this newly reported lincRNA, \textit{PANCR}, was associated with SNPs in the 4q25 region, we first performed characterization experiments. We identified its tissue specific expression by qRT-PCR and determined that it was expressed most abundantly in the left atria and the eye. We found that cardiomyocytes differentiated from H9 ES cells expressed
PANCR as well as PITX2c. We performed siRNA knockdown of PANCR, PITX2c and the two genes combined to determine what effect these genes have on gene expression. We found that knock-down of PANCR not only repressed its own expression but also expression of PITX2c. Additionally, using RNAseq, we found that the expression levels of numerous genes were altered after PANCR knock-down. We also found that using PANCR and PITX2 siRNAs simultaneously, the number of genes significantly changed was greater than the each of the two alone. Lastly, and most importantly, we determined that in 223 human left atrial tissue samples, the 4 AF-associated SNPs in the 4q25 region were not associated with PANCR expression. Therefore, the mechanisms of the AF associated SNPs on chromosome 4q25 remain a mystery.

III.2 Materials and Methods

III.2.1 Human left atrial appendages

Left atrial appendage tissues were obtained from a biorepository of human atrial tissues from patients who underwent cardiac surgery at the Cleveland Clinic and who consented to have discarded tissue used for research under a protocol approved by the Cleveland Clinic Institutional Review Board. AF history, type of AF, structural heart disease, demographics, and other clinical data were collected in a research database and a prospectively collected database of all cardiac surgeries (The Cardiovascular Information Registry) maintained by the Department of Cardiothoracic Surgery. Subjects were categorized as “lone AF” if they had a history of AF and did not have coronary artery disease or valvular disease. AF rhythm status was determined by review of electrocardiograms obtained prior to surgery. Samples were snap frozen in liquid nitrogen.
and kept at -80°C until RNA extraction. All samples were approved by the Cleveland Clinic Institutional Review Board.

III.2.2 RNA isolation

Left atrial appendage tissue was used with one ml of TRIzol® (Invitrogen). RNA was isolated from the homogenate following the manufacturer’s protocol. The RNA pellet was dried and resuspended in 80 µl of DEPC water and the concentration was measured with the NanoDrop ND-1000 spectrophotometer and stored at -80°C.

III.2.3 cDNA preparation

1 µg of RNA from the left atrial appendages was added to 4 µl of Superscript® Vilo™ mastermix (Invitrogen) and water added to bring the reaction volume to 20 µl. The reaction was run in an ABI thermocycler at 25°C for 10 min, 42°C for 120 min and 85°C for 5 min with a 4°C hold temp. After completion, 5 µl of the newly synthesized cDNA was diluted with 90 µl of nuclease free water and stored at -20°C until further use.

III.2.4 Quantitative reverse transcriptase-polymerase chain reaction

Table 3.1 LincRNA Taqman primer and probe

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>lincRNA Forward Primer</td>
<td>5' AAT TCT CCA TAG GAC TGC ATG AG-3'</td>
</tr>
<tr>
<td>lincRNA Reverse Primer</td>
<td>5'-CAC CTC GGT TCC ACT CAA C-3'</td>
</tr>
<tr>
<td>Probe</td>
<td>5'/-56-FAM/CGG TCG TCT /ZEN/TCT CCC AGA ATG AG/3IABkFQ/-3'</td>
</tr>
</tbody>
</table>

* FAM fluorophore with internal ZEN and 3' IOWA BLACK FQ quencher modification
12.5 µl of the TaqMan® gene expression master mix (Applied Biosystems), 1.25 µl of the custom designed lincRNA or PITX2c primer/probe set (Tables 2.1 and 3.1, obtained from IDT) and the primer limited cardiac actin (ACTC1) primer/probe mix (assay number Hs00606316_m1 from Applied Biosystems) was mixed to create a master mix to be added to each sample. A similar qRT-PCR assay was done, but normalized to cyclophilin A (PPIA) instead of ACTC1 (assay number Hs04194521_s1 from Applied Biosystems). This 15 µl master mix was pipetted into individual wells of a 96-well working plate. Using the robot, 10 µl of the diluted cDNA was added. 5 µl of the total mixture from the working plate was pipetted in triplicate to a 384-well assay plate. PCR was performed in a Bio-RAD CRX qRT-PCR machine that had been calibrated for our FAM and VIC fluorescent probes. Thermal cycling was performed with a hot-start at 95°C for 10 minutes, followed by 40 cycles of 95°C for 15 seconds and 60°C for 60 seconds. Delta C(t) values for lincRNA and PITX2c expression levels were calculated relative to ACTC1 expression, and the ΔΔCT method was used to compare expression among samples, yielding log2 based expression values.

III.2.5 lincRNA expression analysis

Relative log2 gene expression levels were corrected for plate and batch effects using three standardized atrial RNA samples on each plate. Relative expression levels were fit to an additive linear model including age, gender, donor/surgical sample, atrial fibrillation history and pre-operative rhythm status, using the R statistical program.
III.2.6 *PANCR* and *PITX2c* tissue panel expression assay

A human total RNA master panel was purchased from Clontech (cat 636643). Additionally, ventricular tissue from the Moravec Lab, eye tissue from the Anand-Apte lab, and atrial RNA obtained from the Van Wagoner lab was used. The custom designed lincRNA and *PITX2c* taqman assays were used with cyclophilin (*PPIA*) as the endogenous control. The reactions were run in triplicate and results were measured in terms of $2^{\Delta\Delta C(t)}$.

III.2.7 Overexpression of *PANCR* by transfection into HEK293 cells and RNA immunoprecipitation

The cDNA of *PANCR* was cloned into the vector pcDNA3.1+ (Invitrogen). This new construct was used to transfct *PANCR* into HEK293 cells, a human embryonic kidney cell line that does not endogenously express *PANCR* (Figure 3.1). Using RNA immunoprecipitation (Millipore) following the standard manufactures protocol, we used the following proteins to determine if *PANCR* may play a role in epigenetic gene regulation; SUZ12 and EZH2, core components of polycomb repressor complex (PRC) 2 (described in results) and JARID1c and HP1b proteins known to interact with the PRC2 complex.

