THE UNIVERSITY OF CHICAGO

LEVERAGING TRANSCRIPTION FACTOR DEPENDENT RNA PROFILING TO INTERROGATE GENE REGULATORY NETWORKS IN HUMAN DISEASE MODELS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES
AND THE PRITZKER SCHOOL OF MEDICINE
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

COMMITTEE ON GENETICS, GENOMICS, AND SYSTEMS BIOLOGY

BY

LINSIN ANN SMITH

CHICAGO, ILLINOIS

AUGUST 2022

TABLE OF CONTENTS

LIST OF FIGURES

Acknowledgements

This dissertation is the product of work involving many members of the Moskowitz lab, as well as collaborators both internal and external to the University of Chicago. First, I'd like to thank my advisor, Dr. Ivan Moskowitz, for all of his guidance and support throughout my graduate school career. He provided immeasurable mentorship on the scientific investigations presented in this dissertation, as well as on many other projects that are not included in this document. I am grateful for the experience to work so closely with Dr. Moskowitz and to have his mentorship throughout my time at the University of Chicago. I would also like to thank my thesis committee including Dr. Marcelo Nobrega, Dr. Alexander Ruthenburg, and Dr. Yang Li. I am extremely grateful for their advice, support, and mentorship.

I am incredibly thankful for the opportunity to work with past and current members of the Moskowitz lab, including many who have contributed directly to this work. I would like to thank Dr. Rajiv Nadadur, Dr. Chul Kim, Dr. Joshua Theisen, Dr. Megan Rowton, Dr. Ozanna Burnicka-Turek, Dr. Alexander Guzzetta, Dr. Holly Yang, Dr. Kohta Ikegami, Carlos Perez-Cervantes, Zhezhen Wang, Zoheb Khan, Jessica Jacobs-Li, Madeline Masengarb, Matthew Stocker, and Sonja Lazarevic for support, reagents, and general help. I would like to give a special acknowledgement to Sonja Lazarevic for her support throughout my time in the Moskowitz lab. She has provided profound insights to my work, assisted in every possible way, and has been a fantastic friend. It has been a great privilege to work with Sonja and to see our scientific investigations grow together over the past six years.

A significant portion of this work, as well as work not presented in this dissertation, has resulted from collaborations with scientists outside of the Moskowitz Lab. I would like to acknowledge Dr. Constance Cepko and Dr. Nicolas Lonfat, without whom the first chapter of

this dissertation would not exist. I would also like to acknowledge members of Dr. Alan Attie's group, namely Dr. Alan Attie, Dr. Shane Simonett, and Dr. Mark Keller for our fruitful collaboration resulting in the work presented in the second chapter of this dissertation. Finally, I would like to thank Dr. Sebastian Pott and Dr. Xin He for the opportunity to work together, resulting in work not presented in this dissertation.

I would like to thank the administrative support provided to me in the Genetics, Genomics, and Systems Biology program; namely by Susan Levison who has been a great support throughout my graduate career. I would also like to thank the administrators of the Cardiovascular Sciences Training Grant and the Genetics and Regulation Training Grant which provided funding for much of my time at the University of Chicago.

Finally, I would like to thank my cohort, my friends, and my family for their constant support throughout my graduate career. While we all have focused on very different elements of genetics in our dissertation research, it has been a pleasure to experience graduate school alongside my cohort, and I look forward to seeing how we all continue to develop scientifically in the future. To my friends, thank you for being a part of this journey – your support over the years means so much to me. To my family, thank you all for your support over the years; even though I am one of the few scientists in the family, you always found a way to support me and encourage me to succeed.

Abstract:

Gene regulation is a tightly controlled process that when disrupted can cause a multitude of cellular defects, increased disease risk, and in some cases disease itself. In order to understand the molecular basis of disease, a deep understanding of the gene regulatory networks that underlie healthy and diseased states is necessary. One approach to understand gene regulatory networks in cell-type and context-dependent states is through the application of transcriptional profiling in the presence or absence of key transcription factors (TFs). While TF dependent transcriptional profiling of coding genes has been commonly adopted to quantitatively relate genes in transcriptional hierarchies; identifying non-coding regulatory elements that drive context-specific gene regulatory programs is a relatively new focus. This paradigm change is driven in part by the finding that the majority of genetic variants associated to disease through genome wide association studies reside in the non-coding portion of the genome and are predicted to affect gene regulation. This dissertation outlines the application of TF dependent coding and non-coding transcriptional profiling to three independent contexts. In the first chapter, we identify cell-type specific gene regulatory programs in the developing retina and nominate regulatory elements that expand our knowledge of retinal biology. In the second chapter, TF-dependent transcriptional profiling reveals key players for inducing proliferation and insulin secretion in human beta islet cells, and defines a proliferative gene regulatory network. And in the third, TF dependent transcriptional profiling disentangles the functional relationships between atrial fibrillation, inflammation, and fibrosis at the level of gene expression in models of disease and rescue. While each context is unique, the three chapters of this dissertation revolve around the central theme of altered gene regulatory networks and leveraging transcriptional profiling to unveil deep insights on the mechanisms of disease and development.

Chapter 1: Introduction:

The question of how one's genetic code results in the profoundly diverse spectrum of phenotypes observed in the human population has been a central question in biology for decades. Understanding the link between genotype and phenotype is critical to our understanding of biological processes, ranging from development to disease. The central dogma of biology posits that DNA is transcribed to generate RNA molecules, which are then translated into functional protein products. The process of going from DNA to RNA is gene expression, which is controlled dynamically, in cell-type and context-dependent patterns. Appropriate regulation of gene expression is necessary to produce the correct amount of protein products at the correct times and in the correct cell types; therefore gene regulation is an extremely critical component of cellular processes.

Gene regulation is an intricate and complex process, requiring the coordinated effort of many cellular components to create functional gene products. Gene expression is controlled through many interconnected regulatory levels; including the basal transcriptional machinery, chromatin accessibility, histone modifications, transcription factors, and non-coding regulatory elements. The basic requirements for gene expression start with the promoter region of a gene, which is bound by transcription factors and activating co-factors to recruit RNA polymerase. In order for a gene to be expressed, it must have accessible chromatin, and nucleosome remodeling may occur in order for the transcription start site (TSS) of the gene to be available to the basal transcriptional machinery. In addition to direct gene regulation at a gene's promoter, transcriptional regulation is also driven by non-coding regulatory elements, such as enhancers, which loop to target gene promoters to induce gene expression.

Enhancer elements reside in the non-coding portion of the genome: that is the stretches of DNA that do not code for genes encoding protein products. Out of the nearly three billion base pairs of DNA in the human genome, less than 3% encodes protein coding genes, leaving the remainder as non-coding DNA [1]. These non-coding regions, once thought to be largely inconsequential, have been the subject of deep investigation. It has been shown that what was previously thought to be "junk" DNA, actually is comprised of critical regulatory elements that play a profound role in establishing appropriate gene expression programs. While there are only ~20,000 protein coding genes in the human genome, there are approximately 1 million putative non-coding enhancer elements, in addition to thousands of documented non-coding RNAs which can also play a role in gene regulation [1]. Enhancer elements play a large role not only in driving gene expression, but specifically in driving expression in cell-, time-, and context-dependent ways. Enhancer elements are bound by transcription factors, which are thought to provide cell-type and context specificity to enhancer activity, alongside chromatin conformation.

Genes are regulated by transcription factors: proteins that contain DNA-binding elements, which can bind to promoters and enhancers to direct gene expression. Transcription factors (TFs) are expressed in a cell-type dependent manner to control the transcription of genes and drive the gene expression program of the cell, specifically regulating genes which are necessary for that cell's fate or function. TFs can provide activating or repressing signals to genes, and are often dynamically expressed themselves in cell-type and time-dependent ways. Out of the ~1,200 identified TFs, a select group are considered "master regulators" and are largely responsible for establishing the gene regulatory program within a specific cell type [2]. Within specific cell types, TFs often operate in groups, with suites of key TFs controlling the

pattern of gene expression in the cell, and this regulatory system is often coined a gene regulatory network (GRN).

Understanding gene regulatory networks is at the foundation of understanding developmental and disease processes. Disease can arise from aberrant gene expression as well as from DNA mutations. While many diseases are caused by single gene mutations (monogenic diseases) the vast majority are caused by subtle additive changes in genes or in their regulation, which confer increased disease risk. This has been shown through the findings of genome wide association studies (GWAS) which link single nucleotide polymorphisms to disease traits of interest. Over 80% of GWAS hits reside in the non-coding regions of the genome, as opposed to within genes themselves [3]. These non-coding variants are hypothesized to affect gene regulation, where small changes in the regulation of genes confer additive increases in disease risk through the disruption of gene regulatory networks. One example is in cardiovascular disease, where it has been well documented that loss of function mutations in key TFs can cause a multitude of cardiovascular deficiencies [2], [4]–[7]. In addition to these loss of function mutations, GWAS highlight the role of how smaller changes to the regulatory network can also result in transcriptional dysregulation that leads to disease. This observation dictates that the path to understanding molecular mechanisms of disease requires investigating gene regulatory function, and the determination of the cis-regulatory elements (CREs) that compose gene regulatory networks.

With the understanding that gene regulatory networks, TF activity, and non-coding genetic variation can have a profound impact on the cellular processes that lead to disease, much work has been done to develop novel approaches to study gene regulation. In order to understand the gene regulatory programs active inside a cell, we have utilized transcriptional profiling

experiments such as RNA sequencing (RNAseq). RNA sequencing harnesses next generation sequencing technology to assay all of the RNA molecules inside of a cell. Further, one can perform these experiments in cell-type and context-dependent ways in order to resolve cell-type and context-dependent differences in gene expression. For example, one can utilize disease models and compare transcriptional output in diseased cells against healthy cells in order to identify disease-specific changes in gene expression. This approach can also be used to investigate the gene regulatory networks mediated by master TFs, which outlines the regulatory relationships between TFs and genes. By performing TF knockout or overexpression experiments followed by RNA profiling, one can identify differentially expressed genes and understand how their expression is mediated by that specific TF.

In addition to using RNAseq to interrogate the regulatory relationships between coding genes, recently RNAseq has also been applied to understand the non-coding genome. Even though the non-coding genome does not encode functional protein product, it has been discovered that up to 80% of the non-coding genome is transcribed into RNA [8]. These non-coding RNAs (ncRNAs) can play many functional roles within a cell, including structural and regulatory roles [9]–[11]. In particular, it has been documented that active enhancer elements are transcribed and produce ncRNAs [9], [12]–[17]. These ncRNAs have been shown to play functional roles in gene regulation, and some may be required for enhancer activity, through participation in chromatin looping events or in the recruitment of activating factors to the enhancer/promoter complex [10], [13], [18], [19]. Transcriptional profiling of non-coding RNA in context-dependent states allows for the identification of differentially active enhancer elements through the quantification of the ncRNAs they generate [13], [16], [20]. TF-dependent ncRNA profiling identifies enhancer elements that are regulated by a specific transcription

factor, allowing for the construction of TF-dependent gene regulatory networks incorporating both coding and non-coding regulatory elements.

In addition to RNA profiling experiments, in order to understand gene regulatory networks in healthy and diseased states, one must also interrogate the genomic landscape. In order for a gene to be transcribed, it must exhibit a chromatin structure that allows for the binding of the basal transcription machinery and transcription factors. In addition to genes themselves, non-coding regulatory elements such as enhancers also must be accessible in order to be bound by TFs and proteins that enable looping from enhancers to promoter regions. Chromatin accessibility is dynamically regulated, with repressed genes residing in closed chromatin regions and expressed genes residing in regions of accessible chromatin. Chromatin dynamics are regulated through chromatin remodeling proteins which can remodel nucleosomes, as well as deposit activating or repressing histone modifications, such as H3K27ac and H3K27me3. In order to investigate the cell-type and context-dependent nature of chromatin accessibility, methods such as DNase hypersensitivity and ATAC sequencing (ATACseq) are instrumental in identifying regions of accessible chromatin genome-wide; and ChIP sequencing (ChIPseq) is a method imperative in identifying the presence of activating or repressing histone modifications which impact chromatin conformation. In addition to measuring the presence or absence of regulatory histone modifications, ChIPseq can also be applied to map the locations of TF binding genome wide, which is of high significance for understanding gene regulation. Utilizing these approaches in a cell-type and TF-dependent manner allows for the elucidation of TF-dependent chromatin accessibility, which in turn improves our understanding of TF-dependent gene regulation.

With advances in technology, we are able to utilize genomic profiling experiments to construct the gene regulatory networks that underlie healthy and diseased states. This dissertation will apply transcriptional profiling to three distinct contexts: the developing mouse retina to identify cell-type specific gene regulatory networks, human beta islet cells to identify gene regulatory networks underlying beta cell proliferation, and the mouse atrium to identify gene regulatory changes in atrial fibrillation and rescue. While each context is markedly different from the others, all three chapters of this thesis revolve around the central themes of altered gene regulatory networks, understanding the genetic basis of development and disease, and leveraging transcriptional profiling to unveil deep insights on the molecular mechanisms of disease.

Chapter 2: Enhancer transcription identifies cis-regulatory elements for photoreceptor cell types

2.1 Introduction:

Identification of tissue and context specific *cis*-regulatory elements (CREs) is critical for defining the transcriptional networks that govern physiology and disease across biological contexts. While high-throughput sequencing methods to monitor histone modifications, chromatin status, and transcription factor (TF) occupancy have led to the identification of many CREs, it has been shown that many putative regulatory elements do not appear to affect gene expression, making them unlikely to represent functional CREs [21]–[26]. The development of massively parallel reporter assays (MPRAs) has allowed for high-throughput analysis of potential CREs for enhancer activity [27]–[29]. However, most of these approaches depend on *a priori* knowledge of candidate CREs and often test CREs in non-native chromatin configurations. To accurately interrogate context-dependent CRE function, complementary strategies for the identification of functional enhancers are necessary. One such approach the interrogation of differentially expressed non-coding RNAs (ncRNAs) to identify active CREs in context- specific states [13], [20], [30].

Non-coding RNAs are often transcribed from active regulatory elements and serve as a hallmark of functional enhancers [11], [13], [14], [19], [31], [32]. While ncRNAs are specifically produced from active regulatory elements, the genome is pervasively transcribed, leading to an abundance of non-coding transcripts, making identification of CRE-specific transcription potentially difficult. For example, several classes of functional ncRNAs have been described in the retina, including microRNAs (miRNAs) and long ncRNAs (lncRNAs) [33]–[35]. To isolate CRE-based ncRNA transcripts, transcription factor (TF)-dependent ncRNA profiling has been

utilized in other systems to identify highly active regulatory elements at loci critical for cardiac rhythm[13]. We hypothesized that TF-dependent ncRNA profiling may afford the identification of TF- dependent retinal CREs.

Over the past two decades, researchers have outlined the GRNs that regulate photoreceptor development [36]. Retinal progenitor cells (RPCs) that express *Otx2* develop into both rod and cone photoreceptors, which are subsequently distinguished by the expression of additional TFs [37], [38]. Cones are specifically produced from RPCs that express *Olig2* and *Oc1*, in addition to *Otx2* [39]–[42]. Rods are specified by the expression of *Nrl*, a basic leucine zipper TF. *Nrl* is required for the formation of rods as it activates a rod specific GRN and represses a cone-specific GRN [43]–[46]. The pan-retinal photoreceptor TF, *Crx*, is also required for the differentiation and maintenance of both rods and cones.

Defining cone-specific CREs has been challenging due to the small number of cones in the mouse retina, where rods comprise the vast majority of photoreceptors [47]. However, in *Nrl*$^{-/-}$ mutant mice, cells normally fated to become rods are transformed into cone-like cells [44]. The abundance of cones in the *Nrl*$^{-/-}$ mouse has been used to investigate cone-specific gene regulation. For example, previous work has profiled rod- and cone-enriched coding and non-coding transcripts [35], [48], [49], and distinguished the regulatory landscape in the rod-predominant wild-type retina from the cone-predominant *Nrl*$^{-/-}$ mutant retina, by ATAC-seq and CRE-seq [48], [50], [51].

Models for combinatorial TF activity in rod and cone GRNs have been generated by examining TF binding in wild-type (rod-dominant) and *Nrl*$^{-/-}$ (cone-dominant) retinas. CRX plays a key role in driving photoreceptor GRNs by promoting chromatin remodeling at specific target sites and recruiting additional TFs to direct photoreceptor differentiation and gene

expression [48], [50], [52]–[58]. The effect of CRX binding changes dramatically when co-bound with NRL. In the presence of NRL, CRX and NRL act cooperatively to robustly induce gene expression, whereas in the absence of NRL, CRX appears to drive gene expression to a lesser extent [29], [48], [49], [51], [56], [59]–[62]. These findings are based on biochemical studies that interrogate physical interactions between NRL and CRX, their genomic localization, the impact of *Nrl* and *Crx* mutant alleles on retinal gene expression, and examination of the activity of NRL and CRX on photoreceptor CREs [29], [48], [49], [59], [62], [63]. These data suggest a heterotypic model for photoreceptor-specific GRN selection in which combinatorial binding of NRL and CRX drive expression of a rod-specific GRN and CRX binding alone drives the expression of a cone-specific GRN.

Here, we applied ncRNA profiling to identify rod- and cone-specific regulatory elements active in the wild-type and $Nrl^{-/-}$ mutant retina, respectively. We identified thousands of *Nrl*-dependent ncRNAs, an order of magnitude more than *Nrl*-dependent ncRNAs identified in previous analyses[35]. Differential ncRNA expression strongly correlated with local coding gene expression, consistent with the observation that altered ncRNA abundance identifies active CREs [13]. Comparing *Nrl*-dependent ncRNAs with retinal genomic signatures established these loci as rod- and cone-specific ncRNA-defined CREs. We examined the effect of specific TF localization patterns on ncRNA abundance genome-wide and found that ncRNA abundance provides quantitative support for heterotypic interactions between NRL and CRX as drivers of cell-type specific CRE activity. Direct assessment of *Nrl*-repressed ncRNA-defined candidate CREs showed that they are active regulatory elements in developing cones. This work provides a resource for the identification of retinal gene regulatory networks and highlights ncRNA profiling as a robust method for the detection of cell-type specific CREs.

2.2 Methods:

Animals

Nrl$^{-/-}$ mice were generated as previously described [48], [64]. Mouse husbandry and all procedures (including euthanasia by CO2 inhalation and cervical dislocation) were conducted in accordance with the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health, and were approved by the Washington University in St. Louis Institutional Animal Care and Use Committee. For ex vivo enhancer testing, wild-type embryos were obtained from timed pregnant CD1 mice (Charles River Laboratories). All animal studies were approved by the Institutional Animal Care and Use Committee at Harvard University.