III.2.8 Differentiation of H9 cells to cardiomyocytes

![Figure 3.1: Overexpression of lincRNA in HEK293 cells. The lincRNA cDNA construct was successfully transiently transfected into HEK293 cells.](image)
A modified protocol from Q. Zhang et al. was used to differentiate H9 hES cells into cardiomyocytes with the following modifications, instead of using 6-well plates coated with gelatin, we used 12-well plates that were coated with diluted growth factor reduced matrigel (BD). On a dish~90% confluent, 2 x 12-well plates could be made. The cells were plated in MEF media (R&D systems) for 3 days to prime them for differentiation and growth factors were added as previously published. The cells were maintained in RPMI 1640 with B27 (Invitrogen 17504044) until experimentation.

III.2.9 siRNA knockdown of lincRNA and *PITX2*

Table 3.2 Custom sequence for lincRNA siRNA.

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>lincRNA Sense Sequence</td>
<td>5'-CGG UUC CAC UCA ACC GAU U-3'</td>
</tr>
<tr>
<td>lincRNA Anti-sense Sequence</td>
<td>5'-AAU CGG UUG AGU GGA ACC G-3'</td>
</tr>
</tbody>
</table>

100 pmol of a custom lincRNA siRNA (Ambion silencer select table 3.2), 33 pmol of each of 3 *PITX2* siRNA (Ambion silencer select cat#4392420, ids: S10557, S10558, and S10559) and 100 pmol of a control scramble siRNA (Ambion silencer select cat #4390843) were used with RNAiMax (Invitrogen) transfection reagent according to manufactures specifications to knockdown the genes of interest. The siRNA complexes were incubated with cells for 48 hrs, followed by RNA isolation. cDNA was prepared and qRT-PCR was run as described above.

III.2.10 RNAseq and analysis
Library preparation and sequencing was completed at the University of Chicago on Illumina’s HiSeq platform. The 100bp paired-end reads were aligned to hg19 using STAR aligner\textsuperscript{57} and the Ensembl 71 transcript annotation.\textsuperscript{88} Read counts were summed up using htseq-counts. Differential gene expression was determined using the edgeR package in R.\textsuperscript{89} Exon analysis was done using the bioconductor package DEXseq. Geneset enrichment was done in the R-package edge R.

**III .3 Results**

**III.3.1 Discovery of a lincRNA adjacent to the *PITX2* gene in human left atria**

![Figure 3.2 Left atrial specific uncharacterized lincRNA discovered by RNAseq. RNAseq of 4 left/right atrium pairs revealed an uncharacterized lincRNA. This lincRNA is located upstream of the *PITX2* gene.](image)

RNAseq was previously performed in four pairs of human left-right atrial samples.\textsuperscript{65} We detected a left atrial expressed RNA adjacent to *PITX2* that corresponds to Ensembl transcript RP11-380D23.2-002 (ENST00000513690, release 74), which has now been classified as a long intergenic noncoding RNA (Figure 3.2). We will henceforth refer to this transcript PANC (PITX2 adjacent long intergenic noncoding RNA). PANC expression was left atrial specific with an average of 6.9 reads per kilobase per million mapped reads (RPKM) while no reads above background mapped to this transcript in the
right atria (p-value = 0.002, Figure 3.3). The most common transcript isoform was 446 nucleotides encoded by two-exons derived from a gene of ~19 kb (ENSG00000250103). Our analysis confirmed the lincRNA annotation, with the longest open reading frame (ORF) encoding only 35 amino acid residues, typical for other lincRNAs. Other minor splice junctions were observed and confirmed upon sequencing additional left atrial appendages, with a total of 4 exons detected, but the read coverage over these alternate exons was low (Figure 3.2).

This ORF and the entire PANC sequence is well conserved in primates such as chimpanzee and orangutan (100 and 98.9% identity, respectively), but only small fragments are conserved in non-primate mammalian genomes. For example, the mouse shares a small 55bp region of identity (54/55 nucleotides conserved) located in the orthologous region on mouse chromosome 3 between the Enpep and Pitx2 genes. Only the first 19 nucleotides of the ORF were conserved perfectly. However, RNAseq of mouse left

\[\text{Figure 3.3 PANC share left/right expression pattern in human atria. RNAseq of four left/right atrial pairs showed that PANC (ncRNA near PITX2c) shares the same left/right expression pattern that has been shown in PITX2c. ([157 Hsu, J. 2012])}\]
atria did not yield detectable expression overlapping the 55bp region of identity with human PANCNR.

III.3.2 PANCNR and PITX2c expression in human tissues

To determine the tissue distribution of PANCNR expression, qPCR was performed on 33 human tissue samples including the left atria as a positive control. Expression was highest in the left atria, with ~5-fold lower expression in the eye and lower levels detected in fetal heart, total heart, placenta, and small intestine (Figure 3.4). PANCNR was not detected in the right atria, left or right ventricle, or any other tissue examined. We also determined the expression of PITX2c in these same tissue samples. PITX2c was expressed highest in skeletal muscle (~5.6-fold vs. left atria), followed by eye (~1.8-fold vs. left atria) and left atria (Figure 3.5). Although lower than in the left atria, PITX2c expression was also
detected in placenta >> colon, small intestines, fetal and adult total heart, prostate, and adipose tissue.

III.3.3 Identification of eQTLs for PANC2 in human left atria.

Expression of the lincRNA and PITX2c was measured by qPCR in 223 human left atrial appendages obtained during surgery. Table 3.3 shows the clinical characteristics of these subjects. Overall, there was a robust and significant positive correlation between PANC2 and PITX2c expression ($r^2 = 0.158$, $p<0.0001$, Figure 3.6).

Previously, we determined that common SNPs in the 4q25 region were not associated with PITX2c expression in human adult left atrial appendages. In order to determine if common genetic variants in this locus were

![Figure 3.5 Tissue specific expression of PITX2c](http://genome.ucsc.edu)
Table 3.3: Left atrial appendage surgical patient characteristics

<table>
<thead>
<tr>
<th>Patient Characteristics</th>
<th>Total N=223</th>
<th>No History of AF n=24, 11%</th>
<th>History of AF/SR n=78, 35%</th>
<th>History of AF/AF n=121, 54%</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex, Female, %</td>
<td>22%</td>
<td>27%</td>
<td>25%</td>
<td>19%</td>
<td>0.45$</td>
</tr>
<tr>
<td>Age (years)*</td>
<td>62 (55, 69)</td>
<td>68 (62,75)</td>
<td>59 (52, 65)</td>
<td>64 (55, 70)</td>
<td>0.0035$</td>
</tr>
<tr>
<td>BMI (kg/m$^2$)*</td>
<td>27.8 (24.7, 31.6)</td>
<td>26.5 (23.4,28.3)</td>
<td>27.0 (24.0, 31.4)</td>
<td>28.2 (25.2, 32.2)</td>
<td>0.064$</td>
</tr>
<tr>
<td>Hypertension, %</td>
<td>51%</td>
<td>60%</td>
<td>41%</td>
<td>56%</td>
<td>0.085$</td>
</tr>
<tr>
<td>CAD, %</td>
<td>34%</td>
<td>50%</td>
<td>29%</td>
<td>34%</td>
<td>0.31$</td>
</tr>
<tr>
<td>MVD, %</td>
<td>49%</td>
<td>70%</td>
<td>53%</td>
<td>43%</td>
<td>0.076$</td>
</tr>
<tr>
<td>Lone AF, %</td>
<td>16%</td>
<td>0%</td>
<td>17%</td>
<td>18%</td>
<td>0.78**</td>
</tr>
</tbody>
</table>

*, Median (interquartile range)

$^1$, p-value by chisquare test

$^3$, p-value by Kruskal Wallis nonparametric ANOVA

$^{**}$, p-value by chi-square comparing only AF/SR and AF/AF groups

associated with PANCRC expression, we performed a cis-

eQTL analysis by calculating the association of PANCRC expression with SNP genotypes obtained from microarrays (+/- 500 kb) from PANCRC. First we examined the four independent AF associated SNPs on chromosome 4q25 for

Figure 3.6 PANCRC and PITX2c are positively correlated in human adult left atrial appendages. RNA from human adult left atria appendages tissue were measured for PANCRC and PITX2c expression, they were found to be significantly and positively correlated with a p < 0.0001.
association with the expression of the *PANCR*. These four SNPs (Table 2.3) were identified from the Cleveland Clinic lone AF GWAS, and were recently confirmed in a large conditional meta-analysis where the same four SNPs, or proxies in perfect linkage disequilibrium (HapMap 22) were identified. None of these SNPs were associated with *PANCR* expression in unadjusted data or after adjustment for sex, age, body mass index, hypertension, coronary artery disease, mitral valve disease, and AF history/rhythm (Table 3.4). Then we examined 169 genotyped SNPs in this region, and none were associated with the expression of *PANCR* at the Bonferroni corrected p-value threshold of 0.05.

<table>
<thead>
<tr>
<th>SNP</th>
<th>location</th>
<th>PANCR p-value uncorrected</th>
<th>PANCR p-value phenotype corrected*</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2200733</td>
<td>111929618</td>
<td>0.87</td>
<td>0.80</td>
</tr>
<tr>
<td>rs3853445</td>
<td>111980936</td>
<td>0.60</td>
<td>0.68</td>
</tr>
<tr>
<td>rs1448818</td>
<td>111789672</td>
<td>0.26</td>
<td>0.25</td>
</tr>
<tr>
<td>rs10033464</td>
<td>111940210</td>
<td>0.35</td>
<td>0.31</td>
</tr>
</tbody>
</table>

*, corrected for sex, age, BMI, hypertension, CAD, MVD, and AF history/rhythm

Thus, the AF associated SNPs do not appear to regulate *PANCR* expression in adult human left atrial appendages; however, these SNPs may regulate *PANCR* expression at another time in development or another region of the left atria.

III.3.4 *PANCR* binds to SUZ12, a polycomb repressive complex 2 proteins
A number of lnc/lincRNAs have been shown to interact with components of the PRC2 components. The PRC2 complex is responsible for numerous functions in the cell such as methylation of H3k27me2/3, maintenance of specific gene expression pattern, and repression of progenitor programming during differentiation. To determine if PANCR may play a role in epigenetic gene regulation through the PRC2 complex, RNA immunoprecipitation was performed using HEK293 cells transiently transfected with a PANC PR C expression vector. We tested two core components, SUZ12 and EZH2 as well as two proteins known to interact with PRC2, JARID1c and HP1b, and found that PANC only binds to SUZ12 (Figure 3.7).

III.3.5 PANC and PITX2c coordinately expressed during cardiomyocyte differentiation of human H9 embryonic stem cells
H9 ES cells were differentiated to cardiomyocytes and were harvested at different time points to determine when PANC and PITX2c were expressed. Expression levels of PANC, PITX2c and cardiac Troponin T2, all normalized to cyclophilin (PPIA), were measured by qPCR at different times during cardiomyocyte differentiation. PANC and PITX2c were both induced at day 2-3, while cardiac troponin was not induced until day 8 (Figure 3.8) indicating that PANC and PITX2c were coordinately induced during differentiation prior to the expression of cardiomyocyte structural proteins.

III.3.6 PANC and PITX2c knockdowns in H9 derived cardiomyocytes

We demonstrated above that the expression of PANC and PITX2c were positively correlated in human adult left atrial appendages. We hypothesized that this noncoding gene and the adjacent transcription factor coding gene might regulate each other’s expression, and both might also regulate gene expression in trans. Using the siRNAs shown in table 3.2, PANC and PITX2c were knocked down independently and in combination in H9 differentiated cardiomyocytes, and global gene expression was ascertained by RNAseq. We first looked at the expression of these two adjacent genes.
**Figure 3.9** Knock down of PANCR, PITX2 and the combined by siRNA in differentiated cardiomyocytes. By RNAseq, levels of PANCR and PITX2c were measured. The knock down of PITX2 (green) resulted in significant reduction of PITX2c expression (p< 0.01). The knockdown of PANCR (blue) resulted in a significant reduction of PANCR and PITX2c expression (p <0.01). When both genes were knocked down (purple) this was a greater decrease of expression of both genes. (p< 0.001). Significant difference in expression is notated by different numbers.