Coding RNA-Seq library preparation and Data analysis

Libraries were prepared from RNA starting with 1 ug per sample and using the mRNA seq Sample Prep Kit (Illumina) as per recommended instructions. After Ribozero purification and removing only ribosomal RNA, barcoded libraries were prepared according to Illumina's instructions (2013) accompanying the TruSeq RNA Sample prep kit v2 (Part# RS 122 2001). Libraries were quantified using the Agilent Bio analyzer (model 2100) and pooled in equimolar amounts. The pooled libraries were sequenced with stranded 50-bp single-end reads on the HiSeq2500 in Rapid Run Mode following the manufacturer's protocols (2013). RNA library preparation was performed as previously discussed [13]. Briefly, 22M to 30M reads were mapped to mouse genome with STAR (v 2.5.3a). Reads mapped to the mitochondrial genome, and with phred score < 30 were excluded. Counts were retrieved with HTseq (v.0.6.0) [65] in union mode. Lastly, counts were analyzed for differential expression with R (3.4) package DEseq2 [66].

Noncoding RNA-Seq library preparation

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

Noncoding RNA-Seq Data analysis

About 170–186 million high-quality reads (quality score>30) for each sample were obtained. Fastq files were aligned to UCSC genome build mm10 using STAR and between 168 million and 174 million reads were successfully mapped. Transcript assembly was performed by Stringtie (version 1.3.3) and merged with Stringtie “merge” function. Counts were retrieved from the merged transcriptome for differential expression testing, performed as above. False discovery rate (FDR) was calculated after removing the coding-gene transcripts, ncRNA that overlapped coding genes in the same strand, and RNAs shorter than 100 base pairs. Significance was considered to have been reached when FDR was <0.05 and fold change was > 2. Deeptools multiBamCompare and plotHeatmap were used to make and plot sample correlations.

GIGGLE index and search

We downloaded retina methylation, ChIP-seq and ATAC-seq regions from GSE72550 [51], DNAse GSE37074 [22]. TF ChIP-seq (CRX: GSE20012 [60], NRL: https://datashare.nei.nih.gov/nnrlMain.jsp[61], OTX2: GSE54084 [67]) and ATAC-seq

(GSE83312 [48]) were processed as below. Regions were placed in a directory, sorted and gunzipped with 'giggle sort_bed', and then indexed with 'giggle index' which uses tabix (samtools 1.5) [68]. This was deemed the 'retinal index'. NcRNA regions were also GIGGLE processed (sort index), and used to compare against the retinal index with command 'giggle search -I <index> -q <ncRNA.bed> -g 272552137 -s .' Results were ranked by score which is calculated as odds ratio multiplied by -log10(P-value). Heatmap was generated with R pheatmap (v 1.012) CRAN package [69].

Nearest gene assignment

Mouse ESC Topologically associated domains (TAD) were downloaded from ENCODE (chromosome.sdsc.edu, GSE29184 [70]). Empty regions were filled in with R and assigned as "inter-TAD" as described by Szabo et al 2018 [71]. The 5' regions of all RNAs were intersected with TADs using Bedtools intersect.

GO enrichment analysis

Enrichment of GO Biological Process terms from closest genes that share a TAD with ncRNAs was performed with Bioconductor package ClusterProfiler version 2.4670 [72].

ATAC-seq and ChIP-seq data processing

Fastq files from previously generated ChIP-seq (GSE20012 [60]) and ATAC-seq (GSE83312 [48]) datasets were downloaded from GEO and processed identically as previously described [48]. Briefly, adapter sequences were clipped from reads using Cutadapt [73], then aligned to UCSC mouse genome mm10 with bowtie version 2.3.4 [74] in end- to-end mode. Mismatched reads, PCR duplicates, ENCODE blacklisted regions, and reads with quality < 30 were removed with Samtools version 1.5 [75]. For ATAC-seq, fragments with width > 147 base pairs were

removed to enrich for nucleosome free reads using a custom script. Peaks for both assays were called with Macs 2.11 [76].

Associating ncRNAs and regulatory elements

Open chromatin peaks and TF-binding peaks were intersected with ncRNAs using Bioconductor package GenomicRanges [77] allowing for a 500bp gap upstream the 5' end of the RNA and 50bp into the RNA.

Metagene analysis

To compare the coverage of ATAC-seq regions in ncRNAs we used Bioconductor Package Metagene (2.12.1). Coverage was normalized to reads per million. We binned the position of each region to 100 bp. We modified the current Metagene source code to output boxplots as opposed to ribbons and we then tested the difference of means with R base function TukeyHSD.

Identification of differentially accessible peaks

ATAC-seq peak regions were combined and sorted with bash commands (cat sort). Counts were retrieved from each alignment file using Bedtools multicov (2.26.0) and tested for differential expression with DESeq2. Peaks were considered differentially accessible when log2 fold change greater than 1 and p-adjusted value less than 0.05 and these regions were considered cone-specific and rod-specific. To assess the relationship between differentially expressed ncRNAs and differentially accessible open chromatin we performed a global overlap of all ncRNA regions and combined ATAC-seq regions. Then differential regions from both sets were highlighted.

TF Footprinting and Motif Enrichment Analysis

We used the RGT suite [78] (v 0.12.1) for footprint analysis, and sites were matched with JASPAR CORE 2016 database to assign a motif. From the differential ATAC vs differential ncRNA overlap in rods and cones (Figure 2B), we asked what TF footprint was associated with each ATAC peak from whole retina and in sorted rods vs cones. We then plotted the ncRNA fold change and labeled the top 10 families.

Known and de novo motif scanning was performed with HOMER (4.3) on the 5' end the ncRNAs extended by 500bp and 50 bp into the RNA. We used regions that failed to overlap with any ATACseq or ChIPseq peak.

CRX ChIP-seq comparison and heterotypic binding

Wild-type CRX, Nrl-/- CRX , and NRL ChIPseq regions refs were combined into a master list and each region was assigned a binding category (NRL alone, NRL and CRX, etc.). All regions were overlapped with the 5' region of ncRNA background, the foldchange distribution was plotted with ggplot2, and color-coded by binding category. For each category, the average expression by genotype of the ncRNA background was compared with R by student T-test and cohen's d.

Comparison with previous ncRNA analysis

De novo transcriptome and Fastq files were downloaded from GSE74660 [79]. Fastq files were aligned with STAR, and alignment files were compared with our ncRNA background using bedtools multicov, Counts were normalized with R and plotted with pheatmap. GSE74660 transcriptome was filtered by fpkm > 1 and 2 > exons, and overlapped with ncRNA background using R genomicRanges.

Retina electroporation, AP staining, and imaging

For a full description of these methods please see Perez-Cervantes and Smith, 2020 [16]

2.3 Results:

Nrl-dependent coding and non-coding transcriptional profiling identifies photoreceptor CREs

We interrogated *Nrl*-dependent coding and non-coding transcription in the mouse retina. We performed mRNA transcriptional profiling to determine *Nrl*-dependent coding gene expression [16] (Perez-Cervantes and Smith, Figure 1A). We sequenced cDNA libraries made from polyA+ selected RNA from retinas of litter-matched WT and $Nrl^{-/-}$ adult mice at postnatal day 21 (P21) (WT vs $Nrl^{-/-}$,n=5 and n= 6 resp.). By P21, mouse photoreceptor differentiation is largely complete. Differential expression analysis revealed 4,315 misregulated genes (Perez-Cervantes and Smith, Figure 1A). *Nrl* expression was absent from $Nrl^{-/-}$ samples, along with numerous known *Nrl* targets and rod-specific genes, including *Rho*, *Gnat1*, and *Nr2e3*. Conversely, genes whose expression is normally absent in rods, including those involved in cone phototransduction, were upregulated in $Nrl^{-/-}$ samples, such as *Gnat2, Gngt2*, and *Opn1sw* [56]. These expression changes are consistent with previously identified *Nrl*-dependent gene expression and rod- versus cone-specific gene expression patterns in the adult retina [48], [51], [56], [59], [80], [81].

We performed non-coding transcriptional profiling to identify *Nrl*-dependent ncRNAs by deep sequencing non-polyadenylated RNA from the same control and $Nrl^{-/-}$ retinal samples (Perez-Cervantes and Smith, Figure 1B). This approach identified approximately 30,000 retinal non- coding transcripts by *de novo* transcript assembly (Perez-Cervantes and Smith, Supplemental Figure 1A). 5,900 of these transcripts were *Nrl*-dependent intergenic ncRNAs (FDR<0.05, FC>2). Of the *Nrl*- dependent ncRNAs, 2,553 were significantly downregulated, (referred to henceforth as *Nrl*-activated), and 3,489 were significantly upregulated, (referred to

henceforth as *Nrl*- repressed), and the remainder of ncRNAs were not significantly misregulated ('shared'). (Perez-Cervantes and Smith, Figure 1B). We found a strong linear relationship between samples by genotype, confirming reproducibility (Perez-Cervantes and Smith, Figure 1C).

We and others have hypothesized that quantitative changes in CRE transcription mirror quantitative changes in CRE activity [13], [20], [82]. Accordingly, we examined the quantitative correlation between the change in *Nrl*-dependent ncRNA transcription and the change in the most proximal *Nrl*- dependent gene. The average change in expression of *Nrl*-dependent ncRNAs and the nearest mRNAs within a shared TAD was strongly correlated (Perez-Cervantes and Smith, Figure 1D, Cor=0.76, P=2.2e-16). GO analysis of the nearest genes were enriched for ontology terms associated with phototransduction, sensory perception, and response to light stimulus (Perez-Cervantes and Smith, Figure 1E).

We examined the correlation between epigenomic landscape of mouse photoreceptors and retinal ncRNAs, applying GIGGLE to calculate and rank the significance of the overlap between genomic locations with *Nrl*-dependent ncRNAs and features of retinal enhancers including accessible chromatin, decreased methylation, and H3K27Ac [22], [48], [51], [68]. *Nrl*-activated ncRNAs originated within regions enriched for wild-type and rod enhancer features while *Nrl*-repressed ncRNAs originated within regions enriched for cone and $Nrl^{-/-}$ enhancer features (Perez-Cervantes and Smith, Figure 1F). Specifically, *Nrl*-activated ncRNAs originated from 1,340 genomic regions that overlapped with whole Retina accessible chromatin as defined by Assay for Transposase-Accessible Chromatin (ATAC)-seq (p= 9.642e-201) [51], 1,088 locations of sorted rod ATAC-seq (p= 1.10 e-200) [48], and 820 locations of low methylation in rods (p = 5.32e-108) [51]. *Nrl*-repressed ncRNAs originated from 1,580 genomic regions that

overlapped with whole Retina ATAC-seq (p= 1.01e-200) [51], 1,517 locations of sorted cone ATAC-seq (p= 3.51e-201) [48], and 1,300 locations of low methylation in cones (p = 1.25 e-200)[51]. Shared ncRNAs originated within regions that displayed a lower enrichment score (see Methods) for retinal enhancer features relative to *Nrl*-activated or *Nrl*-repressed regions. Together, these observations indicate that *Nrl*-dependent ncRNA profiling correlates with the retinal epigenomic landscape and local *Nrl*-dependent gene expression. The association of chromatin accessibility with define over 1000 candidate rod- and cone- specific CREs. These findings further indicate that *Nrl*-activated and *Nrl*-repressed ncRNA-defined loci specifically correlate with rod- and cone-specific features of retinal enhancers.

Nrl-dependent ncRNA transcriptional profiling identifies distinct gene regulatory signatures specific to rod and cone photoreceptors

Changes in photoreceptor chromatin accessibility and changes in corresponding rod and cone gene expression are correlated [29], [51], [61]. We hypothesized that the *Nrl*-dependent change in ncRNA abundance may correlate with cell type specific changes in chromatin accessibility. We therefore compared the *Nrl*-dependent changes in ncRNA transcription with the changes in chromatin accessibility in wild-type retina vs $Nrl^{-/-}$ retina and in chromatin accessibility in sorted rods vs cones (Perez-Cervantes and Smith, Figure 2A, B) [48], [51]. We defined differential open chromatin regions as those with a fold change greater than 1 and FDR less than 0.05, identifying 1,040 rod-specific regions at *Nrl*-activated ncRNAs and 990 cone-specific regions at *Nrl*-repressed ncRNAs.We observed a strong positive correlation between ncRNA fold change and open chromatin fold change for whole retina, and for rods vs cones (Perez-Cervantes and Smith, Figure 2A, Cor=0.82, 0.77 resp.). This result indicated that the direction and quantitative degree of *Nrl*-dependence of ncRNA expression correlated with the

relative cell-type specificity of ATAC-seq signal, providing further evidence for the hypothesis that ncRNA expression is a useful predictor of CRE activity in specific cell types. We next assessed *Nrl*-activated and *Nrl*-repressed ncRNA association with ATAC-seq read density in whole retina and sorted rods and cones [48], [51]. We found that *Nrl*-activated ncRNAs emanated from regions associated with higher ATAC-seq signal in wild-type retina and in rods than in cones or $Nrl^{-/-}$ retina, and the inverse was true for *Nrl*-repressed ncRNAs (Perez-Cervantes and Smith, Figure 2B). These observations are consistent with the hypothesis that *Nrl*-activated and *Nrl*- repressed ncRNA-defined regions represent rod- and cone-specific CREs, separately highlighting 1000s of genomic locations as strong candidate *Nrl*-dependent cell-type specific CREs.

We next searched for TF motif sequences that were enriched in ncRNA- associated ATAC-seq peaks. In *Nrl*-activated regions, we observed NR2E3, ESRRB, and RORB motifs, previously identified as enriched in rod versus cone open chromatin regions [29] (Perez-Cervantes and Smith, Figure 2C and D). In *Nrl*-repressed regions we observed enrichment of the Neurod and RXRG motifs, previously identified as enriched in cone- versus rod-specific open chromatin regions [29] (Perez-Cervantes and Smith, Figure 2 C and D). This recapitulation of previously identified motifs indicates that *Nrl*-dependent ncRNAs accurately identify genomic locations likely to be active cell-type specific CREs. We also observed motifs for the CUT, EN, and the brain POU domains in the *Nrl*-repressed regions, suggesting that these families of retinal TFs may play a specific role in cone gene regulation [40], [83]–[86]. Genes in these families also play a role in retinal development, suggesting that ncRNA profiling of mature tissue may also detect enhancers found in the developing retina. Finally, we interrogated putative promoter regions of *Nrl*- dependent ncRNAs that did not overlap open chromatin or TF-bound peaks. At

these elements we found binding sequences for both rod and cone elements, such as bZip, CUT, Homeobox (K59 HD, Q50 HD), Hox (HD), and ESSRB (NR). This observation suggests that ncRNA-profiling identifies candidate regulatory regions undetected by other assays (Perez-Cervantes and Smith, Figure 2 supplement). We conclude that *Nrl*-dependent ncRNA profiling identifies distinct gene regulatory signatures, including cell-type specific TF motifs, at CREs that may participate in cell-type specific gene regulation.

Nrl-dependent ncRNA expression identifies functional output of combinatorial TF action

CRX and NRL cooperatively control photoreceptor gene expression [48]–[50], [56], [59], [61]. Since the combinatorial action of CRX and NRL at least partially determines the specificity of rod vs cone CREs, the binding of these TFs also may be reflected in cell-type specific ncRNA expression. We utilized data from previous CRX ChIP-seq experiments [60] in the wild type and $Nrl^{-/-}$ retina, and tested whether CRX binding events were associated with global changes in ncRNA expression. First we evaluated the overlap of CRX binding with ncRNA regions in the wild-type and $Nrl^{-/-}$ retina (Figure 2.1A, top) and next compared ncRNA fold change in each category of CRX binding (Figure 2.1A, bottom). On average, ncRNAs from regions with wild-type specific CRX binding were more highly expressed in the wild-type retina, while ncRNAs from regions with Nrl -specific CRX binding were more highly expressed in the $Nrl^{-/-}$ retina (Figure 2.1B, red compared to blue). This relationship becomes even clearer when we specifically examined Nrl-dependent ncRNAs; Nrl-dependent ncRNAs within regions bound by CRX in wild-type were more highly expressed in wild-type (Cohen's d=0.51, p<2.2e-16) and Nrl-dependent ncRNAs within regions bound by CRX in $Nrl^{-/-}$ were more highly expressed in $Nrl^{-/-}$ retina (Cohen's d=0.78, p < 2.2e-16). ncRNAs in regions bound by CRX in both wild type

and *Nrl*$^{-/-}$ retina displayed negligible differences in expression between the mutant and wild type (Cohen's d=0.13, p =0.005) (Figure 2.1B).

We next tested whether coordinate binding by CRX and NRL correlated with a greater effect on ncRNA expression in either the wild-type or *Nrl*$^{-/-}$ state than CRX binding alone. We observed that ncRNAs from regions with NRL and CRX binding in the wild-type retina had a larger decrease in ncRNA expression in the *Nrl*$^{-/-}$ retina compared to those from regions with either NRL or CRX binding alone (Cohen's d: 0.99; p-value: <2.2e-16) (Figure 2.1C, column 4). ncRNA expression from regions with CRX binding in the absence of NRL demonstrated the inverse relationship, with ncRNA expression displaying a large increase in the *Nrl*$^{-/-}$ retina near CRX-bound regions (Cohen's d: 0.83; p-value: <2.2e-16) (Figure 2.1C, column 7). These relationships suggest that, at the genome scale, co-binding of CRX and NRL drive Nrl-activated ncRNA- defined CRE expression in a rod-predominant context, and that in the absence of NRL, unique CRX binding at CREs drive NRL-dependent ncRNA activity in a cone- predominant (*Nrl*$^{-/-}$) context [48], [51]. Together, these findings support the model that combinatorial NRL and CRX action defines cell-type specific expression patterns by providing quantitative support of the output of TF binding events on cell-type specific CREs.