PITX2c knock-down reduced its own expression by 33% (p<0.01 by ANOVA posttest), but expression of PANCR was not affected significantly. However, PANCR knock-down decreased expression of both itself and PITX2c (34% and 37%, respectively, p<0.001 for both by ANOVA posttest). Knock-down of both PANCR and PITX2c reduced their expression by 42% and 34% respectively (p<0.001 by ANOVA posttest).

**III.3.7** **PANCR and PITX2c KO induce significant global gene expression changes.**

Using RNA-seq, six genes were significantly altered after PANCR knockdown correcting for multiple testing using the Benjamini-Hochberg false discovery method (significant FDR of < 0.05). However, there are 2,030 genes that are altered at a significant p-value of <0.05, of which 1,304 are down-regulated and 726 are up regulated (top 20 genes, Table 3.5). ATP1A2 and ATP1B2 are the top 2 genes and both are differentially down-regulated. ATP1A2 is a Na+/K+ ATPase found to be associated with hemiplegia, which is paralysis of
the trunk and limbs on one side of the body.\textsuperscript{93} ATP1B2, an ATPase, Na+/K+ transporting, β 2 polypeptide has been studied to determine its role in breast and ovarian cancers.\textsuperscript{94,95} PITX2 knockdown showed no significant changes after multiple testing correcting (FDR <0.05); however, there are 761 genes that are significant at the p-value threshold of <0.05, of which 461 are down-regulated and 300 are up-regulated (top 20 genes, Table 3.6). The top two gene differentially regulated by PITX2 knock down are CST1, which was up-regulated and CHRDL2 which was down-regulated. CST1 is a member of the cystatin proteins which are a class of cysteine peptidase inhibitors present in human saliva.\textsuperscript{96} CHRDL2 which stands for chordin-like 2 is a bone morphogenetic protein antagonist.\textsuperscript{97} Knocking down both PANCR and PITX2 resulted in 440 differentially regulated genes at a FDR threshold of 0.05 and 2,811 genes being significantly changed at a p < 0.05, of which 1,615 are down-regulated and 1,196 are up-regulated (top 20 genes, Table 3.7). HK2 and PYGM are the top two genes and both are differentially down-regulated by the double knock down. HK2, which stands for hexokinase 2, is an isozyme of hexokinase important in glucose phosphorylation in
skeletal muscle. PYGM is a myophosphorylase gene implicated in McArdle’s disease, a metabolic myopathy that causes exercise intolerance, muscle cramps, myalgia, and early fatigue. To determine how many genes (significant p< 0.05) overlapped between, PANCR and PITX2c, PANCR and the double knock down, PITX2c and the double knockdown, and all three categories combined we constructed a Venn diagram (Figure 3.10). Of the genes differentially altered by PITX2c knock down, 75% of them overlapped with one or both of the other conditions. In genes differently altered by PANCR knock down, 65% overlapped with one or both of the other conditions. When both PANCR and PITX2c were knocked down simultaneously, 40% of the genes differently altered overlapped with one or both of the individual knock downs. We speculate that the decrease in overlap between the combined knock-down and the individual knock-downs could due to an additive effect of these PANCR and PITX2c knock-downs. If they are acting in the same pathway as we suspect, a knock-down of one gene alone may not cause significant reduction of a particular gene as may be found when there is a knock-down of both genes.

III.4 Discussion

RNAseq of 4 left/right atria pairs revealed an uncharacterized lincRNA, which we have since named PANCR, located upstream of PITX2c and the chromosome 4q25 region. The chromosome 4q25 region has been previously identified as a strong locus for AF. We found that PANCR is expressed in humans strongly in the left atria and in the eye. We performed a similar study with PITX2c and discovered that it too was expressed strongly in the left atria and the eye, but it was most strongly expressed in the skeletal muscle. We were intrigued with the expression pattern similarities of PANCR and PITX2c, but are not
sure how they are connected. It is known that in Axenfeld–Rieger syndrome (RGS), which is due to mutations in \textit{PITX2}, patients are characterized with abnormal development of the anterior eye and of the heart, among other malformations.\textsuperscript{43,100} Additionally in mice that are that are gene deleted for \textit{Pit2}, eye and heart defects have been shown, similar to what is seen in patients with RGS.\textsuperscript{42,48,100} To date, no experiments have been done to determine if \textit{PANCR} may have developmental implications.

To determine what genes/pathways \textit{PANCR} may be regulating, we performed siRNA knockdown of \textit{PANCR}, \textit{PITX2}, and both gene together in cardiomyocytes differentiated from H9 human ES cells. Upon initial analysis, we were surprised to find that when \textit{PANCR} was knocked down, expression of \textit{PITX2c} was significantly reduced, suggesting that \textit{PANCR} may be regulating \textit{PITX2c} in cis. We have also shown that \textit{PANCR} interacts with SUZ12, a core component of PCR2. This is similar to what was found in Klattenhoff et al. in which they showed \textit{Braveheart (Bvht)}, a heart associated lncRNA in mice, bound to SUZ12 at different times during cardiomyocyte differentiation.\textsuperscript{85} They went on to show that in cells that were depleted for Bvht, SUZ12 and its associated modification were enriched at the promoter of cardiac genes; however, they were not able to determine whether Bvht was acting in a direct or indirect matter.\textsuperscript{85} There are several potential roles \textit{PANCR} may play in regards to the PRC2 complex, it can act as a RNA decoy, competitively binding to SUZ12, preventing the PRC2 complex from binding its normal transcriptional target, as in the case of Gas5 during apoptosis.\textsuperscript{56,63} \textit{PANCR} may act as a scaffold molecule by interacting with SUZ12 and the PRC2 complex. ANRIL and HOTAIR have both been shown to interact with the PRC2 complex and induce remodeling of the chromatin
landscape resulting in silencing of specific genes.\textsuperscript{56} We performed RNAseq on the RNA derived from the siRNA knockdown experiments and found that numerous genes were significantly differentially regulated when PANC\textsubscript{R} was reduced. We cannot definitively say whether PANC\textsubscript{R} is acting directly or indirectly on these genes, we can only speculate that if PANC\textsubscript{R} is acting as a scaffold molecule for PRC2, it may be acting in a pathway that under normal circumstances would act as an enhancer or repressor for these genes.

In addition to determining PRC2 complex binding partners, we have showed that during differentiation, PANC\textsubscript{R} and PITX2\textsubscript{c} were coordinately induced prior to cardiac troponin, which is a cardiac muscle-specific gene, indicating that these genes are induced prior to the cells becoming cardiomyocytes. This is in line with the understanding that PITX2\textsubscript{c} and now maybe PANC\textsubscript{R} may be most active during development.

In a paper we previously published, we found that AF SNPs in the 4q25 region were not associated with PITX2\textsubscript{c} expression in human adult left atrial appendages.\textsuperscript{83} We wanted to determine if there was association with expression of PANC\textsubscript{R} in these same tissues. 223 surgically obtained tissues were used and we found that the AF SNPs in the 4q25 region were not associated with expression of PANC\textsubscript{R}. Our suspicions as to why this is the case remains the same, PITX2\textsubscript{c} and now PANC\textsubscript{R} regulation may occur at an earlier time in development and/or in a different location of the heart.
Table 3.5: PANC1R KD top 20 differentially expressed genes

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Chromosome</th>
<th>log$_2$FC</th>
<th>log$_2$CPM</th>
<th>PValue</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATP1A2</td>
<td>1</td>
<td>-1.00</td>
<td>5.32</td>
<td>2.25E-07</td>
<td>4.25E-03</td>
</tr>
<tr>
<td>ATP1B2</td>
<td>17</td>
<td>-0.93</td>
<td>4.90</td>
<td>2.19E-06</td>
<td>1.26E-02</td>
</tr>
<tr>
<td>WNK2</td>
<td>9</td>
<td>-0.71</td>
<td>5.88</td>
<td>2.37E-06</td>
<td>1.26E-02</td>
</tr>
<tr>
<td>HK2</td>
<td>2</td>
<td>-0.70</td>
<td>6.80</td>
<td>2.66E-06</td>
<td>1.26E-02</td>
</tr>
<tr>
<td>NAT8L</td>
<td>4</td>
<td>-1.05</td>
<td>4.87</td>
<td>5.25E-06</td>
<td>1.99E-02</td>
</tr>
<tr>
<td>ADD2</td>
<td>2</td>
<td>-0.79</td>
<td>5.27</td>
<td>1.44E-05</td>
<td>4.55E-02</td>
</tr>
<tr>
<td>ENSG00000248362</td>
<td>5</td>
<td>2.69</td>
<td>-1.07</td>
<td>2.53E-05</td>
<td>6.83E-02</td>
</tr>
<tr>
<td>IGF9</td>
<td>1</td>
<td>-0.98</td>
<td>5.34</td>
<td>3.46E-05</td>
<td>7.32E-02</td>
</tr>
<tr>
<td>MMP9</td>
<td>20</td>
<td>1.26</td>
<td>4.24</td>
<td>3.94E-05</td>
<td>7.32E-02</td>
</tr>
<tr>
<td>COL2A1</td>
<td>12</td>
<td>-0.57</td>
<td>8.72</td>
<td>4.07E-05</td>
<td>7.32E-02</td>
</tr>
<tr>
<td>DPP4</td>
<td>2</td>
<td>-1.74</td>
<td>2.29</td>
<td>4.26E-05</td>
<td>7.32E-02</td>
</tr>
<tr>
<td>TPPP</td>
<td>5</td>
<td>-1.30</td>
<td>2.98</td>
<td>6.80E-05</td>
<td>1.07E-01</td>
</tr>
<tr>
<td>ASTN1</td>
<td>1</td>
<td>-0.99</td>
<td>3.55</td>
<td>9.22E-05</td>
<td>1.34E-01</td>
</tr>
<tr>
<td>PYGM</td>
<td>11</td>
<td>-0.75</td>
<td>4.43</td>
<td>1.19E-04</td>
<td>1.53E-01</td>
</tr>
<tr>
<td>ENSG00000258123</td>
<td>12</td>
<td>2.41</td>
<td>-1.76</td>
<td>1.26E-04</td>
<td>1.53E-01</td>
</tr>
<tr>
<td>CCDC85C</td>
<td>14</td>
<td>-0.55</td>
<td>5.61</td>
<td>1.36E-04</td>
<td>1.53E-01</td>
</tr>
<tr>
<td>ITGB6</td>
<td>2</td>
<td>0.80</td>
<td>7.30</td>
<td>1.38E-04</td>
<td>1.53E-01</td>
</tr>
<tr>
<td>PER1</td>
<td>17</td>
<td>-0.56</td>
<td>4.80</td>
<td>1.94E-04</td>
<td>1.57E-01</td>
</tr>
<tr>
<td>ZNF710</td>
<td>15</td>
<td>-0.51</td>
<td>5.18</td>
<td>1.95E-04</td>
<td>1.57E-01</td>
</tr>
<tr>
<td>CNTN5</td>
<td>11</td>
<td>0.79</td>
<td>5.84</td>
<td>1.95E-04</td>
<td>1.57E-01</td>
</tr>
</tbody>
</table>

Log$_2$FC = log$_2$ of fold change, reduction after knockdown

Log$_2$CM = log$_2$ of counts per million of gene level, negative sign (-) indicates that gene is lowly expressed
Table 3.6: PITX2 KD top 20 differentially expressed genes