Figure 2.1: Unique CRX binding events correlate with context-specific ncRNA expression

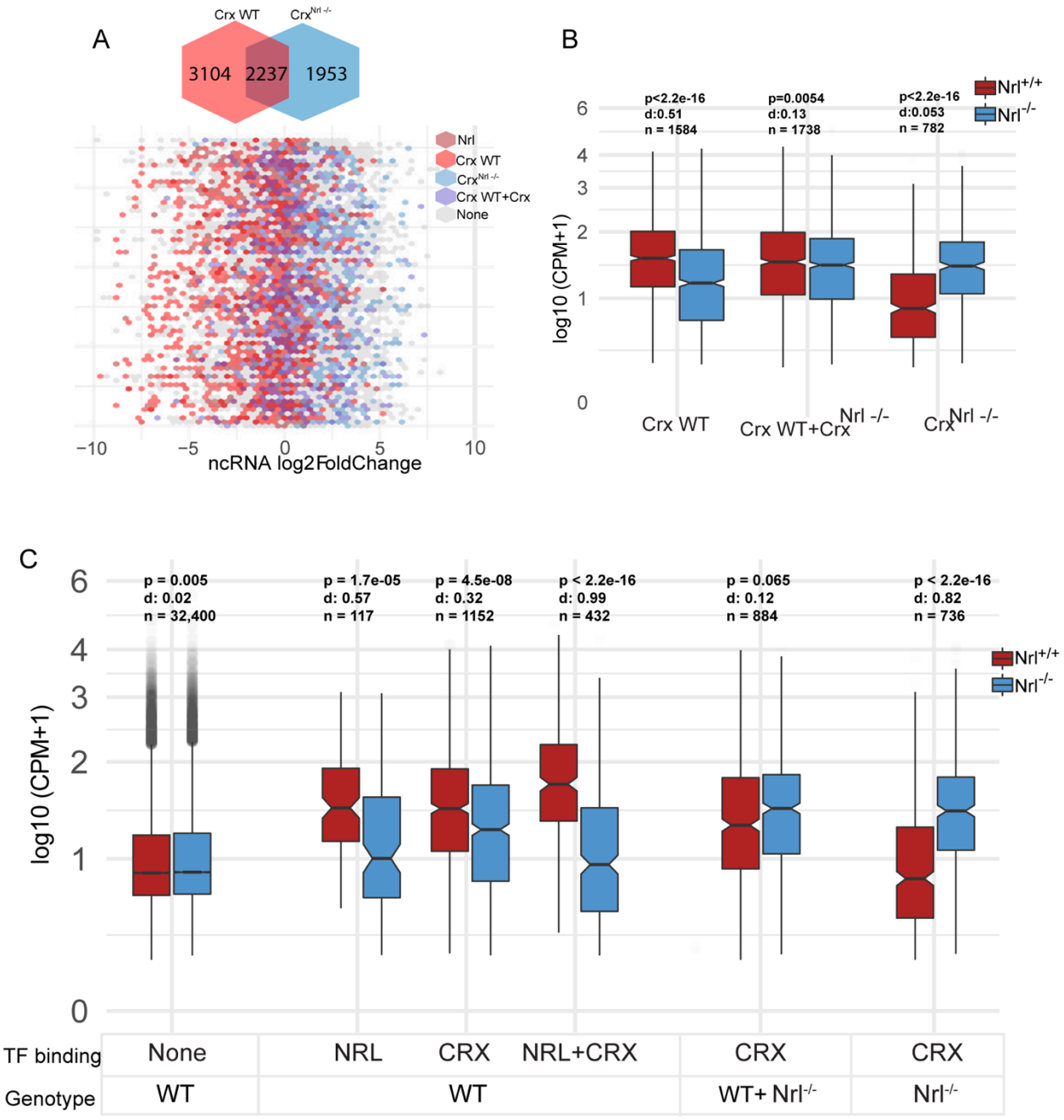


Figure 2.1: **Legend**

A) Venn diagram of the overlap of Crx wild-type ChIP-seq (red) and $Crx^{Nrl-/-}$ ChIP- seq (blue) peaks (top). Scatterplot in hexagonal binning depicting ncRNA fold change x-axis. CHIP-seq Binding events that intersect at the 5' end of the RNA highlighted: red for wild-type Crx ChIP-seq peak, blue for $CRX^{Nrl-/-}$ ChIP-seq, and purple for shared Crx sites (Corbo et al., 2010).

Figure 2.1: **Legend, cont.**

B) Boxplots depicting mean expression differences between global *Nrl*-dependent ncRNAs categorized by CRX ChIP-seq binding. P-value calculated by student T- test and effect size by Cohen's d test.

C) Boxplots depicting mean expression differences between global *Nrl*-dependent ncRNAs categorized by CRX ChIP-seq binding and NRL ChIP-seq. P-value calculated by student T-test and effect size by Cohen's d test.

Nrl-dependent ncRNAs define functional photoreceptor regulatory elements

Our data suggest that *Nrl*-dependent ncRNAs locate 1000s of functional CREs with cell-type specific activity. Specifically, *Nrl*-activated ncRNA-defined regions with rod-specific ATAC defined 1,040 candidate rod enhancers and *Nrl*-repressed ncRNA- defined regions with cone-specific ATAC define 990 candidate cone enhancers (Perez-Cervantes and Smith, Figure 2B, Supplementary Table 1). We tested the regulatory capacity of *Nrl*-dependent ncRNA-defined regions in developing mouse retinas. As *Otx2* is required for the genesis of both rods and cones, understanding its regulation is of interest [37], [38], [40]. We identified 24 DNaseI hypersensitivity sites (HS) at the *Otx2* locus using P0 mouse retina data from ENCODE (Perez-Cervantes and Smith, Figure 4A) [22]. Of these sites, 4 overlapped *Nrl*-repressed ncRNAs and cone-specific ATAC- seq peaks (FC > 2; Perez-Cervantes and Smith, Figures 2A, 4B). We examined the activity of all 24 DNAse1 HSs to determine if the CRE activity of these sites would correlate with areas defined by *Nrl*- repressed ncRNAs. We cloned sequences corresponding to each DNaseI HS into an eGFP-IRES-AP reporter [87], [88] and co-electroporated the constructs with a control plasmid containing a ubiquitously expressing CAG-Cherry reporter. Electroporations were done into explants from embryonic day 14.5 (E14.5), a developmental time point during which cones are being generated [89]. Alkaline phosphatase

(AP) activity was assayed on whole mounts, as it provides a quick and sensitive detection of CRE activity. Nine out of the 24 DNase I HSs showed AP activity, with five of these sites being extremely active (Perez-Cervantes and Smith, Figure 4C). The three most active sites overlapped with *Nrl*-repressed ncRNAs, suggesting that ncRNA-defined regions are often the most active CREs, compared to regions defined by open chromatin alone [13]. Cone development is also regulated by Onecut1 (*Oc1*) [40], [85], leading us to test several DNaseI HS near *Oc1.* Similarly to *Otx2*-associated CREs*,* we found that the Oc1 region with the strongest CRE activity was defined by ncRNA expression (Perez-Cervantes and Smith, Figure 4D, OC1 C compared to OC1 A and B).

We predicted that *Nrl*-repressed ncRNA-defined regions would represent CREs active in cones, but not rods. Such regions were found near other genes important for cone development and function, including *Opn1sw, Rxrg, Gngt2, Gnat2,* and *Sall3* (Perez-Cervantes and Smith, Supplementary Table 1). ncRNA-defined regions from near these loci were thus interrogated for cone-specific CRE activity using electroporation of embryonic explants. To determine cone activity, we co-electroporated the ThrbCRM1-tdTomato plasmid, a reporter with activity predominantly in cones, with some activity also in horizontal cells (HCs) and the cone/HC progenitor cells, but no activity in rods [40]. Ten of the 12 ncRNA-defined regions were able to drive reporter expression, including those with CRE activity from the *Otx2* locus (Perez-Cervantes and Smith, Figure 4E).

To determine if the *Nrl*-repressed ncRNA regions with CRE activity were active in cones, but not rods, we examined tissue sections for expression of eGFP driven by the eGFP-IRES-AP reporters and tdTomato expression driven by the ThrbCRM1 reporter. The majority of tdTomato

positive cells were located in the apical region of the electroporated E15.5 retinas (Perez-Cervantes and Smith, Figure 5A), the location of developing cones [90]. tdTomato positive cells also expressed the Rxrg protein, a known cone marker (Perez-Cervantes and Smith, Figure 5B) [91]. Cells that expressed GFP from 7 of the *Nrl*-repressed ncRNA constructs were located in the apical region. Importantly, they also expressed tdTomato (Perez-Cervantes and Smith, Figure 5C). The only exception was the *Socs3* putative CRE (Perez-Cervantes and Smith, Figure 5C), which marked cells with a morphology and position matching those of RPCs. We further tested the expression from the *Otx2* ncRNA-defined enhancer regions 1 and 2 relative to that of the OTX2 protein (Perez-Cervantes and Smith, Figure 5D). Both *Otx2*-ncRNA-defined CREs showed strong co- localization with the Otx2 protein. Together, these findings indicate that the genomic regions nominated by *Nrl*-repressed ncRNAs have CRE activity in developing cones. These findings validate the hypothesis that *Nrl*-dependent ncRNA profiling identifies functional cell-type specific regulatory elements.

2.4 Discussion:

Enhancer transcription defines context-specific CREs

Non-coding transcriptional profiling from specific biological contexts was evaluated as an approach for nominating candidate regulatory elements in the mouse retina. A strong positive quantitative relationship between TF-dependent ncRNA expression and transcriptional activation in enhancer assays, as well as with local mRNA expression, and retinal epigenomic markers showed that differential ncRNA-seq provides a robust genome-wide assay to identify functional CREs. This approach specifically nominates candidate regulatory elements that are most likely to be functional, as shown by the electroporation assays for CRE activity. Furthermore, adding quantitative ncRNA data to the existing genomic metrics at retinal CREs informed TF-dependent models of retinal gene expression. When incorporated with existing genomic information, ncRNA expression provided contextual and quantitative evidence for functional output of CREs and the GRNs they comprise, as demonstrated through our analysis of the transcriptional effect of combinatorial TF action (Figure 2.1). TF-dependent ncRNA profiling provides an alternative approach to large scale reporter assays to identify putative regulatory regions, which can be difficult to apply *in vivo*, or when the cells of interest are not accessible, as is the case for cones. The example herein using the retina, and the previous study using cardiomyocytes, show that TF-dependent ncRNA profiling is successful as a general approach for the definition of regulatory elements.

Regulatory regions active in photoreceptors and their progenitor cells

An understanding of the GRNs that govern photoreceptor production is particularly important for therapeutic applications [92]. In particular, cone photoreceptors, the cell type that

mediates color and high acuity vision, degenerate in a variety of blinding disorders, retinitis pigmentosa or aged-related macular degeneration (AMD) [93], [94]. Therapeutic approaches focus on cone-replacement using cells derived from stem cells, or utilize gene therapy to specifically target cones [92]. Both approaches would benefit from a greater understanding of cone vs rod gene regulatory networks [95]–[97]. The conversion of rods to cone-like cells in the *Nrl*$^{-/-}$ mouse has provided access to an enriched source of cone-like cells, allowing for the elucidation of molecular aspects of rods vs cones, including the identification of rod- and cone-enriched coding and non-coding transcripts and open chromatin regions [35], [48], [49], [51], [82]. These datasets have provided an excellent background for our assessment of the rod- and cone-specific transcriptome and quantitative analysis of CRE activity changes in a rod-predominant state (wild-type retina) and a cone-predominant state (*Nrl*$^{-/-}$ retina).

The activity of ncRNAs at regulatory regions is associated with their enhancer activity [20], [98], suggesting that the CREs identified in this study at P21, after retinal development is complete, should be active in mature cones. Interestingly, functional examination of enhancer activity for a subset of ncRNA-defined CREs regulating relevant cone genes showed cell-type specific activity in developing cones, and often identified the most active regulatory regions (Perez-Cervantes and Smith, Figure 4). These observations suggest that the presence of ncRNAs is a signature for strong enhancers and/or enhancers that are active throughout development to adult stages. Accordingly, some of the tested ncRNA-defined regions overlap with previously published ncRNAs from earlier developmental time points (P2/4/6/10) in the *Nrl*$^{-/-}$ retina, suggesting sustained activity of those regions [35] (Perez-Cervantes and Smith, Supplemental Figure 4).

A small number of tested elements had activity in other cell types or had no specific activity in retinal explants. This may be due to the fact that the CREs identified by ncRNA profiling and ATAC-seq peaks are fairly large, and thus may harbor multiple cell-type specific TF binding sites, or may rely on higher order chromatin structure for proper regulation. Elements that did not show activity in retinal explants may be solely active in mature cones, and not in the cone progenitor cells or immature cones that were assayed here. Interestingly, *Oc1*, which is upstream of *Nrl* [40], is not expressed in mature photoreceptors [99], however we identified a candidate *Nrl*-dependent ncRNA-defined CRE at this locus that displayed strong activity in the early retina (Perez-Cervantes and Smith, Figure 5D). It is possible that ncRNA transcription may reflect an epigenetic memory of embryonic activity, similar to previous findings that vestigial enhancers may remain undermethylated in adult tissues [51], [100], [101].

While many of the genes associated with *Nrl*-repressed ncRNAs were related to cone development or function (*Otx2, Rxrg, Gngt2, Gnat2, Opn1sw, Sall3*) [38], [91], [102]–[104], we also identified CREs for genes that are generally involved in retinal development that currently have no known cone-specific role, except for their enrichment in adult cones and $Nrl^{-/-}$ retinas [48], [105], [106]. For example, the ncRNA-defined CRE associated with *Six6*, an eye-field TF [84], and the element near *En2,* a gene important for ganglion cell differentiation [83], both displayed cone-specific reporter activity. Recently, *Six6* has been found to regulate middle-wavelength opsins in zebrafish [107]. These results suggest that non-coding transcriptional profiling of the wild-type and $Nrl^{-/-}$ retina uncovered not only rod- and cone-specific regulatory programs, but potential shared regulatory programs that warrant further investigation. Overall, the regulatory elements nominated in this analysis represent potent candidates that may be relevant for cone-replacement strategies and gene therapy, may shed insight on developmental

and mature CRE usage, and may add novel players to our understanding of developmental cone GRNs.

Heterotypic NRL/CRX localization defines ncRNA abundance at rod vs cone CREs

In retinal development, CRX has been shown to interact with NRL and recruit additional TFs, including NR2E3, to regulate photoreceptor genes [52], [58], [108]. The combinatorial action of CRX and NRL has been hypothesized to target expression specifically to rods [48], [58], suggesting a model of heterotypic TF action, where CRX alone generally targets photoreceptor enhancers and the addition of NRL provides rod- specificity. Using ncRNA abundance and differential expression in the wild-type and *Nrl*$^{-/-}$ retina, we provide quantitative transcriptional evidence for the combinatorial action of CRX and NRL. The binding of CRX and NRL together correlated with rod-specific ncRNA transcription, and in the absence of NRL, CRX was preferentially located at CREs with strong cone-specific ncRNA transcription (Figure 2.1). These rod- and cone- specific binding events are distinguished from shared CRX binding in both rods and cones, which did not display cell-type specific non-coding transcription. This heterotypic model, supported by relative ncRNA abundance as a measure of altered CRE activity, may drive rod vs cone CRE choice and thereby regulate gene expression that determines cell type during retinal development. Overall, this finding highlights the utility of differential ncRNA profiling for nominating potent CREs and deepening our understanding of context-specific gene regulatory networks.

2.5 Acknowledgement of Work Performed:

I'd like to acknowledge the individuals who contributed to this work. This work was conceptualized by Rajiv Nadadur, Carlos Perez-Cervantes, Linsin Smith, Nicolas Lonfat, Constance Cepko and Ivan Moskowitz. Methodology was constructed by Rajiv Nadadur, software by Carlos Perez-Cervantes, and experimental validation by Nicolas Lonfat. Formal analysis was conducted by Carlos Perez-Cervantes, Linsin Smith, and Nicolas Lonfat. Writing of the manuscript (original draft) was performed by Linsin Smith, Carlos Perez-Cervantes, Rajiv Nadadur, Constance Cepko, Nicolas Lonfat, and Ivan Moskowitz. Manuscript review and editing was performed by Linsin Smith, Carlos Perez-Cervantes, Nicolas Lonfat, Constance Cepko, Ivan Moskowitz, and Joseph Corbo. Data curation was performed by Carlos Perez-Cervantes, and Carlos Perez-Cervantes, Linsin Smith, and Nicolas Lonfat all contributed to data visualization.

The figure presented in this chapter (Figure 2.1) was constructed from data originally analyzed by Linsin Smith and Carlos Perez-Cervantes, and both Linsin Smith and Carlos Perez-Cervantes contributed to data visualization and interpretation. The writing presented in this chapter is composed by Linsin Smith, Carlos Perez-Cervantes, and Nicolas Lonfat; overseen by Constance Cepko, and Ivan Moskowitz. Funding, administration, and supervision details can be found in the published manuscript at https://doi.org/10.1242/dev.184432.

Chapter 3: Identification of direct transcriptional targets of NFATC2 that promote β cell proliferation

3.1 Introduction:

Diabetes mellitus is a chronic metabolic disorder characterized by hyperglycemia and the inability of the body to assimilate glucose and maintain glucose homeostasis. The persistent hyperglycemia attributed to diabetes mellitus can have drastic consequences on the body, leading to long term organ damage and organ failure; particularly of the eyes, kidneys, nerves, heart, and blood vessels. The prevalence of diabetes mellitus has been increasing, and represents a large burden on healthcare systems worldwide. It is estimated that there were 382 million people worldwide living with diabetes in 2013, and that by 2035 this number will have risen to 592 million [109]. Diabetes mellitus is associated with premature death due to cardiovascular disease and cancer as well as other causes, and has been classified by the WHO as the seventh leading cause of death worldwide [110], [111].

Type 2 Diabetes (T2D) is the most prevalent form of diabetes mellitus, comprising 80-90% of all reported diabetes cases [112]. T2D is an insulin-resistance condition which develops when pancreatic β cells dysfunction and are unable to produce insulin, do not produce enough insulin, or are unable to release sufficient insulin to overcome the resistance. β cell dysfunction may be in part caused by reductions in β cell mass or by functional defects in β cells, which have both been documented in T2D patients, however the relationship between β cells mass reduction and dysfunction has not been fully elucidated [113]. Many cellular stresses can lead to impaired insulin biosynthesis and secretion from β cells in response to glucose, culminating in β cell failure, apoptosis, or necrosis [114]. There is no cure for T2D, however it is often treated with a

combination of lifestyle changes such as adopting a healthy diet and increasing exercise, as well as medications that aim to reduce glucose production and increase sensitivity to insulin.

T2D has been documented to have genetic, epigenetic, and environmental causes, and is largely thought to result from a complex interplay between all three. Many lifestyle factors, including diet, exercise, and smoking, as well as environmental factors such as pollutants and toxins have all been associated with T2D. Outside of these environmental and non-genetic contributions to T2D, there have also been genetic associations to T2D and T2D risk. The heritability of T2D varies widely across populations, however family studies have revealed that the risk of developing T2D is 40% for individuals who have 1 parent with T2D, indicating a clear genetic component [115]. Many genes have been connected to T2D through candidate gene studies, linkage studies, and GWAS; which to date have linked over 200 genetic loci with T2D risk [111], [112], [116]. However there still is no clear understanding of exactly how these associated variants contribute to T2D risk, as well as how these genetic associations can be leveraged to develop better therapeutics.