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Chromosome</th>
<th>log₂FC</th>
<th>log₂CPM</th>
<th>PValue</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>CST1</td>
<td>20</td>
<td>8.44</td>
<td>-1.41</td>
<td>1.94E-05</td>
<td>3.67E-01</td>
</tr>
<tr>
<td>CHRDL2</td>
<td>11</td>
<td>-3.15</td>
<td>0.79</td>
<td>1.43E-04</td>
<td>9.98E-01</td>
</tr>
<tr>
<td>SLAMF8</td>
<td>1</td>
<td>1.87</td>
<td>-0.80</td>
<td>1.72E-04</td>
<td>9.98E-01</td>
</tr>
<tr>
<td>HS6ST3</td>
<td>13</td>
<td>1.92</td>
<td>-0.14</td>
<td>3.11E-04</td>
<td>9.98E-01</td>
</tr>
<tr>
<td>ENSG00000261286</td>
<td>16</td>
<td>2.35</td>
<td>-1.47</td>
<td>3.73E-04</td>
<td>9.98E-01</td>
</tr>
<tr>
<td>MT2A</td>
<td>16</td>
<td>-1.43</td>
<td>1.27</td>
<td>4.14E-04</td>
<td>9.98E-01</td>
</tr>
<tr>
<td>DRD2</td>
<td>11</td>
<td>-1.75</td>
<td>1.39</td>
<td>4.22E-04</td>
<td>9.98E-01</td>
</tr>
<tr>
<td>ENSG00000248362</td>
<td>5</td>
<td>2.22</td>
<td>-1.07</td>
<td>6.24E-04</td>
<td>9.98E-01</td>
</tr>
<tr>
<td>ENSG00000257576</td>
<td>12</td>
<td>1.63</td>
<td>-1.00</td>
<td>7.09E-04</td>
<td>9.98E-01</td>
</tr>
<tr>
<td>PLA2G2A</td>
<td>1</td>
<td>-1.44</td>
<td>1.93</td>
<td>7.26E-04</td>
<td>9.98E-01</td>
</tr>
<tr>
<td>BMPER</td>
<td>7</td>
<td>-1.24</td>
<td>3.89</td>
<td>7.97E-04</td>
<td>9.98E-01</td>
</tr>
<tr>
<td>SLITRK2</td>
<td>X</td>
<td>-1.00</td>
<td>4.77</td>
<td>8.57E-04</td>
<td>9.98E-01</td>
</tr>
<tr>
<td>BEX5</td>
<td>X</td>
<td>1.97</td>
<td>-0.80</td>
<td>9.52E-04</td>
<td>9.98E-01</td>
</tr>
<tr>
<td>LIN28A</td>
<td>1</td>
<td>1.65</td>
<td>1.73</td>
<td>9.69E-04</td>
<td>9.98E-01</td>
</tr>
<tr>
<td>DPP4</td>
<td>2</td>
<td>-1.39</td>
<td>2.29</td>
<td>9.72E-04</td>
<td>9.98E-01</td>
</tr>
<tr>
<td>CMKLR1</td>
<td>12</td>
<td>-2.26</td>
<td>2.58</td>
<td>1.00E-03</td>
<td>9.98E-01</td>
</tr>
<tr>
<td>RSPO2</td>
<td>8</td>
<td>-1.96</td>
<td>3.48</td>
<td>1.04E-03</td>
<td>9.98E-01</td>
</tr>
<tr>
<td>RAB6C</td>
<td>2</td>
<td>1.83</td>
<td>-1.14</td>
<td>1.04E-03</td>
<td>9.98E-01</td>
</tr>
<tr>
<td>ENSG00000251429</td>
<td>4</td>
<td>1.96</td>
<td>-1.36</td>
<td>1.07E-03</td>
<td>9.98E-01</td>
</tr>
<tr>
<td>COL15A1</td>
<td>9</td>
<td>-1.58</td>
<td>5.91</td>
<td>1.13E-03</td>
<td>9.98E-01</td>
</tr>
</tbody>
</table>

Log₂FC = log₂ of fold change, reduction after knock-down

Log₂CPM = log₂ of counts per million of gene level, negative sign (-) indicates that gene is lowly expressed
Table 3.7: PANC and PITX2 simultaneous KD top 20 differentially expressed genes

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Chromosome</th>
<th>log(_2)FC</th>
<th>log(_2)CPM</th>
<th>PValue</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>HK2</td>
<td>2</td>
<td>-1.18</td>
<td>6.80</td>
<td>7.52E-15</td>
<td>1.42E-10</td>
</tr>
<tr>
<td>PYGM</td>
<td>11</td>
<td>-1.45</td>
<td>4.43</td>
<td>2.25E-13</td>
<td>2.13E-09</td>
</tr>
<tr>
<td>HRC</td>
<td>19</td>
<td>-1.61</td>
<td>5.72</td>
<td>1.58E-12</td>
<td>9.97E-09</td>
</tr>
<tr>
<td>PPFIA4</td>
<td>1</td>
<td>-1.09</td>
<td>6.87</td>
<td>1.48E-10</td>
<td>6.98E-07</td>
</tr>
<tr>
<td>HMHA1</td>
<td>19</td>
<td>-1.15</td>
<td>4.74</td>
<td>4.44E-10</td>
<td>1.45E-06</td>
</tr>
<tr>
<td>AHNAK2</td>
<td>14</td>
<td>-1.36</td>
<td>6.92</td>
<td>4.61E-10</td>
<td>1.45E-06</td>
</tr>
<tr>
<td>ATP1B2</td>
<td>17</td>
<td>-1.20</td>
<td>4.90</td>
<td>1.24E-09</td>
<td>3.35E-06</td>
</tr>
<tr>
<td>ENO2</td>
<td>12</td>
<td>-0.90</td>
<td>7.06</td>
<td>2.61E-09</td>
<td>6.16E-06</td>
</tr>
<tr>
<td>WNK2</td>
<td>9</td>
<td>-0.89</td>
<td>5.88</td>
<td>2.93E-09</td>
<td>6.16E-06</td>
</tr>
<tr>
<td>PLEKHA2</td>
<td>8</td>
<td>-0.81</td>
<td>6.15</td>
<td>3.75E-09</td>
<td>7.10E-06</td>
</tr>
<tr>
<td>RCAN2</td>
<td>6</td>
<td>-1.33</td>
<td>6.21</td>
<td>5.63E-09</td>
<td>9.68E-06</td>
</tr>
<tr>
<td>SLC16A3</td>
<td>17</td>
<td>-0.82</td>
<td>6.13</td>
<td>3.64E-08</td>
<td>5.74E-05</td>
</tr>
<tr>
<td>TRH</td>
<td>3</td>
<td>-1.11</td>
<td>6.52</td>
<td>1.09E-07</td>
<td>1.59E-04</td>
</tr>
<tr>
<td>DDN</td>
<td>12</td>
<td>-2.28</td>
<td>0.37</td>
<td>1.46E-07</td>
<td>1.97E-04</td>
</tr>
<tr>
<td>SPHKAP</td>
<td>2</td>
<td>-1.36</td>
<td>5.65</td>
<td>2.91E-07</td>
<td>3.67E-04</td>
</tr>
<tr>
<td>GJA3</td>
<td>13</td>
<td>-1.12</td>
<td>5.96</td>
<td>4.09E-07</td>
<td>4.83E-04</td>
</tr>
<tr>
<td>ADAM11</td>
<td>17</td>
<td>-1.24</td>
<td>5.18</td>
<td>5.70E-07</td>
<td>6.25E-04</td>
</tr>
<tr>
<td>ADCK3</td>
<td>1</td>
<td>-0.89</td>
<td>4.49</td>
<td>6.54E-07</td>
<td>6.25E-04</td>
</tr>
<tr>
<td>PDK1</td>
<td>2</td>
<td>-0.79</td>
<td>7.63</td>
<td>6.73E-07</td>
<td>6.25E-04</td>
</tr>
<tr>
<td>KCNJ11</td>
<td>11</td>
<td>-1.47</td>
<td>1.82</td>
<td>6.91E-07</td>
<td>6.25E-04</td>
</tr>
</tbody>
</table>