Deeper investigation into the genetic underpinnings of T2D is necessary, not only to broaden our understanding of the heritability of T2D, but also the pathophysiology of T2D, with the goal of applying that knowledge to therapeutics. As GWAS have identified a multitude of variants associated with T2D risk, one approach is to explore the potential relationships between risk associated genes. Previous work from the Attie lab asked if variation in the expression of candidate genes located at T2D associated genetic variants were regulated by one or more transcription factors, thereby attempting to resolve a T2D gene regulatory network [117]. Using an obese F2 mouse intercross segregating for T2D, they showed that the expression of ~40% of the T2D-associated genes was linked to a broad region on mouse chromosome 2. Using

mediation analysis they were then able to identify the transcription factor Nfatc2 as a potent transcriptional regulator of these genes.

To further understand the role of Nfatc2 as a transcriptional regulator of T2D-associated genes, they overexpressed a constitutively active form of the protein in both mouse and human islets. They found that NFATC2 overexpression lead to enhanced insulin secretion and β cell proliferation, highlighting the role of this TF in β cell function [117]. This finding necessitates deeper investigation into the molecular mechanisms of NFATC2-dependent β cell proliferation and insulin secretion, as understanding the mechanisms of this relationship to shed light on the genetic pathways that induce proliferation and insulin secretion in islets.

3.2 Methods:

For a full list of methods, please see Simonett et al, 2021 10.1172/JCI144833. Methods included below are methods specific to analyses presented in this dissertation.

Nfatc2 ChIP-seq in human islets

48 hours after Nfatc2 adenoviral treatment, ~17,000 human IEQs were prepared for ChIPseq. For more details, please check Simonett et al, 2021.

Sequencing was performed at the University of Wisconsin-Madison Biotechnology Center. DNA concentration and size distribution were verified using the Qubit dsDNA HS Assay Kit (Invitrogen, Carlsbad, California, USA) and Agilent DNA HS chip (Agilent Technologies, Inc., Santa Clara, CA, USA), respectively. Samples were prepared according the TruSeq ChIP Sample Preparation kit (Illumina Inc., San Diego, California, USA) with minor modifications. Libraries were size-selected for an average insert size of 350 bp using SPRI-based bead selection. Quality and quantity of the finished libraries were assessed using an Agilent DNA1000 chip and Qubit dsDNA HS Assay Kit, respectively. Libraries were standardized to 2 nM. Paired end 125bp sequencing was performed to a depth of 68 – 154 million reads per sample, using SBS chemistry (v4) on an Illumina HiSeq2500. Images were analyzed using the standard Illumina Pipeline, version 1.8.2.

The raw reads from all samples were aligned to human reference genome (hg19) with Bowtie 2 (version 2.3.1). We used Bowtie2 default reporting mode with the argument *-X 550 --no-mixed --no-discordant* being specified. Picard MarkDuplicates tool (version 2.9.2) was used to remove duplicate reads from the alignment. We performed peak calling from mapped and properly paired reads with Bioconductor package MOSAiCS (version 3.5). The peak sets called

from the six samples were merged and 8,635 peaks were retained based on the following criteria. We included (i) peaks reproducible in at least two samples and (ii) peaks identified in one sample, whose summit signals are higher than the third quantile of the summit signals of the reproducible peaks. Further, peaks in ENCODE DAC blacklist regions were dropped from the peak list.

De novo motif analysis was performed using MEME-ChIP and identified TTTCCR as the most significant motif present at all Nfatc2 binding sites[118]. This motif is similar to a known NFATC2 motif with a Fisher's exact test p-value = 2.0e-668. To identify motifs that co-occur at Nfatc2 binding sites, we performed spaced motif analysis (SPAMO) on the sequences around the summits of the merged peaks. We discovered the FOXP1 motif, MA0481.2, significantly co-occurred with the NFATC2 motif (E-score of 0.0022). We used FIMO to confirm the co-occurrence of NFATC2 and FOXP1 motifs, and to compute the binding proximity. Co-occurrence from the dependency between NFATC2 and FOXP1 motif occurrences that Pearson's chi-square testing supports with p-value less than $2.2e^{-16}$. The partner TFs may bind to DNA sequences of the 4,349 peaks where the both motifs are identified, within a window of 500 bp at most, having 26.9% of them sharing binding sites (1170/4349). The ChIPseeker R package (v1.14.2) was used to annotate peaks and their transcriptional start sites (TSS) within 3kb[119].

Sequencing of human islet transcriptome

48 hours after adenoviral treatment, ~6,000 human IEQs per condition (Ad-GFP, Ad-caNfatc1 and Ad-caNfatc2), islets were collected, and prepared for sequencing. For full methodology, please see Simonett et al 2021 at 10.1172/JCI144833.

The human islet RNA-sequencing reads were mapped via bowtie against the refseq hg19 reference (19261174). Gene expression values were then estimated via RSEM (21816040). Expected counts were normalized using median-by-ratio normalization (23975260). EB-seq was used to classify genes into 5 distinct patterns of differential expression (23428641). Differential expression (DE) was defined by between the three conditions: (GFP-Control, C1-Nfatc1, and C2-Nfatc2). Pattern 1: C1 = C2 = GFP (no DE); Pattern 2: C1 = C2 ≠ GFP (DE for C1 and C2 equally); Pattern 3: C2 ≠ C1 = GFP (DE for C2 only); Pattern 4: C1 ≠ C2 = GFP (DE for C1 only); and Pattern 5: C1 ≠ C2 ≠ GFP (DE for C1 and C2 unequally). For each gene, EB-seq computed a posterior probability (PP) associated with each expression pattern. The higher the PP for a particular pattern, the more likely the gene is following that pattern.

Human-mouse ortholog genes were annotated using biomaRt (v2.33.4). We identified genes as DE overall as those with posterior probability of Pattern 1 < 0.05 in both human and mouse datasets. Orthologs were considered DE for a specific pattern if its posterior probability was > 0.50 for that pattern. All statistical analyses were carried out using R version 3.4.1.

ATAC-seq sample preparation:

Please see 10.1172/JCI144833 for a full description of ATAC seq sample methodology.

ATAC-seq data analysis

Illumina Nextera adapters were trimmed with cutadapt (version 2.0) (Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal.) with option "-q 30 --minimum-length 36". Paired-end ATAC-seq reads were aligned to the human genome assembly (hg19) with bowtie2 (version 2.3.4.1) with option "-X 800 --no-mixed --no-discordant"[74]. For each sample, unmapped reads were filtered out by using SAMtools view

(version 1.8)[75], with option "-F 4" and mitochondrial reads were removed. Duplicate reads were removed with Picard tools (version 2.9.2) ("Picard Toolkit." 2019. Broad Institute, GitHub Repository. http://broadinstitute.github.io/picard/; Broad Institute). We obtained on average 105 ± 20 million reads per sample.

In total, 231,232 master ATAC peaks (mean width 663 ± 393 bp) were called for pooled GFP and Nfatc2 samples with MACS2[76]. We utilized "getCounts()" function in chromVAR to obtain the count matrix of Tn5 cuts. We then performed differential accessibility analysis with DESeq2[66] and obtained Nfatc2 opening and closing chromatin. The genomic compartment annotations were obtained using ChIPseeker[119]. Motif enrichment analysis was performed using with MEME-CHIP[118].

***Noncoding RNA-seq library preparation*:**

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

***Noncoding RNA-Seq Data analysis*:**

Non-coding RNAseq data analysis was performed as described in Perez-Cervantes and Smith et al. 2020[16]. Briefly, Fastq files were aligned to UCSC genome build hg19 using STAR (v2.5.1), transcript assembly was performed by Stringtie (version 1.3.3) and merged with the

Stringtie ‘merge’ function. Counts were retrieved from the merged transcriptome with HTseq (v0.6) differential expression was performed with R (3.4) package DEseq2[66]. False discovery rate (FDR) was calculated after removing the coding-gene transcripts, ncRNA that overlapped coding genes in the same strand (bedtools intersect 2.29.0), and RNAs shorter than 100 base pairs. Significance was considered to have been reached when FDR was <0.05 and |log2 fold-change| was >0.5. Open chromatin peaks and TF-binding peaks were intersected with ncRNAs using Bioconductor package GenomicRanges[77] allowing for a 1000 bp gap upstream the 5’ end of the RNA and 50 bp into the RNA.

3.3 Results:

To understand the mechanism by which NFAT overexpression stimulates β-islet proliferation, our collaborators performed RNAseq following the overexpression of a constitutively active form of *Nfatc1* or *Nfatc2*. 8 distinct human islet donors were treated with *Nfatc1* and *Nfatc2* overexpression, leading to the identification of 19,000 islet transcripts, 5,589 of which were differentially expressed in response to overexpression of either NFATC1 and/or NFATC2 [17] (Simonett et al, Figure 1A). Our collaborators found that more than 50% of the differentially expressed genes were regulated concordantly by NFATC1 and NFATC2, indicating a large overlap in their regulatory activity. In order to look more closely at the induction of β-cell proliferation in response to NFAT overexpression, they next explored a gene set of 1,700 human genes associated with regulation of cell cycle (Gene Ontology GO:0007049). Out of these 1,700 genes, 1,642 are expressed in human islets, of which 723 are DE in response to NFATC1 or NFATC2 overexpression. This indicates that NFAT is directly or indirectly regulating a large suite of cell cycle genes, further suggesting its role in regulating a β islet proliferation GRN. Previous work had identified *ASF1B* as a cell cycle gene that is sufficient to drive β-cell proliferation in human islets. *ASF1B* was identified as upregulated in response to NFATC1 or NFATC2 overexpression, suggesting that NFAT may be a master regulator of β-cell proliferation.

To further understand the gene regulatory networks mediated by NFATC1 and NFATC2, GO analysis was performed on genes that were misregulated in both overexpression contexts. They found that downregulated genes were enriched for terms related to nuclear function ($z>12$) and upregulated genes were enriched for terms related to mitochondrial function ($z>9$). They also found NFATC2-specific upregulated genes were enriched for terms related to insulin secretion

(z=4.5), mirroring the functional findings that NFATC2 overexpression led to enhanced insulin secretion, but NFATC1 overexpression did not. These findings highlight the role of NFATC2 in regulating gene expression to induce β-cell proliferation and insulin secretion in human islets.

To identify direct transcriptional targets of NFATC2, our collaborators performed ChIPseq for NFATC2 on 6 distinct human islet donors, identifying approximately 8,600 binding sites genome wide [17] (Simonett et al, Figure 2). 35% of these binding sites were located in promoter regions, and many were extremely close to the TSS of the nearest gene, indicating NFATC2 is likely directly regulating the transcription of these target genes (Simonett et al, Figure 2A). Additionally, a large proportion of NFATC2 binding occurred in intergenic and intronic regions, which often harbor non-coding regulatory elements, further indicating the role of NFATC2 in regulating gene expression through regulatory elements such as enhancers. To more deeply investigate the role of NFATC2 in regulating gene expression, our collaborators analyzed the extent to which NFATC2 binding overlaps with other TF binding and epigenome modifications indicative of active transcription. They found that more than 60% of NFATC2 binding sites also contained strong signals for islet TFs such as NKX2-2, FOXA1, and PDX1 (Simonett et al, Figure 2C). Additionally, they observed that more than 70% of NFATC2 binding sites contained strong signals for H3K4me1, a known feature of active enhancer elements [23]. These results further indicate NFATC2 may work in concert with other key islet TFs at gene promoters as well as critical regulatory loci in human islets.

To elucidate the direct transcriptional networks regulated by NFATC2 that lead to enhanced β-cell proliferation, our collaborators integrated the NFATC2 overexpression RNAseq with the ChIPseq data. Out of the ~5,500 DE genes and 8,600 NFATC2 binding sites, they identified approximately 2,200 DE genes that had one or more NFATC2 binding events within

50Kbp of the TSS (Simonett et al, Figure 3A). Of these 2,200 genes, 1,425 of them were upregulated in response to NFATC2 overexpression and 786 were downregulated, highlighting the role of NFATC2 as both a transcriptional activator and a repressor (Simonett et al, Figure 3B).

In order to gain insight to the biological pathways that may be directly regulated by NFATC2, GO analysis was performed on NFATC2-bound target genes that were differentially expressed in response to NFATC2 overexpression. Our collaborators identified a significant enrichment for genes involved in the unfolded protein response and the heat shock response in the 786 directly bound genes downregulated in response to NFATC2 ($z > 8$). In contrast, the 1,425 directly bound upregulated genes were enriched for regulation of cellular communication by electrical coupling ($z \approx 6$), histone methyl-transferase complex ($z \approx 5$), and pathways associated with ciliary transport ($z \approx 5$).

To further characterize the transcriptional effect of Nfatc2 overexpression, we interrogated the non-coding epigenomic landscape. While ~35% of identified NFATC2 binding sites were promoter-proximal, approximately 60% of Nfatc2 binding events occurred in intergenic and intragenic regions, regions of the genome that often contain non-coding regulatory elements. As Nfatc2 overexpression results in large transcriptional changes, we hypothesized that non-coding NFATC2 binding events may identify Nfatc2-dependent enhancers that mediate the observed changes in gene expression. To further understand these potential Nfatc2-dependent enhancers we used non-coding transcriptional profiling to identify active Nfatc2-dependent regulatory elements.

It has been well documented that non-coding regulatory elements, including enhancers, often produce non-coding transcripts [9], [12], [13], [16], [120]. These transcripts have been

shown to be reliable predictors of enhancer activity in a variety of systems, and in some cases to be required for enhancer activity. In order to interrogate Nfatc2-dependent enhancers, we performed non-coding RNA sequencing on 3 distinct human beta islet donors, each treated with either control adenovirus (Ad-GFP), or the adenovirus expressing ca-Nfatc2. Out of ~920,000 identified ncRNAs, ~12,000 of them were dependent on Nfatc2 expression, with ~6,000 downregulated and ~6,000 upregulated in response to Nfatc2 overexpression (Figure 3.1A**)**. We term these downregulated and upregulated ncRNAs as Nfatc2-dependent ncRNAs, as their expression changes in response to Nfatc2 overexpression.

In addition to ncRNA expression, a hallmark of enhancer activity is the presence of open chromatin. As shown earlier, ~45% of the Nfatc2 binding sites co-occurred with previously identified β-cell ATAC-seq peaks, indicating that Nfatc2 commonly binds to open chromatin sites, but may also pioneer relatively closed chromatin, highlighting its potential importance in non-coding regulatory region activity. To further characterize the effect of Nfatc2 overexpression on chromatin accessibility we performed ATAC-sequencing on islets from three separate human donors, each treated with either control adenovirus (Ad-GFP), or the adenovirus expressing ca-Nfatc2. Among >230,000 ATAC peaks we identified genome-wide, ~58,000 showed differential accessibility ($\log_2 > 0.5$, FDR < 0.05) in response to Nfatc2 (Figure 3.1B**)**.

We next interrogated the relationship between Nfatc2-dependent ncRNAs and Nfatc2-dependent open chromatin. Overall, 21,017 of our identified ncRNAs emanate from open chromatin regions. We observed a strong positive correlation between ncRNA fold change and open chromatin fold change, indicating that the direction and quantitative degree of Nfatc2-dependence of ncRNA expression and open chromatin signal is tightly linked (Pearson's Corr = .74, Figure 3.1C). This finding provides further support that Nfatc2-dependent ncRNAs are

active non-coding regulatory regions and reflect the epigenomic landscape in response to Nfatc2 overexpression.

We next separated the 21,017 ncRNAs that emanate from open chromatin into their respective bins of open chromatin: closing chromatin (chromatin that loses accessibility in the Nfatc2 overexpression compared to WT), shared open chromatin (chromatin that stays relatively accessible in both the overexpression and in WT), and opening chromatin (chromatin that gains accessibility in the overexpression compared to WT) (Figure 3.1D). We hypothesized that these bins of accessible chromatin represent different categories of enhancers: closing chromatin being WT-dominant enhancer regions that lose activity in the Nfatc2 overexpression, shared chromatin being enhancers that are relatively active in both states, and opening chromatin representing Nfatc2 overexpression-dominant enhancers that gain activity in the Nfatc2 overexpression. In order to test this hypothesis, we measured the expression of ncRNAs in each bin, as it has been shown that ncRNA abundance is a reliable metric for enhancer activity. We observed a clear trend in the mean expression of ncRNAs in each bin, where ncRNAs emanating from closing chromatin tended to be more highly expressed in the WT and ncRNAs emanating from opening chromatin tended to be more highly expressed in the Nfatc2-overexpression, while ncRNAs emanating from shared open chromatin displayed no significant difference in expression between WT and Nfatc2 overexpression (Figure 3.1E).

To further understand the effect of Nfatc2-overexpression on gene regulation, we wished to look specifically at enhancers that are directly regulated by Nfatc2. To this end, we subset ncRNA-defined enhancer regions binned by open chromatin by the presence or absence of a NFATC2 binding event within a 1000bp window around their 5' TSS (Figure 3.1F). These ncRNA-defined enhancer regions, marked with open chromatin and an NFATC2 binding event,

represent direct transcriptional targets of Nfatc2 that may play a role in establishing the gene regulatory network mediated by Nfatc2. Interestingly, the addition of NFATC2 binding shifted the relative proportion of enhancers in each bin to dramatically favor newly opening chromatin (green), going from approximately 18% of all ncRNA-defined open chromatin regions to 41% of all NFATC2-bound ncRNA-defined open chromatin regions (comparison of Figure 3.1D to 3.1F). This suggests that NFATC2 may be pioneering open chromatin events.

Similar to above, we examined the mean expression of NFATC2 bound ncRNAs in each bin of open chromatin (Figure 3.1G). In this case, we found that NFATC2 bound ncRNAs emanating from shared open chromatin are more highly expressed in the overexpression than in the WT, indicating that these shared putative enhancers are activated in response to NFATC2 binding. This is in direct contrast to the trend we find without direct NFATC2 binding, suggesting that the subset of shared enhancers bound by NFATC2 represent pre-existing regulatory elements that are co-opted by NFATC2, which may potentially play a role in the regulatory pathways underlying Nfatc2-mediated beta islet proliferation. In addition, we see that the difference between the mean expression of NFATC2-bound ncRNAs emanating from opening chromatin is much greater than the difference between the mean expression without NFATC2 binding, indicating that the strongest newly activated enhancers are directly bound by NFATC2. Taken together (Figure 3.1D to 3.1G), these data imply that Nfatc2 overexpression results in the activation of novel and pre-existing regulatory elements, as defined by newly opening or shared open chromatin respectively, and the activation is strongest when NFATC2 binds these regulatory elements directly. These identified direct Nfatc2-dependent enhancers may play a role in regulating critical genes involved in the proliferative response we observe as a result of Nfatc2 overexpression.

We observed that many of the direct NFATC2-bound ncRNAs lay proximal to genes that have known roles in beta cell proliferation and/or function, as well as are differentially expressed in response to Nfatc1 and Nfatc2 overexpression in human beta islets, supporting the hypothesis that ncRNA-defined NFATC2 bound regions are active enhancers that regulate the expression of these genes. We specifically asked whether any of the genes in the GO term "regulation of type B pancreatic cell proliferation" (GO: 0061469) were proximal to any of our identified Nfatc2-dependent ncRNA defined enhancers. We found that of the genes in this term, BIRC5 and NR4A1 displayed differential expression in response to Nfatc2 overexpression and had a proximal Nfatc2 dependent ncRNA (Figure 3.1H and 3.1J). Both of these ncRNAs are bound directly by NFATC2 and emanate from newly opening chromatin, the bin that displays the largest transcriptional effect by Nfatc2 overexpression (Figure 3.1G, right violin).

As shown in the genome browser view, the gene BIRC5 has a proximal Nfatc2-dependent ncRNA (MSTRG.178739, shown in green), that is directly bound by NFATC2 (shown in grey) (Figure 3.1H). BIRC5 is the only gene in this region that is differentially expressed (log2FC = 1.53) in response to Nfatc2 over-expression. The pictured ncRNA is highly upregulated in response to Nfatc2 overexpression (log2FC = 1.97), is bound by NFATC2 (grey peak), and emanates from opening chromatin (comparison of OE ATACseq (red) to WT ATACseq (blue)). This ncRNA shows specific upregulation in response to Nfatc2 overexpression, visualized in the bar chart (Figure 3.1I). The confluence of these factors: transcriptional upregulation of the ncRNA, a direct NFATC2 binding event, and the chromatin around the TSS of the ncRNA becoming more accessible in response to Nfatc2 overexpression, makes this locus a highly probable Nfatc2-dependent regulatory element that may regulate the expression of BIRC5. Interestingly, BIRC5 itself does not appear to be directly regulated by

Nfatc2, as indicated by the lack of NFATC2 binding events around its TSS or gene body. This indicates that the identified ncRNA-defined regulatory element may be responsible for the upregulation of BIRC5 we observe in response to Nfatc2 overexpression.

The gene NR4A1 also has a proximal Nfatc2-dependent ncRNA (MSTRG.93455) that is directly bound by NFATC2 (grey peak) (Figure 3.1J). NR4A1 is the most highly differentially expressed gene in this region (log2FC = 1.15) in response to Nfatc2 overexpression. The pictured ncRNA is highly upregulated in response to Nfatc2 overexpression (log2FC = 1.57), is bound by NFATC2 (gray peak), and emanates from newly opening chromatin (comparison of OE ATACseq (red) to WT ATACseq (blue)), which appears to be completely closed in the WT. This pattern of chromatin accessibility suggests that NFATC2 may be pioneering open chromatin at this locus. This ncRNA shows specific upregulation in response to Nfatc2 overexpression, visualized in the bar chart (Figure 3.1K). The confluence of these factors: transcriptional upregulation of the ncRNA, a direct NFATC2 binding event, and the presence of newly opening chromatin around the TSS of the ncRNA, makes this locus a highly probable Nfatc2-dependent regulatory element that may regulate the expression of NR4A1. In contrast to BIRC5, NR4A1 displays multiple NFATC2 binding events at its TSS and throughout the gene body.

Taken together, these results indicate that Nfatc2 overexpression results in global changes to the gene regulatory landscape of beta islet cells, as shown through the transcriptional changes of ncRNA expression and changes in chromatin accessibility. Utilizing both measures of open chromatin and ncRNA expression we are able to identify putative Nfatc2-dependent gene regulatory elements, and are able to understand nuance in their expression by binning their activity based on open chromatin. As ncRNA expression has been shown to be a reliable indicator of enhancer activity, we are able to interrogate changes in enhancer activity based on

direct NFATC2 binding events and gain insight on how Nfatc2 mediates its proliferative effect by activating pre-existing and novel regulatory elements. Two of the identified novel Nfatc2-dependent regulatory elements lie proximal to genes with annotated roles in beta islet proliferation: BIRC5 and NR4A1, and are prominent candidate enhancers for their regulation. As the enhancer element at NR4A1 is strongly upregulated in response to Nfatc2 overexpression, and the gene itself is a direct transcriptional target of Nfatc2, we decided to further interrogate the role of NR4A1 in Nfat-mediated β islet proliferation.

Figure 3.1: NFATC2 overexpression leads to global changes in ncRNA production and identifies putative enhancers regulating beta islet proliferation

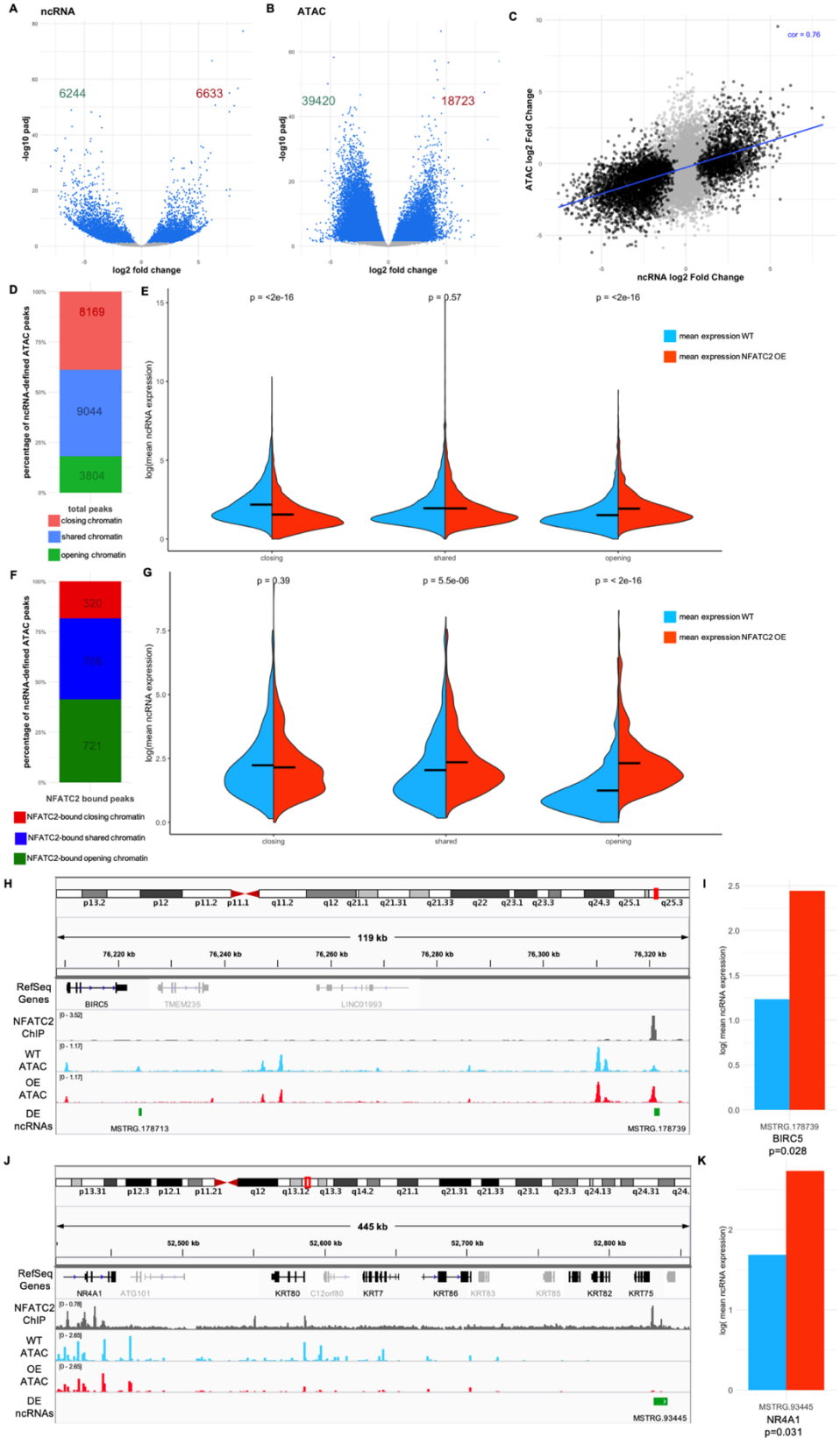


Figure 3.1: continued

A) Volcano-plot (log2 (fold-change) vs –log10 (Padj)) of WT vs NFATC2^{OE} non-coding RNAs (ncRNAs). Significantly misregulated ncRNAs (PAdj < 0.05 |log2 fold change| > 0.5) are depicted in blue and non-significant in grey. Number of significantly misregulated ncRNAs are shown in green for downregulated (6244) and red for upregulated (6633) respectively.
B) Volcano-plot (log2 (fold-change) vs—log10 (Padj)) of WT vs NFATC2^{OE} ATAC-seq peaks. Significantly differentially accessible ATAC peaks (PAdj < 0.05 |log2 fold change| > 0.5) are depicted in blue and non-significant in grey. Number of significantly differentially accessible peaks are shown in green for downregulated (39420) and red for upregulated (18723)
C) Scatterplot depicting correlation of ncRNA fold change with ATAC-seq fold change of WT vs NFATC2 OE samples. Non-significantly changed transcripts in gray. Best fit line in blue.
D) Proportion of ncRNAs emanating from different bins of open chromatin: closing chromatin (log2FC < -0.5) in red, shared chromatin (-0.5 < log2FC < 0.5) in blue, and newly opening chromatin (log2FC > 0.5) in green. The numbers of ncRNAs present in each bin is written in red, blue, and green for each bin of ATAC, respectively.
E) Violin plots depict the mean expression of ncRNAs in each bin of open chromatin: closing, shared, and opening, separated by genotype, with mean expression in WT in blue (left of violin) and mean expression in NFATC2 OE in red (right side of violin). P-values shown for t-test comparing the mean expression of ncRNAs in WT vs OE per bin.
F) Proportion of ncRNAs emanating from different bins of open chromatin that also contain an NFATC2 binding event. The numbers of ncRNAs bound by NFATC2 present in each bin is written in red, blue, and green for each bin of ATAC, respectively.
G) Violin plots depict the mean expression of NFATC2-bound ncRNAs in each bin of open chromatin separated by genotype. P-values shown for t-test comparing the mean expression of NFATC2-bound ncRNAs in the WT vs OE per bin.
H) Genome Browser view of the BIRC5 gene and Nfatc2-dependent ncRNA locus. Tracks are as follows from top to bottom: RefSeq genes (DE genes (|log2FC| > 0.5) shown in black, non-DE (|log2FC| < 0.5) genes in grey, NFATC2 ChIP signal (dark grey), wild-type open chromatin signal (blue), Nfatc2 overexpression open chromatin signal (red), significantly DE ncRNAs (green).
I) Bar plot depicts the expression of the ncRNA distal to BIRC5 (MSTRG.178739) in WT (blue) vs Nfatc2 overexpression (red). Adjusted p-value represents significance of this ncRNA being differentially expressed.
J) Genome Browser view of the NR4A1 gene and Nfatc2-dependent ncRNA locus. Tracks are as follows from top to bottom: RefSeq genes (DE genes (|log2FC| > 0.5) shown in black, non-DE (|log2FC| < 0.5) genes in grey, NFATC2 ChIP signal (dark grey), wild-type open chromatin signal (blue), Nfatc2 overexpression open chromatin signal (red), significantly DE ncRNAs (green).
K) Bar plot depicts the expression of the ncRNA distal to NR4A1 (MSTRG.93445) in WT (blue) vs Nfatc2 overexpression (red). Adjusted p-value represents significance of this ncRNA being differentially expressed.

NR4A1 has previously been shown to regulate β-cell proliferation in mouse and human islets[121] and in this work has been shown to be induced by NFATC2 in both contexts. As indicated above, this gene is directly regulated by NFATC2, as indicated by binding events at the NR4A1 TSS and the presence of an NFATC2-dependent enhancer element. These results suggest that NR4A1 is a key transcriptional target of NFAT that underlies the ability of NFATC2 to induce β-cell proliferation in mouse and human islets.

To experimentally validate whether NR4A1 is required for NFATC2-induced β-cell proliferation, our collaborators tested whether NFATC2 could induce proliferation in *Nr4a1* knockout ($Nr4a1^{KO}$) mice. They found that islets from $Nr4a1^{KO}$ mice had a significantly reduced proliferation in response to NFATC2 overexpression compared to WT islets (Simonett et al, Figure 4D). This finding suggests that NR4A1 is a key downstream activator of β cell proliferation that is directly regulated by NFATC2, making both key players in a GRN for β cell proliferation. To confirm the cell-type specificity of NFATC2 induced β cell proliferation, they also confirmed the major islet cell types stimulated to proliferate in WT and $Nr4a1^{KO}$ mice using a suite of immunostains for insulin, glucagon, and somatostatin; as well as the thymidine analog 5-ethynyl-2′-deoxyuridine (EdU) to measure cellular proliferation. Their results suggest that regardless of the presence of NR4A1, β cells are the primary cell type induced to proliferate in response to NFATC2 overexpression, highlighting that the relationship between NFATC2 and NR4A1 mediated proliferation is specific to β-cells.

Taken together, this work establishes NFATC2 as a potent regulator of β cell proliferation and insulin secretion in human islets. NFATC2 transcriptionally regulates a network of genes related to proliferation, and direct targets of NFATC2 reveal key players in the proliferative GRN such as NR4A1. NFATC2 appears to regulate gene expression through direct

promoter binding events as well as through non-coding mechanisms, including activating enhancer elements and potentially through modulating chromatin structure, as indicated by the changes in chromatin accessibility that occur with NFATC2 overexpression. This work also confirms the role of NR4A1 in β cell proliferation and confirms its transcriptional dependence on NFATC2, establishing a β cell proliferation GRN driven by NFATC2.

3.4 Discussion:

Previous studies, including those from our collaborators, have implicated NFAT as a potent β-cell mitogen in rodent and human islets [117]. Our collaborators have shown that overexpression of constitutively active forms of Nfatc1 and Nfatc2 are capable of driving human β-cells to proliferate. Using transcriptomic profiling, we were able to identify a small set of genes that are transcriptionally induced through NFATC2 overexpression, that are directly bound by NFATC2, and are potentially regulated by proximal NFATC2-dependent enhancer elements. The addition of ncRNA profiling specifically nominated two candidate genes downstream of NFATC2 that play a direct role in mediating the proliferative effect of NFATC2 overexpression. One of these genes, BIRC5 has been shown to be sufficient to regulate β-cell proliferation [122]–[124]. BIRC5 does not contain a TSS-proximal NFATC2 binding site, but with the integration of ChIPseq, ATACseq, and ncRNA profiling we were able to identify an NFATC2-dependent regulatory element upstream of the BIRC5 TSS that explains its transcriptional upregulation in response to NFATC2 overexpression. This NFATC2-dependent regulatory element actually lies within a previously identified islet-specific stretch enhancer [125], giving deeper understanding to our finding of an islet-specific NFATC2-responsive regulatory element that is likely to regulate BIRC5.

The other candidate gene identified by our analysis, NR4A1, has also been shown to regulate β-cell proliferation in mouse and human islets [121]. NR4A1 is transcriptionally induced by NFATC2 overexpression and also has a proximal NFATC2-dependent regulatory element marked by the presence of an upregulated ncRNA, open chromatin, and NFATC2 binding. To experimentally validate the connection between NFATC2, NR4A1, and β-cell proliferation, our collaborators asked if Nfatc2 can induce β-cell proliferation in islets from

Nr4a1 knockout (Nr4a1KO) mice, and found that from Nr4a1KO mice yielded significantly reduced β-cell proliferation in response to Nfatc2 compared to islets from WT mice [17]. This work highlights the role of Nr4a1 operating downstream of Nfatc2 in order to mediate β-cell proliferation in mouse and human islets.

The integration of chromatin binding, chromatin accessibility, and transcriptional profiling has allowed us to identify direct transcriptional targets of Nfatc2 that regulate β-cell proliferation in human islets. The use of non-coding RNA profiling in particular allows for the identification of differentially active regulatory elements that are responsive to NFATC2, allowing for the elucidation of an NFATC2-dependent gene regulatory network. This work demonstrates that Nfatc2 is a key transcription factor for genes that regulate β-cell proliferation and function, such as Birc5 and Nr4a1. Specifically, Nr4a1 is an Nfatc2 target gene that mediates much of its ability to stimulate β-cell proliferation. Finally, this work shows that Nfatc2 can act as a transcriptional enhancer to direct gene expression and affect target genes that mediate β-cell proliferation and T2D susceptibility.

3.5: Acknowledgement of Work Performed

I'd like to acknowledge the individuals who contributed to this work. The analyses presented in this chapter (Figure 3.1) are conceived, generated, and visualized by Linsin Smith. Carlos Perez-Cervantes performed original bioinformatic analysis of ncRNAseq, and downstream analysis was performed by Linsin Smith. Original bioinformatic analysis of coding RNAseq and ATACseq was performed by Sunyoung Shin with contributions from Linsin Smith and Carlos Perez-Cervantes. All experimentation was performed by individuals from Alan Attie's group. Written description of results presented in Figure 3.1 (page 41-47) is original to this dissertation and separate from published work, while the remainder of this chapter is summarized and paraphrased from the published work, available at 10.1172/JCI144833.

Chapter 4: Transcriptional profiling reveals rescue of inflammation, fibrosis, and hypertrophy in response to rebalancing calcium handling in atrial fibrillation

4.1 Introduction:

Heart function relies on coordinated electrical activity, which is controlled by distinct cell populations in the heart. The first step of the cardiac cycle begins with autonomous depolarization of cardiomyocytes in the sinus node, often termed the cardiac pacemaker. This spontaneous electrical activity propagates to the atria, where atrial cardiomyocytes depolarize, resulting in a coordinated atrial contraction. This contraction pushes blood from the atria to the ventricular chambers, which in turn contract and pump blood to the body through the aorta and the pulmonary arteries. The path of electrical current, originating in the pacemaker, moving through the atria, to the atrioventricular node (AV node), the Bundle of His and Purkinje fibers, and finally to the ventricles relies on the coordinated activity of ion handling proteins, which when altered result in cardiac arrhythmias.