\( \log_{2}FC \) = \log_{2} of fold change, reduction after knock-down

\( \log_{2}CPM \) = \log_{2} of counts per million of gene level, negative sign (-) indicates that gene is lowly expressed
CHAPTER IV:

Unraveling the genetics of Atrial Fibrillation

IV.1 Conclusions and future directions

IV.1.1 AF SNPs in the 4q25 region and there possible effects on *PITX2c* expression

We have begun to find ways to understand if the AF SNPs in the 4q25 loci are associated with expression of *PITX2c*. We have shown that in human adult left atrial appendages, the four AF independent SNPs are not associated with expression of *PITX2c*. However, we are not convinced that this issue is resolved. We suspect that we may be looking at the wrong place in the heart and/or at the wrong time. Preliminary data of 27 tissue derived from trimmings from the left atrium of donor hearts removed during the anastomosis of the heart to the recipient left atrial cuff (table 4.1), confirms that we may in fact be using the wrong place in the heart. One of the four AF SNPs, rs3853445, a SNP considered protective for AF (OR = 0.59), is significantly associated with *PITX2c* expression in the PV-
LA cuff trimmings (Table 4.2). We also found that in subjects homozygous for the minor allele, the expression of PITX2c was significantly higher (p = 0.014) than patients with only one or no copies of the minor alleles (Figure 4.1). This is very promising data because as shown in Wang et al., post-natal mice express PITX2c in the PV-LA cuff trimmings as well as the left atrium, and that mice hemizygous for PITX2 can experience pacing induced atrial arrhythmias. The PV-LA cuff trimmings in our experiments were taken from healthy donor hearts that were being used for heart transplants. We determined that expression of PITX2c in PV-LA cuff trimmings was statistically higher when compared to atrial appendages from AF/SR subjects, but only a trend towards significance was found when compared to

Figure 4.1 Effect size of minor allele on PITX2c expression in PV-LA cuff trimmings. The expression of PITX2c in subjects homozygous for the minor allele (2) of SNP rs3853445 is significantly higher (p = 0.014) than in subjects with only one copy of the minor allele (1) or zero copies of the minor allele (0). Using a linear additive model, each allele increases PITX2c expression by 6.25%. The R² of this model is 0.22 suggesting that 22% of the variance is explained by the model.

Figure 4.2: PITX2c expression in PV-LA cuff trimmings is significantly higher than AF/SR. Unadjusted levels of PITX2c in PV-LA cuff trimmings is significantly higher than the AF/SR group (p < 0.0001), however there is a trend toward significance in the No AF group.
AF/AF (Figure 4.2). Therefore, we can speculate that increased levels of PITX2c may be protective in preventing AF. To confirm these proposed results, additional tissue from the PV-LA cuffing trimming will need to be obtained.

IV.1.2 PANCR and its regulation of gene expression

There is still much to learn about PANCR. Thus far we have found what tissue it is predominately expressed in, that its expression is positively correlated with PITX2c, and that during differentiation of H9 hES cells it is coordinately expressed with PITX2c, which occurs prior to expression of cardiac troponin T2. We have also shown that in human adult left atria appendages and PVTs, the 4 AF SNPs located on chromosome 4q25 are not associated with expression of PANCR (Tables 3.4 and 4.1). In cardiomyocytes, differentiated from H9 ES cells, the siRNA knock-down of PITX2c, PANCR and the genes combined, we found that PANCR not only reduced expression of itself, it also reduced expression of PITX2c suggesting that PANCR may exhibit cis-enhancer activity on PITX2c expression. RNAseq of these siRNA knock-downs also revealed numerous genes, many of which overlapped between the different treatments that were differentially altered. It appears that PANCR and PITX2c may function in the same pathways but additional experiments to directly test this are necessary. With all that we have discovered about PANCR; the location of PANCR in the cell, how it may regulate gene expression, and whether it plays a role in AF remain largely unknown.

To begin understanding how PANCR regulates gene expression, we performed RNA-protein binding assays by RIP assay (as discussed in the chapter 3 results section). This assay would tell us which proteins interact with PANCR. Of the four proteins we tested,
we found that SUZ12, a core component of the PRC2 complex, interacted with PANCER. As previously stated, these results do not definitively indicate if PANCER is directly or indirectly affecting gene expression of other genes, only that it may interact with the complex important in chromatin remodeling and epigenetic regulation.