One such arrhythmia is atrial fibrillation (AF). AF is the most common human arrhythmia and is characterized by irregularly irregular atrial electrical activity, resulting in asynchronous atrial contraction, that lead to a multitude of downstream effects. AF manifests as an irregularly irregular heart rate on surface ECG with the absence of a clear P wave, indicative of uncoordinated atrial contractions. The pathophysiology of AF is thought to consist of two arrhythmogenic components: an ectopic trigger in the form of ectopic depolarization events and a vulnerable substrate that abnormally propagates these triggers throughout the atria [126].

Normal cardiac conduction relies on the appropriate path of electrical signals through the heart. Cardiomyocytes are capable of holding a resting membrane potential until an electrical

threshold is reached, firing an action potential. When threshold is reached, the action potential begins with rapid cellular depolarization, mediated by voltage-gated sodium channels. This depolarization leads to calcium release from the sarcoplasmic reticulum (SR) through the ryanodine receptor (encoded by *Ryr2*). This released calcium binds the contractile apparatus of the cardiomyocyte, causing contraction of the cell. Around 80% of released calcium is taken back into the SR by the sarco/endoplasmic reticulum calcium-ATPase SERCA2 (encoded by *Atp2a2*). After depolarization and contraction, the cardiomyocyte is repolarized by potassium channels, enters a refractory period, and is primed for the next action potential.

In the healthy heart, the only source of atrial depolarizations is the sinus node; however in AF there are ectopic depolarizations forming the triggering events in AF, contributing to electrical remodeling. Similarly, in the healthy heart, cardiomyocytes enter a refractory period after firing, which limits re-entrant electrical activity. In AF, the refractory period can be affected, leading to a vulnerable fibrillogenic substrate that is able to inappropriately propagate triggered depolarizations through the normally refractory myocardium. The formation of a vulnerable substrate, or structural remodeling, is thought to be induced by atrial enlargement or hypertrophy, atrial fibrosis, as well as atrial inflammation: all of which delay electrical propagation, allowing time for refractory cells to recover and fire abnormally [126]–[128]. Interestingly, the structural atrial remodeling induced by inflammation and fibrosis can also impact electrical cardiomyocyte function, leading to enhanced electrical remodeling [127]. While the precise nature of the relationship between AF, inflammation, and fibrosis is yet to be fully understood; it is clear that AF can induce inflammation and fibrosis, which can also induce and enhance AF, leading to a vicious cycle. Understanding the interplay between AF, inflammation,

and fibrosis is critical for our understanding of the pathophysiology of AF, and also to motivate novel therapeutic strategies to treat AF and its underlying causes.

AF represents a large burden on healthcare worldwide, affecting over 33 million people and approximately 25% of Americans over the age of forty [129], [130]. AF is associated with significant morbidity and mortality due to thromboembolic events, sudden cardiac death, and heart failure; and patients with atrial fibrillation had a 3.7-fold increased risk of all-cause death compared with the general population [131]. Numerous risk factors have been identified for AF, including age and heart failure, however these factors have little capacity to predict the onset of AF. This necessitates further research into the mechanisms of AF and AF susceptibility in order to provide improved diagnostic, preventative, and therapeutic strategies for AF treatment.

AF has a large genetic component, with estimates of heritability of AF ranging between 40%-60% [132], [133]. Interestingly, most AF presents in the context of concomitant heart disease, such as heart failure, however the high heritability of AF and genetic predispositions identified by genome wide association studies (GWAS) underscore a strong genetic component separable from other cardiac diseases. While most AF presents in the context of pre-existing cardiac diseases, “lone” AF can also manifest [4], [134]. Utilizing family studies on patients with lone AF, it has been shown that rare variants in ion channel subunits, gap junction proteins, and cardiac transcription factors can all play a causal role [4], [5].

GWAS have associated over 140 genetic variants to AF progression and risk, nominating hundreds of genes with known and unknown roles in cardiac biology [135], [136]. Unsurprisingly, many of these associations are for variants in critical ion channels and structural cardiac genes. However, GWAS also implicate cardiac transcription factors, which indicates a strong gene regulatory component contributes to AF. It is known that increased or decreased ion

channel expression can cause cardiac arrhythmias, highlighting the role of transcriptional regulators of ion channels in AF genetics [7], [126], [135], [137]. One such transcription factor is *TBX5*, which has been associated to AF susceptibility and PR interval prolongation through GWAS, and dominant mutations in Tbx5 are known to cause Holt-Oram Syndrome, characterized by limb and heart defects and an increased risk of AF [138]–[141]. Deletion of *TBX5* from the adult mouse heart causes primary spontaneous atrial fibrillation as well as the transcriptional dysregulation of multiple calcium handling genes including *Atp2a2, Ryr2,* and *Pln* [7].

Tbx5 null mice display all the hallmarks of AF: irregularly irregular heart rate, disorganized atrial activity, the absence of a stereotypical P-wave, AP prolongation, and abnormal cellular calcium handling [7]. The action-potential abnormalities present in *Tbx5* null mice were able to be rescued by chelating free calcium, highlighting the role of calcium handling in the molecular mechanism of *Tbx5*-dependent AF [7]. To elucidate the role of abnormal calcium handling in the mechanism of *Tbx5*-dependent AF, an analysis of the differentially expressed genes in response to *Tbx5* deletion was performed. Numerous genes critical to atrial rhythm control and around half of AF-associated genes were downregulated in the Tbx5 null left atrium [7]. Specifically, there was significant downregulation of genes required for appropriate calcium handling in the atrium, namely *Ryr2, Atp2a2,* and *Sln.* Taken together, these results nominate aberrant calcium handling as a considerable contribution to *Tbx5* induced AF. Interestingly, out of the group of dysregulated calcium handling genes, one gene *Pln,* was upregulated in response to *Tbx5* deletion. *Pln,* also known as phospholamban, is a negative regulator of the sarco/endoplasmic reticulum calcium-ATPase SERCA2, encoded by *Atp2a2*.

Since Tbx5 deletion decreases the expression of *Atp2a2,* leading to decreased SERCA function and abnormal calcium handling in atrial cardiomyocytes, while simultaneously increasing the expression of *Pln,* a negative regulator of SERCA, it was hypothesized that reducing *Pln* levels would normalize SERCA function and potentially rescue calcium handling in the *Tbx5* mutant atrium [137], [142]. Interestingly, it was found that *Pln* deletion on its own resulted in few calcium handling defects, with increased SR load and SERCA activity compared to controls, and normal AP duration [142]. However, *Pln* deletion in the *Tbx5* mutant resulted in rescued AP duration and rescued AF susceptibility as measured by intracardiac burst pacing [142]. While control animals and *Pln* single null animals were not able to be paced into AF, all of the *Tbx5* single null animals were able to be paced into AF. The double knockout (DKO) animals showed significantly reduced AF susceptibility, with only 1 out of the 11 DKO animals being able to be paced into AF. These results show that *Tbx5*-deficiency induced AF is due to calcium handling abnormalities, specifically decreased SR load and SERCA activity, and that modulation of the SERCA2 inhibitor, *Pln*, normalized SERCA activity and AF susceptibility caused by *Tbx5* loss.

With the understanding that normalization of calcium handling due to *Pln* deletion rescues *Tbx5*-dependent AF, we next wished to understand the transcriptional consequences of AF rescue. In particular, we were curious to understand how rescue of calcium handling would impact the fibrotic and inflammatory response typically seen in AF. While it is known that inflammation and fibrosis contribute to AF by providing a vulnerable substrate, as well as by directly affecting atrial cardiomyocyte electrophysiology, it is unknown how aberrant calcium handling leads to or stems from these disease processes. The experimental system of *Tbx5*

induced AF and rescue through *Pln* deletion provides an appropriate platform to interrogate the effect of calcium mishandling and rescue on the activation of inflammation and fibrosis.

To understand the relationship between AF, inflammation, and fibrosis, we performed transcriptional profiling in control, *Tbx5* single null, *Pln* single null, and DKO animals. We show that rescue of *Tbx5*-dependent AF through deletion of *Pln* also rescues the transcriptional activation of the fibrotic and inflammatory response prior to the onset of AF. We find that there is a high degree of transcriptional similarity between the *Tbx5* single null and the DKO, and that in particular, critical ion handling genes are concordantly dysregulated in both contexts. This suggests that even though canonical cardiac rhythm genes are downregulated in the DKO, the rebalancing of calcium handling through Pln deletion is enough to compensate for their lower expression levels. Differentially expressed genes between the DKO and the *Tbx5* single null nominate candidate genes and pathways involved in mediating rescue, as well as in pathogenic structural atrial remodeling. In addition to the transcriptional rescue of fibrosis and inflammation, *Pln* deletion in the *Tbx5* null mouse also rescued development of atrial hypertrophy associated with AF. Based on our findings that rescuing calcium handling also rescued the fibrotic and inflammatory response, we also wished to understand whether modulating inflammation could also rescue AF and its associated calcium handling defects. Surprisingly, modulating inflammation through IL6 deletion enhanced and worsened *Tbx5*-induced AF, while exhibiting no phenotype on its own. This work provides a starting point for the elucidation of the relationship between fibrosis, inflammation, and AF.

4.2 Methods:

Animals

The $Tbx5^{fl/fl}$, $Pln^{-/-}$ and $Rosa26^{CreERt2}$ lines have all been previously described and were kept in a mixed genetic background [139], [143], [144]. Double knockout mice were generated by crossing $Tbx5^{fl/fl}$;$R26^{CreERt2}$ mice with germline $Pln^{-/-}$ mice. All experiments were done using age- and genetic strain-matched littermate controls. $R26^{CreERt2}$, $Tbx5^{fl/fl}$;$R26^{CreERt2}$, $Pln^{-/-}$;$R26^{CreERt2}$, and $Tbx5^{fl/fl}$; $Pln^{-/-}$;$R26^{CreERt2}$ animals were submitted to a tamoxifen regimen of 100 uL 20mg/mL tamoxifen over 3 days at 6-8 weeks of age. All experiments were performed in accordance to The University of Chicago Institutional Animal Care and Use Committee (IACUC) approved protocol.

The $IL6^{-/-}$ line has been previously described [145]. Mice were generated by crossing $Tbx5^{fl/fl}$;$R26^{CreERt2}$ mice with germline $IL6^{-/-}$ for multiple generations. Mice were maintained in a double homozygous background $Tbx5^{fl/+}$; $IL6^{-/+}$;$R26^{CreERt2}$ to generate genetic strain-matched littermate controls. All experiments were done using age- and litter-matched controls. $R26^{CreERt2}$, $Tbx5^{fl/fl}$;$R26^{CreERt2}$, $IL6^{-/-}$;$R26^{CreERt2}$, and $Tbx5^{fl/fl}$; $IL6^{-/-}$;$R26^{CreERt2}$ animals were submitted to a tamoxifen regimen of 100 uL 20mg/mL tamoxifen over 3 days at 6-10 weeks of age. All experiments were performed in accordance to The University of Chicago Institutional Animal Care and Use Committee (IACUC) approved protocol.

Coding RNA-Seq library preparation and Data analysis

Total RNA was extracted from left atrial tissue from $R26^{CreERt2}$, $Tbx5^{fl/fl}$;$R26^{CreERt2}$, $Pln^{-/-}$;$R26^{CreERt2}$, and $Tbx5^{fl/fl}$; $Pln^{-/-}$;$R26^{CreERt2}$ animals using the Macherey-Nagel Nucleospin DNA, RNA, and protein extraction kit. RNA was quantified and 350 ng RNA / sample were used for

library generation. RNA was submitted to PolyA selection using the Lexogen CORALL PolyA selection module, and only PolyA+ RNA was used for library preparation. Libraries were prepared with the Lexogen CORALL Total RNA Library kit and sequenced on the Illumina NovaSeq SP flowcell with 50bp paired-end reads.

Paired-end RNA-seq with 10M to 20M reads were mapped to mouse genome with STAR (v 2.5.3a). Reads with MAPQ score < 30 were removed and sorted by position with samtools/1.9. Counts were retrieved with HTseq (v.0.11.2) UCSC known genes annotation. Lastly, counts were analyzed for differential expression with R (4.0.3) package DEseq [66].

GO enrichment analysis

GO enrichment analysis was performed using Metascape (https://metascape.org) using the single gene list and multiple gene list functions [146].

Motif scanning

Known and de novo promoter motif scanning was performed using HOMER (4.3) on the promoters of differentially expressed genes . Promoter regions were defined as 1000bp upstream of the TSS and 50bp downstream of the TSS.

Imaging and Histology

Hearts were dissected from animals in ice cold sterile 1xPBS, imaged, and then fixed in a 4% formaldehyde solution for 72 hours. After 72 hours, hearts were washed 3 times with ice cold sterile PBS and then bisected for embedding. After bisection, hearts were washed 3 times with ice cold 70% ethanol. Histology was performed by the University of Chicago Human Tissue

Resource Center. Formaldehyde fixed tissues were embedded in paraffin wax and sectioned to 5uM thickness as a four chamber view. Tissue was stained with H&E and trichrome.

Cardiomyocyte widths were measured using ImageJ. For quantification, five cardiomyocyte widths were measured per mouse then averaged to compute average cardiomyocyte width and standard deviation (control, Tbx5 single null, DKO n=2, Pln single null n=1). For Pln single null, two images from separate sections were used to compute average cardiomyocyte width and standard deviation. Significance was calculated using two tailed paired t-test.

Electrophysiology

Control, Tbx5 single null, IL6 single null, and DKO animals were treated with tamoxifen at 6-10 weeks of age as previously described [7], [142]. Animals were anesthetized with inhaled isoflurane (~2%) delivered via nose cone. Body temperature was maintained with a heated platform and surface ECG readings were recorded for 3 minutes using LabChart software (ADInstruments) with subcutaneous needle electrodes placed just under the skin. P wave duration, PR interval, QRS duration, and RR interval were calculated at 3 time points (30s, 1m, 2m) and averaged per mouse.

4.3 Results:

Rescue of atrial fibrillation due to *Pln* deletion also rescues transcriptional immune and fibrotic response

In order to assay the transcriptional changes in the context of AF and rescue, we performed RNA profiling on litter matched control, *Tbx5* single null, *Pln* single null, and double knockout (DKO) animals 2 weeks post-tamoxifen injection, at the onset of AF susceptibility. We saw a clear separation by genotype using principle component analysis, with control and *Pln* single null animals clustering together and *Tbx5* single null and DKO animals clustering in distinct, but nearby groups (Supplemental Figure 4.1A). We next performed differential expression testing between each genotype and the control (Figure 4.1 A-C). In the comparison between the *Tbx5* single null and the control we identified 971 differentially expressed genes ($|log2FC| > 0.5$, $padj < 0.05$, Figure 4.1A center) that share high concordance with previously identified *Tbx5*-dependent transcripts in other contexts (Supplemental Fig 4.1B, C, D). These 971 DE genes can be split into upregulated and downregulated genes, which in turn identifies subnetworks of *Tbx5*-dependent genes. As shown in previous work, downregulated genes are highly enriched for pathways related to cardiac function, such as "regulation of ion transmembrane transport" and "cell communication in cardiac conduction" (Figure 4.1A, left). Upregulated genes are enriched for disease-associated terms related to fibrosis and inflammation, such as "extracellular matrix organization", "leukocyte cell-cell adhesion", "regulation of cytokine production", and "inflammatory response" (Figure 4.1A, right).

We next compared our DKO animals to the control and identified 445 differentially expressed genes, which can be split into 230 downregulated genes and 215 upregulated genes respectively ($|log2FC| > 0.5$, $padj < 0.05$, Figure 4.1B, center). Similar to the case with *Tbx5*

single null, we saw an enrichment of terms related to calcium handling and cardiac function on the downregulated side, including “regulation of ion transport”, “muscle contraction”, and “regulation of atrial muscle cell membrane depolarization”. However, on the upregulated side, we no longer see the global enrichment of terms related to fibrosis or inflammation, and instead see terms more related to cardiac function, such as “cardiac conduction” (Figure 4.1B, right). This indicates that in the DKO context where we have rescued atrial fibrillation through the deletion of *Pln* that we have also rescued the transcriptional fibrotic and inflammatory response.

Interestingly, we found that in the comparison between the *Pln* single null and the control, the only differentially expressed gene is *Pln* itself, which is significantly downregulated in the *Pln* null as expected (log2FC = -3.5, padj = 1.635706e-42, Figure 4.1C). While this finding is not unexpected, considering that *Pln* is not a transcription factor and is not known to regulate gene expression either directly or indirectly, it highlights the lack of a feedback mechanism from the level of *Pln* function back to gene transcription. Additionally, it allows for much deeper understanding of the relationship between the DKO and the *Tbx5* single, as we can now attribute all transcriptional changes in the DKO to the combinatorial deletion of *Tbx5* and *Pln,* as opposed to baseline changes induced by deleting *Pln* alone.

We next wished to understand the similarities and differences between the Tbx5 single null and the DKO. As a first step we computed the correlation between gene expression in each compared to the control. Globally, gene expression is highly correlated between the *Tbx5* single null vs WT and the DKO vs WT, with all identified genes displaying a correlation of 0.72, genes significant in one comparison but not the other displaying a correlation of 0.88, and genes significantly DE in both displaying a correlation of 0.975 (Figure 4.1D). These results indicate that gene expression changes in the *Tbx5* vs WT and the DKO vs WT are highly similar, even

though the two genotypes display extremely different phenotypic outcomes. To get a closer look at the function of the differentially expressed genes in each condition, we performed comparative GO analysis between the 971 *Tbx5*-dependent DE genes and the 445 DKO-dependent DE genes (Figure 4.1E). Both gene sets are highly enriched for pathways related to cardiac function, as indicated earlier in the individual GO analyses. However this analysis clearly shows that the *Tbx5*-dependent genes are enriched for terms related to disease while the DKO-dependent genes are not enriched for these pathways at all, or to a very small extent.

Taken together the transcriptomic profiling shows that rescuing Ca2+ handling through the deletion of *Pln* also rescues the transcriptional fibrotic and immune response that is normally activated in *Tbx5* single null AF animals. In addition, the finding that *Pln* deletion results in no other transcriptional changes allows us to attribute the differences in gene expression that we see between the *Tbx5* single null and the DKO solely to the DKO context, as opposed to gene expression changes that result from *Pln* single deletion. Further, we see these transcriptional changes at the onset of AF susceptibility, prior to the onset of spontaneous AF itself. This indicates that the transcriptional activation of inflammation and fibrosis in AF occurs much earlier than previously expected, as at this time point there are no phenotypic signs of inflammation or fibrosis in the *Tbx*-dependent AF animals. Together, these findings showcase that *Pln* deletion is an extremely precise way to re-balance calcium homeostasis, resulting in AF rescue and the transcriptional rescue of critical disease processes such as inflammation and fibrosis, prior to the onset of AF.

Figure 4.1: Rescue of atrial fibrillation through Pln deletion also rescues transcriptional immune and fibrotic response

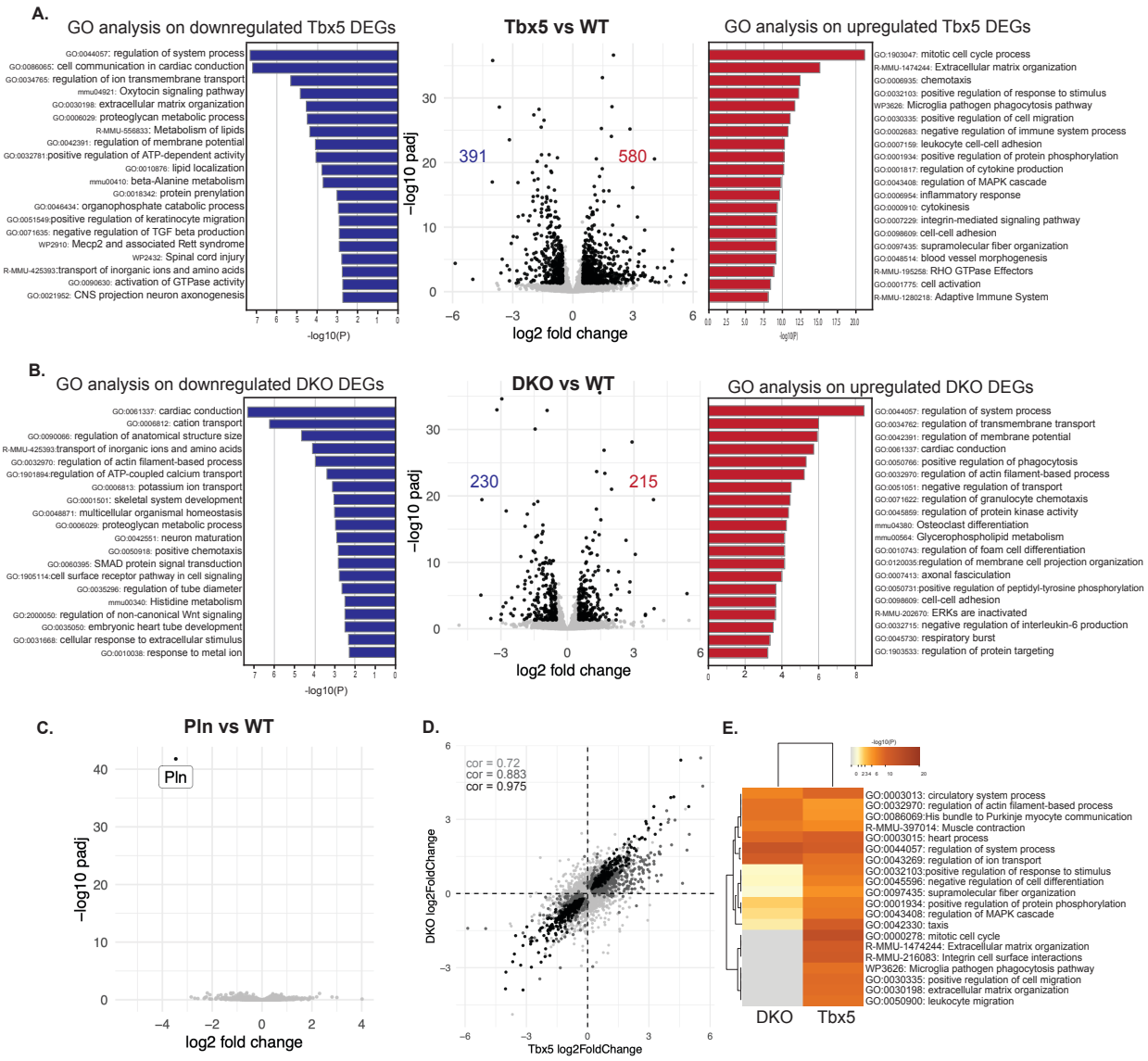


Figure 4.1: Legend

A. **Center:** Volcano plot (log2FoldChange versus -log10(adjusted pvalue)) of Tbx5 single null vs WT coding transcripts. Significantly misregulated genes (|log2FoldChange| > 0.5 and padj <. 0.05) shown in black and non-significant in grey. **Left:** GO term analysis on downregulated genes between Tbx5 single null and WT sorted by -log10(pvalue). **Right:** GO term analysis on upregulated genes between Tbx5 single null and WT sorted by -log10(pvalue).

B. **Center:** Volcano plot (log2FoldChange versus -log10(adjusted pvalue)) of DKO vs WT coding transcripts. Significantly misregulated genes (|log2FoldChange| > 0.5 and padj <. 0.05) shown in black and non-significant in grey. **Left:** GO term analysis on

Figure 4.1: Legend continued

downregulated genes between DKO and WT sorted by -log10(pvalue). **Right:** GO term analysis on upregulated genes between DKO and WT sorted by -log10(pvalue).

C. Volcano plot (log2FoldChange versus -log10(adjusted pvalue)) of Pln single null vs WT coding transcripts. Significantly misregulated genes (|log2FoldChange| > 0.5 and padj <. 0.05) shown in black and non-significant in grey.
D. Scatterplot for genes identified in the Tbx5 single null vs WT and DKO vs WT: log2FoldChange for Tbx5 vs WT on the x-axis and log2FoldChange for DKO vs WT on the y-axis. Light grey dots represent all identified genes, medium grey dots represent genes significantly differentially expressed in at least one comparison, and black dots represent genes significantly differentially expressed in both comparisons. Pearson's correlations for each group (all genes – light grey, significant in one – medium grey, and significant in both - black) are shown in the upper left corner.
E. Comparative GO analysis on DE genes identified between Tbx5 and WT (left) and DKO and WT (right). Heatmap of enriched terms colored by -log10(pvalue).

Pln deletion rebalances calcium homeostasis without transcriptional rescue of critical calcium handling components

We next interrogated the similarities between the *Tbx5* single null and the DKO. We identified 347 DE genes that are shared between the *Tbx5* single null and the DKO state (|log2FC| > 0.5, padj < 0.05 in both compared to WT). Interestingly, we see that all of these 347 genes are concordantly expressed in each state, meaning there are no cases in which one of these shared genes is upregulated in the *Tbx5* single null but downregulated in the DKO, or vice versa (Figure 4.2A). We next interrogated the function of these shared genes by performing GO analysis on shared downregulated genes as well as shared upregulated genes (Figure 4.2B, 4.2D respectively). For the shared downregulated genes we see a strong enrichment for terms related to calcium handling and cardiac conduction, such as "cell communication involved in cardiac conduction" and "regulation of ion transport". The shared upregulated genes are also enriched for cardiac related terms, such as "cardiac conduction", but we also see enrichment of terms

related to ERK1 and ERK2 signaling as well as “negative regulation of interleukin 6 production”.

To understand the potential regulators of these shared genes we performed known motif analysis on the promoters of these genes. For the downregulated genes we see enrichment of motifs of canonical cardiac transcription factors such as MEF2 and GATA, while the upregulated genes show an enrichment of motifs for STAT, zinc finger, and homeobox motifs (Figure 4.2C, 4.2E respectively). These findings highlight a shared group of cardiac genes that are transcriptionally concordant in both the *Tbx5* single null and the DKO. Interestingly, we find that for the downregulated genes, many of these include canonical AF-associated genes, such as the ion channels *Ryr2*, *Scn5a*, and *Scn10a* as well as calcium-handling genes such as *Sln*, and gap junction proteins *Gja1* and *Gja5* (Figure 4.2F). These findings were surprising, as it is well known that downregulation of ion channels can cause AF. This indicates that even though the DKO still has dysregulated expression of critical ion handling components, the deletion of *Pln* is sufficient to functionally rescue AF. There is also a confluence of critical cardiac genes that are upregulated in both states, including troponins, cardiac myosin, and calmodulin (Figure 4.2F).

While the downregulated genes highlight a clear cardiac transcriptional network, as they are enriched for cardiac-related terms as well as cardiogenic transcription factor motifs, the upregulated genes were also of interest. While many of these genes are certainly related to cardiac function, we also see an enrichment of terms related to the potential negative regulation of disease processes, such as “negative regulation of ERK1 and ERK2 cascade”, “negative regulation of IL6 production”, and “negative regulation of cell adhesion”. We were particularly curious about the enrichment of the STAT3 and STAT4 motifs at the promoters of these shared upregulated genes. The Stat family of transcription factors are capable of integrating inputs from

multiple signaling pathways through their interaction with Janus kinases (JAK), and in particular are known to play a critical role in responding to cytokines [147], [148]. The enrichment of the STAT3 motif is of particular interest, as STAT3 has known cardioprotective roles in the heart, in addition to being a downstream signaling molecule for a variety of pro- and anti-inflammatory cytokines, including IL6 [149], [150]. We posit that this group of shared upregulated genes includes genes related to cardiac function, such as Kcnb2, Tnni1, Myl3, and Camk1, but also pathways related to the cardiac stress response, as indicated by STAT motifs and GO analysis.

Taken together, these results indicate that even though the DKO displays phenotypic rescue of AF, critical AF-associated genes are misregulated in these animals. This highlights the precise nature of the *Pln* deletion as functionally rescuing AF on the level of ion channel function, as opposed to ion channel transcription. Further investigation of the shared upregulated genes may nominate novel cardiac stress response pathways that are transcriptionally activated in both models of AF and AF rescue.

Figure 4.2: *Pln* deletion rebalances calcium homeostasis without transcriptional rescue of critical calcium handling components and cardiac genes

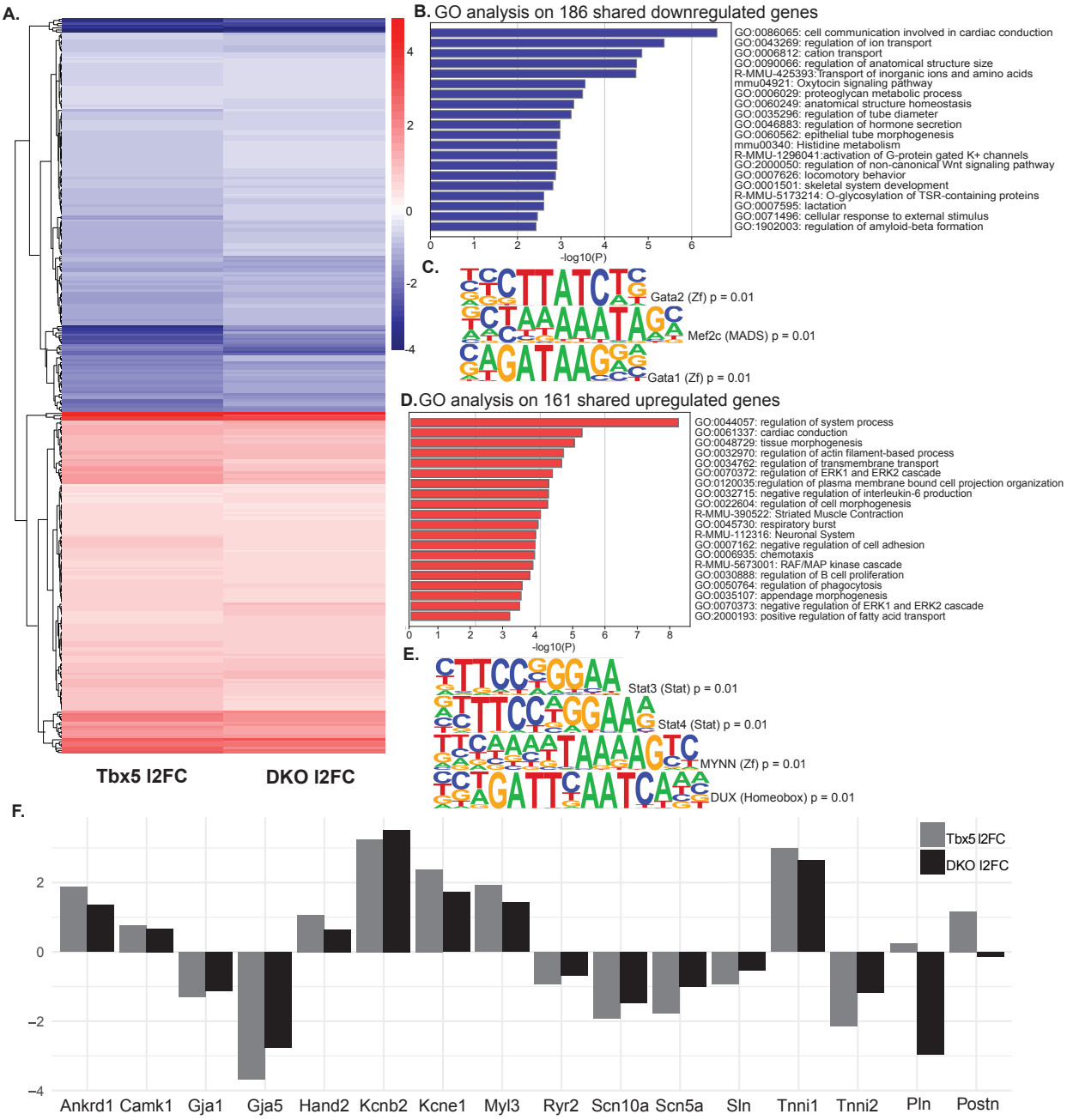


Figure 4.2: Legend:

A. Heatmap of log2FoldChange for 347 shared differentially expressed genes between Tbx5 vs WT (left column) and DKO vs WT (right column) colored by log2FoldChange.
B. GO analysis on downregulated shared genes, sorted by -log10(pvalue).
C. Known motif enrichment analysis on downregulated shared gene promoters. Shows enriched motif position weight matrix and pvalue.
D. GO analysis on upregulated shared genes, sorted by -log10(pvalue).

Figure 4.2: Legend, continued:

E. Known motif enrichment analysis on upregulated shared gene promoters. Shows enriched motif position weight matrix and pvalue.
F. Log2FoldChange of selected shared differentially expressed genes in Tbx5 single v WT (grey) and DKO v WT (black), compared to two examples of discordant genes: Pln and Postn (right).

Differentially expressed genes between *Tbx5* single null and DKO reveal groups of genes involved in disease progression and rescue

We next wished to understand the transcriptional differences between the DKO and *Tbx5* single null, as these may identify genes that are rescued between the *Tbx5* single and the DKO, as well as highlight genes involved in disease progression (DE in Tbx5 single v WT but not in DKO v WT) and involved in rescue (DE in DKO v WT but not in Tbx5 single v WT). To explore this, we performed a direct differential expression test between the DKO and the *Tbx5* single knockout, identifying 89 differentially expressed genes between these two states at an adjusted p-value cutoff of 0.1, and 176 DE genes at 0.2 ($|log2FC| > 0.5$, $padj < 0.1$ (black), 0.2 (medium grey), Figure 4.3A). These DE genes can be split into upregulated and downregulated categories, with downregulated genes being those which are more highly expressed in the *Tbx5* single null than the DKO, and upregulated genes being those which are more highly expressed in the DKO than the *Tbx5* single null. We hypothesized that genes involved in promoting disease processes such as fibrosis and inflammation would be downregulated in the DKO compared to the *Tbx5* single, and genes involved in promoting rescue would be upregulated. The 89 significant DE genes ($padj < 0.1$) are highly enriched for GO terms related to cell cycle, and also

display some disease-associated terms such as “lung fibrosis”, “ECM organization”, and “response to chemokine” (Figure 4.3B).

Based on our hypothesis that directly DE genes between the DKO and the *Tbx5* single may potentially play a role in rescuing AF and its associated phenotypes, such as fibrosis and inflammation; identifying the regulatory relationships between these candidate genes, *Tbx5*, and other transcriptional regulators of great interest. In order to nominate possible regulators of directly DE genes, we performed known motif analysis on the promoter regions of all 89 DE genes between the DKO and the Tbx5 single (Figure 4.3C). Interestingly, we see an enrichment for two T-box motif transcription factors, Eomes and Brachyury, both of which are important developmental transcription factors. Both TFs play a role in cardiac development: Brachyury plays a critical role in cardiac development and left-right patterning, and Eomes is required to develop functional cardiomyocytes [151], [152]. The enrichment of developmental TF motifs is interesting, as it suggests that genes involved in AF progression and rescue may be components of a more developmental gene regulatory network. Alternatively, both motifs also contain similarity to the Tbx5 binding motif, suggesting that perhaps these TFs are enriched due to the similarity between T-box binding motifs. We also see an enrichment for the CTF motif, most commonly seen as CTCF binding. Similar to our results looking at upregulated shared genes between the DKO and Tbx5 single null, we also see a strong enrichment of the Stat motif, implicating the potential role of cytokine signaling in the regulation of these genes.

To gain a deeper understanding of how the directly DE genes between the DKO and the *Tbx5* single null are regulated in comparison to WT, we plotted the log2FoldChange of each of the 89 direct DE genes (right column) against their log2FoldChange between Tbx5 and WT (left column) and the log2FoldChange between DKO and WT (middle column) (Figure 4.3D). This

approach allows for the clear visualization of many transcriptional patterns across these data. One pattern of expression includes genes that are significantly DE between the DKO and *Tbx5* single, but in relation to WT are concordantly expressed. For example, *Cep55*, a gene that plays a significant role in cell cycle exit and cytokinesis, is significantly downregulated in the comparison between the DKO and the *Tbx5* single; however in each condition compared to WT *Cep55* is upregulated. This indicates that while *Cep55* is upregulated in both the DKO and Tbx5 single compared to WT, it is much more strongly upregulated in the *Tbx5* single than it is in the DKO, leading to a negative log2FoldChange in the comparison between DKO and *Tbx5* single. The converse pattern also exists, where a gene such as *Gm1078* is significantly upregulated between the DKO and *Tbx5* single null, but is concordantly downregulated in each condition compared to WT. While these more subtle changes in gene expression are interesting, in order to identify genes involved in disease progression and rescue, we turned to genes displaying discordant patterns of expression across the three comparisons.

Genes that display discordant transcriptional patterns, meaning they are identified as differentially expressed between DKO and *Tbx5* single, and are regulated in opposite directions when compared to WT, represent interesting candidates for functional follow up. One example is *Pnmt*, which is significantly upregulated in the comparison between DKO and *Tbx5* single null, upregulated in the comparison between DKO and WT, but slightly downregulated in the comparison between the *Tbx5* single null and WT. The *Pnmt* gene encodes phenylethanolamine N-methyltransferase which is an important enzyme that synthesizes epinephrine from norepinephrine [153]. Interestingly, levels of *Pnmt* are highest in the left atrium and has been shown to be required for normal blood pressure and cardiac responses during stress [154]. Furthermore, it has been shown that *Pnmt* is expressed in cardiac cells during development in

regions of the heart associated with pace making and conduction, such as the SA node and the atrioventricular canal [155], [156]. Based on the expression pattern of *Pnmt*, it is plausible that this gene is specifically upregulated in the rescued state (DKO v WT and DKO v *Tbx5* single), but is not differentially expressed in AF, nominating *Pnmt* as an interesting candidate for being involved in AF rescue.