A major limitation to the RIP assay study was that it was not performed in a cell type that endogenously expressed PANCER. This was done because HEK293 cells have robust transfection efficiency and grow rapidly allowing for a large amount of cells to be obtained in a relatively short amount of time. Additionally we tested only 4 proteins, SUZ12 and EZH2 are both components of the PRC2 complex and JARID1c, an eraser of histone modification, and HP1b, a heterochromatin protein. We plan to perform RNA immunoprecipitations with cardiomyocytes derived from H9 hES cells, which we’ve shown to endogenously express PANCER, to confirm that PANCER does in fact bind to SUZ12 and also to test the three other components of the PRC2 complex, EZH1/2, EED, and RbAp48, and other chromatin proteins, similar to those used by Guttman et al.54

Members of our lab are working on PANCER fused to a MS2 stem loop tracer that will allow us to determine where in the cell PANCER is located. RNA florescent in situ hybridization (FISH) can also be used for this purpose. We suspect that its primary location is in the nucleus based on the RIP experiments. Simultaneously, we will perform nuclear/cytoplasmic extraction and measure the levels of PANCER in cell extract. Both of these experiments will be performed in cardiomyocytes differentiated from H9 hES cells.

IV.1.3 Final thoughts
Overall, we have shown that the 4 independent AF SNPs in the chromosome 4q25 region are not associated with expression of PITX2c and PANCRT, our recently identified lincRNA. We have also provided important insight into understanding the mechanisms of AF and determining how PANCRT may function; however there is still a great deal we don’t understand about the 4q25 region; we don’t know if these genes are important in AF susceptibility or what genes the AF SNPs in this region affect. To begin to characterize the 4q25 susceptibility region, numerous techniques can be used. ChIP sequencing along with DNase I hypersensitivity sequencing will allow us to map protein-interactions and determine if there are possible transcription factors, enhancers, repressors, promoters and silencers in this region.\textsuperscript{77,101,102} We can also determine chromatin conformation of the cell and locate active and inactive chromatin region by utilizing Hi-C/4C technology.\textsuperscript{103}

Optimally, to answer questions that we have proposed, we would like to obtain early human embryological/fetal left atrial tissue, however due to ethical and/or availability issues that may not be feasible. Additionally, using a mouse model is not possible because as indicated mice do not express the full length PANCRT gene. Therefore, we will need to continue the use of human stem cells or induced pluripotent stem cells differentiated into cardiomyocytes for any future developmental and functional studies. Lastly we have to consider the fact that PITX2c and/or PANCRT may not be the causative genes in AF, and that the AF SNPs in the 4q25 region act on genes not near the 4q25 region.
V APPENDIX

To whom it may concern,

It is the policy of the American Society for Biochemistry and Molecular Biology to allow reuse of any material published in its journals (the Journal of Biological Chemistry, Molecular & Cellular Proteomics and the Journal of Lipid Research) in a thesis or dissertation at no cost and with no explicit permission needed. Please see our copyright permissions page on the journal site for more information.

Best wishes,

Sarah Crespi

American Society for Biochemistry and Molecular Biology
11200 Rockville Pike, Rockville, MD
Suite 302
240-283-6616
JBC | MCP | JLR
License Details

Thank you very much for your order.

This is a License Agreement between Sharonne R Panter ("You") and Elsevier ("Elsevier"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

Get the printable license.

<table>
<thead>
<tr>
<th>License Number</th>
<th>3318860161254</th>
</tr>
</thead>
<tbody>
<tr>
<td>License date</td>
<td>Jan 30, 2014</td>
</tr>
<tr>
<td>Licensed content publisher</td>
<td>Elsevier</td>
</tr>
<tr>
<td>Licensed content publication</td>
<td>Trends in Cell Biology</td>
</tr>
<tr>
<td>Licensed content title</td>
<td>Long noncoding RNAs and human disease</td>
</tr>
<tr>
<td>Licensed content author</td>
<td>Offy Wapinski, Howard Y. Chang</td>
</tr>
<tr>
<td>Licensed content date</td>
<td>June 2011</td>
</tr>
<tr>
<td>Licensed content volume number</td>
<td>21</td>
</tr>
<tr>
<td>Licensed content issue number</td>
<td>6</td>
</tr>
<tr>
<td>Number of pages</td>
<td>8</td>
</tr>
<tr>
<td>Type of Use</td>
<td>reuse in a thesis/dissertation</td>
</tr>
<tr>
<td>Portion</td>
<td>figures/tables/illustrations</td>
</tr>
<tr>
<td>Number of figures/tables/illustrations</td>
<td>1</td>
</tr>
<tr>
<td>Format</td>
<td>both print and electronic</td>
</tr>
<tr>
<td>Are you the author of this Elsevier article?</td>
<td>No</td>
</tr>
<tr>
<td>Will you be translating?</td>
<td>No</td>
</tr>
<tr>
<td>Title of your thesis/dissertation</td>
<td>GENETIC AND FUNCTIONAL STUDIES OF LOCI ASSOCIATED WITH ATRIAL FIBRILLATION</td>
</tr>
<tr>
<td>Expected completion date</td>
<td>Jan 2014</td>
</tr>
<tr>
<td>Estimated size (number of pages)</td>
<td>80</td>
</tr>
<tr>
<td>Elsevier VAT number</td>
<td>GB 494 6272 12</td>
</tr>
<tr>
<td>Permissions price</td>
<td>$0.00 USD</td>
</tr>
<tr>
<td>VAT/Local Sales Tax</td>
<td>$0.00 USD / $0.00 GBP</td>
</tr>
<tr>
<td>Total</td>
<td>$0.00 USD</td>
</tr>
</tbody>
</table>

Copyright © 2014 Copyright Clearance Center, Inc. All Rights Reserved. Privacy statement. Comments? We would like to hear from you. E-mail us at customerinfo@copyright.com


8. Olshansky BM, Chung Mina K. MD, Pogwizd SMM, Goldschlager NM. ARRHYTHMIA ESSENTIALS. first ed. Sudbury, Massachusetts: Jones and Bartlett Learning, LLC; 2011.