Another example of this discordant expression pattern is *Atf3*, which is highly downregulated in the comparison between DKO and *Tbx5* single null, but is upregulated in the *Tbx5* single v WT and is downregulated in the DKO v WT. *Atf3* is a stress-induced transcription factor that can act as both a transcriptional activator and a repressor. In the heart, expression of *Atf3* is induced in response to cardiac stress, in both cardiomyocytes as well as in activated cardiac fibroblasts [157], [158]. Cardiac specific *Atf3* overexpression causes cardiac hypertrophy, increased fibrosis, a reduction of ventricular contractility, and slight conduction abnormalities [158]. Deletion of *Atf3* in cardiomyocytes subject to transverse aortic constriction, a method to induce heart failure, led to decreased levels of cardiac hypertrophy and fibrosis [157]. Taken together, the literature suggests a role for *Atf3* in the cardiac stress response, and suggests that downregulation of *Atf3* is cardioprotective.

In our data, *Atf3* is significantly differentially expressed between the DKO and the *Tbx5* single null, with a log2FoldChange of -2.99, and trends upregulated in the *Tbx5* single vs WT (log2FC 1.7, padj 0.24), and trends downregulated in the DKO vs WT (log2FC -1.29, padj 0.53). It is of interest that this gene is only found as significantly differentially expressed between the DKO and the *Tbx5* single, as it would not have been identified as a gene of interest in our previous analyses. Furthermore, in a *Tbx5*-dependent scRNAseq experiment (data not shown), *Atf3* is identified as significantly upregulated 3 days post tamoxifen in cardiomyocytes (average

log2FC = 1.44, padj = 8.758854e-91) and in the epicardium (average log2FC = 0.55, padj = 4.611575e-21). Based on these results, it is also highly possible that we only recapture significant *Atf3* differential expression due to the lack of cell-resolved data when performing bulk experiments compared to WT, that the similarities between the DKO and the *Tbx5* single allow us to interrogate.

Figure 4.3: Differentially expressed genes between DKO and Tbx5 single null reveal groups of genes involved in disease progression and rescue

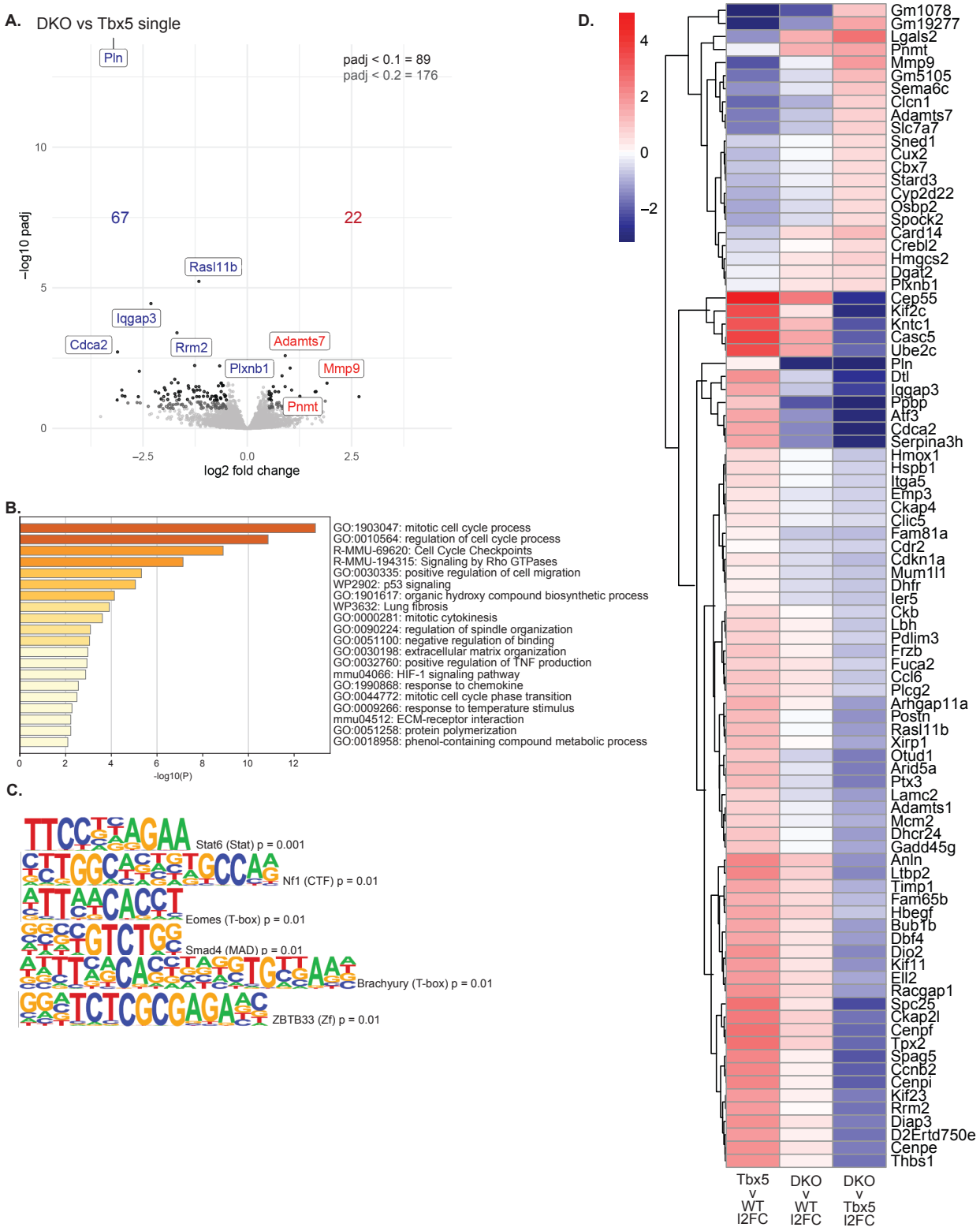


Figure 4.3 continued:

A. Volcano plot (log2FoldChange versus -log10(adjusted pvalue)) of Tbx5 single null vs DKO coding transcripts. Significantly misregulated genes (|log2FoldChange| > 0.5 and padj <. 0.1) shown in black and medium gray (|log2FoldChange| > 0.5 and padj < 0.2), non-significant in light grey.
B. GO analysis on significantly differentially expressed genes between Tbx5 single null and DKO, sorted by -log10(pvalue).
C. Known motif enrichment analysis on differentially expressed gene promoters. Shows enriched motif position weight matrix and pvalue.
D. Heatmap of Log2FoldChange for Tbx5 single null vs WT (left), DKO vs WT (middle), and DKO vs Tbx5 single (left) for the 89 differentially expressed genes identified in the Tbx5 single null vs DKO comparison.

Rescue of calcium handling through *Pln* deletion rescues development of atrial and cardiomyocyte hypertrophy

Our transcriptional profiling results indicated that *Pln* deletion rescues the transcriptional fibrotic and inflammatory response, however we wished to investigate if this transcriptional rescue translated to functional rescue of AF associated phenotypes. We investigated whether *Pln* deletion also rescued the physical atrial hypertrophy associated with AF. We imaged hearts from control, *Tbx5* single null, *Pln* single null, and *Tbx5 Pln* DKO animals 4 weeks post-tamoxifen, a time point when the atria begin to enlarge and *Tbx5* single null animals begin to display spontaneous AF (Figure 4.4A-D). We saw that compared to the control, *Pln* single null animals displayed no signs of atrial hypertrophy at this time point, while the size of the *Tbx5* single null atria were markedly increased (Figure 4.4B,C). In alignment with our transcriptional profiling, we see that atria from the DKO animals also do not show signs of atrial hypertrophy, indicating functional rescue of hypertrophy (Figure 4.4D).

With the understanding that *Pln* deletion rescues the hypertrophic response at the level of the whole atrium, we next investigated whether *Pln* deletion resulted in hypertrophic rescue at the level of individual cardiomyocytes by measuring cardiomyocyte width across all genotypes (Figure 4.4E-H). We found that the average cardiomyocyte width was markedly increased in the *Tbx5* single null compared to control animals (13.2 uM compared to 8.5 uM, pvalue = 9.7E-6) (Figure 4.4I). In contrast, DKO animals showed a significant decrease in average cardiomyocyte width compared to the *Tbx5* single null, indicative of rescue of the hypertrophic phenotype (9.2uM compared to 13.2uM, pvalue = 1.5E-4) (Figure 4.4I). *Pln* single null animals and DKO animals showed no difference in average cardiomyocyte width compared to WT (8.3uM compared to 8.5uM, pvalue = 0.78, and 9.2uM compared to 8.5uM, pvalue = 0.36 respectively). These results corroborate our transcriptional findings: that rescue of calcium handling through *Pln* deletion leads to rescue of disease processes including inflammation, fibrosis, and hypertrophy at both the transcriptional and functional level.

Figure 4.4: **Rescue of calcium handling rescues atrial and cardiomyocyte hypertrophy after AF onset**

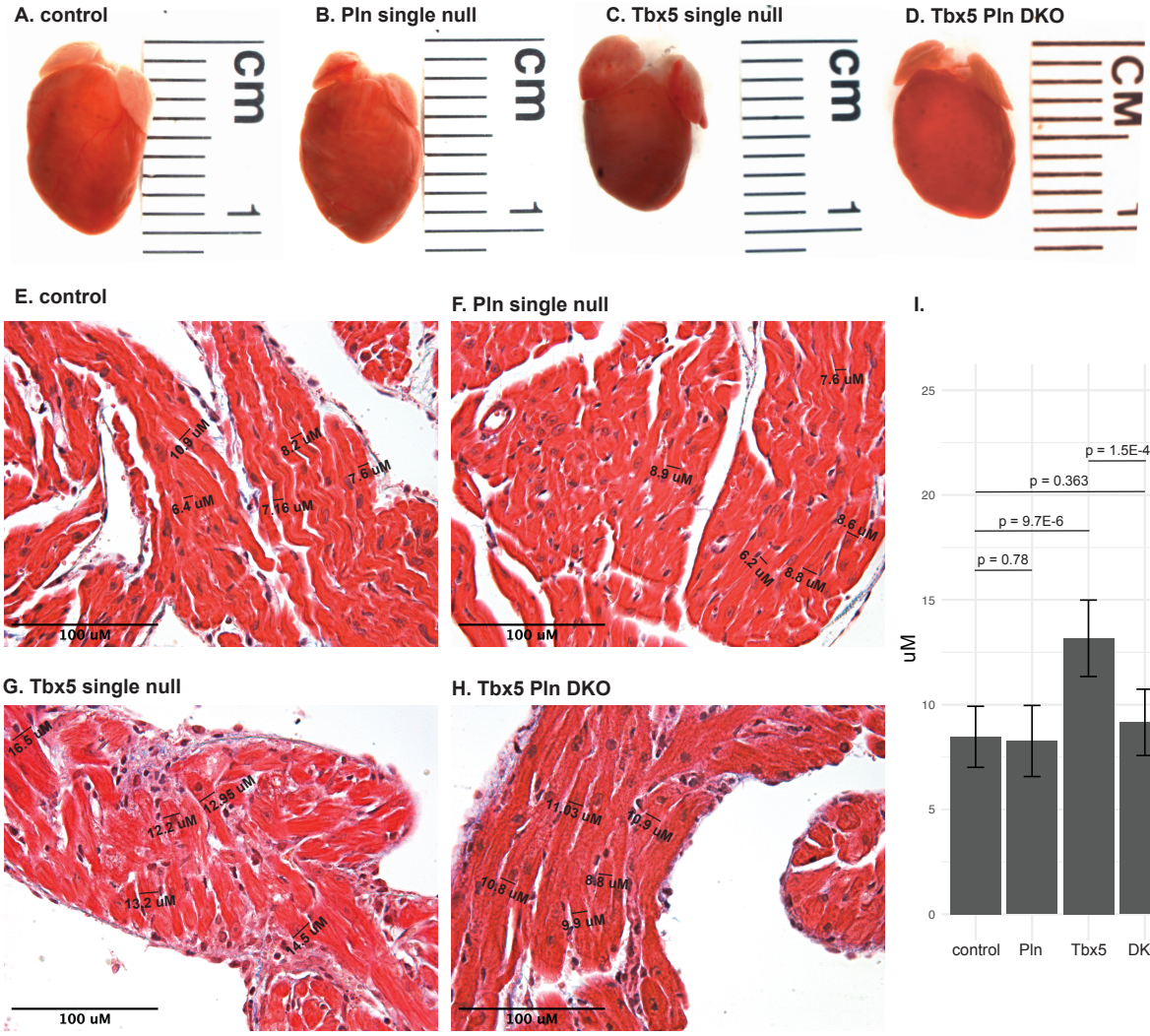


Figure 4.4: Legend

A-D: Whole heart images of control (A), *Pln* single null (B), *Tbx5* single null (C), and DKO animals (D)

E-H: Representative images of H&E and trichrome stained cardiomyocytes from control (E), *Pln* single null (F), *Tbx5* single null (G), and DKO animals (H). Measurements of five cardiomyocyte widths are shown in uM

I: Quantification of cardiomyocyte widths from control (n=2), *Pln* single null (n=1), *Tbx5* single null (n=2), and DKO animals (n=2). 5 measurements were taken per animal and averaged to generate quantification. Error bars are indicative of standard deviation. Pvalues generated by two tailed paired t-test. For *Pln* single null, two images of different sections were used to take cardiomyocyte lengths and analyzed as described above.

Modulation of the inflammatory response through *Il6* deletion exacerbates AF

When taking a closer look at the results across the *Tbx5* single null, the DKO, and the comparison of those two, we noticed a trend of terms, motifs, and genes that are all associated with IL6 production and regulation. Specifically, upregulated genes between the *Tbx5* single and the DKO, as well as in the DKO versus WT, show enrichment of the GO term "negative regulation of interleukin 6 production". Shared upregulated genes between the *Tbx5* single and the DKO also display enrichment for STAT motifs, known members of the IL6 signaling pathway, in the promoters of shared upregulated genes (Figure 4.2E). In addition to these shared findings, we also noticed genes related to IL6 regulation identified as DE between the *Tbx5* single null and the DKO, as well as the enrichment of Stat motifs at promoters (Figure 4.3C,D). Specifically we were curious about the identification of *Arid5a*, an RNA-binding protein that is known to stabilize inflammation-related mRNAs such as *Il6* and *Stat3*, as well as previously described ATF3, a transcription factor that is known to interact with and inhibit IL6[159].

This confluence of gene expression changes related to IL6 regulation highlighted the potential role of IL6 in AF and AF rescue. IL6 is a pleotropic cytokine that is known to have both anti- and pro-inflammatory roles in the heart[160]. IL6 can be produced by most cells, in addition to all immune system cells, and signals through membrane bound receptors (cis, or classical signaling) as well as through a soluble form of its receptor (trans signaling). IL6 signaling in the heart is cardioprotective in the acute response, however chronic upregulation of IL6 signaling is maladaptive, leading to cardiac hypertrophy and a decrease in contractile function [160], [161]. In the acute phase, IL6 signaling protects cardiomyocytes from oxidative stress and activates anti-apoptotic pathways [161], [162].

Research on the role of IL6 in AF has uncovered that elevated levels of circulating IL6 are associated with an increased risk of AF in patients with chronic kidney disease and coronary artery disease [163]–[165]. Furthermore, in patients with persistent and permanent AF, IL6 levels were increased and positively correlated to AF duration as well as left atrial size, implicating elevated IL6 in AF-associated structural atrial remodeling [166]. Another study on patients with inflammatory conditions found that IL6 elevation induces atrial electrical remodeling through the reduction of cardiac connexins 40 and 43, encoded by *Gja5* and *Gja1* [167].

In order to understand the role of IL6 in AF, we crossed *Il6* null animals with our *Tbx5* null animals in order to generate litter matched control, *Tbx5* single null, *Il6* single null, and DKO animals. To test the effect of *Il6* deletion on AF, we subjected the animals to tamoxifen injection and surface ECGs across a time course for 4 weeks. Surprisingly, we found that *Tbx5 Il6* double null animals developed atrial fibrillation much earlier than *Tbx5* single null animals and displayed electrocardiographic defects as early as 1.5 weeks post tamoxifen (Figure 4.5). Specifically, at 1.5 weeks post tamoxifen we see that DKO animals have significantly longer P wave durations than WT animals, while *Tbx5* single null animals and *Il6* single null animals both have normal P wave durations at this time point (Figure 4.5A). At the 2 week time point, *Tbx5* single null animals also begin to display significantly longer P wave durations than WT animals (Figure 4.5A).

In addition to the changes in P wave, we also see significant differences in PR interval (Figure 4.5B), QRS duration (Figure 4.5C), and RR interval (Figure 4.5D) between the DKO and control animals. Due to the early onset of atrial fibrillation and ventricular tachycardia, we were unable to calculate these measurements for all animals at later timepoints. Surface ECG at each

timepoint (Figure 4.5I-L) revealed lengthened P waves in the DKO at 1.5 weeks compared to all other genotypes (Figure 4.5, top row, L compared to I). At 3 weeks post tamoxifen, surface ECG reveals an extreme lengthening of P waves in the DKO compared to all other genotypes (Figure 4.5, third row, L compared to I). At 2 and 4 weeks post tamoxifen, atrial fibrillation can be seen in the DKO (Figure 4.5L, second and fourth row) as well as in the *Tbx5* single null 4 weeks post tamoxifen (Figure 4.5J, fourth row).

In addition to the ECG data, we wished to understand how *Il6* deletion may impact the atrial hypertrophy associated with AF. We imaged hearts from control, *Tbx5* single null, *Il6* single null, and *Tbx5 Il6* DKO animals 4 weeks post-tamoxifen to investigate whether these animals also displayed changes in atrial hypertrophy. Consistent with the ECG results, *Tbx5* single null and *Tbx5 Il6* double null animals both had markedly increased atria, indicative of hypertrophy (Figure 4.5I-L). While more animals are necessary to increase our confidence in the phenotypic differences between *Tbx5* single null, *Il6* single null, and DKO animals, these preliminary results clearly show significant differences in the development of atrial fibrillation and hypertrophy across genotype.

Figure 4.5: Modulation of the inflammatory response through IL6 deletion exacerbates AF and atrial hypertrophy

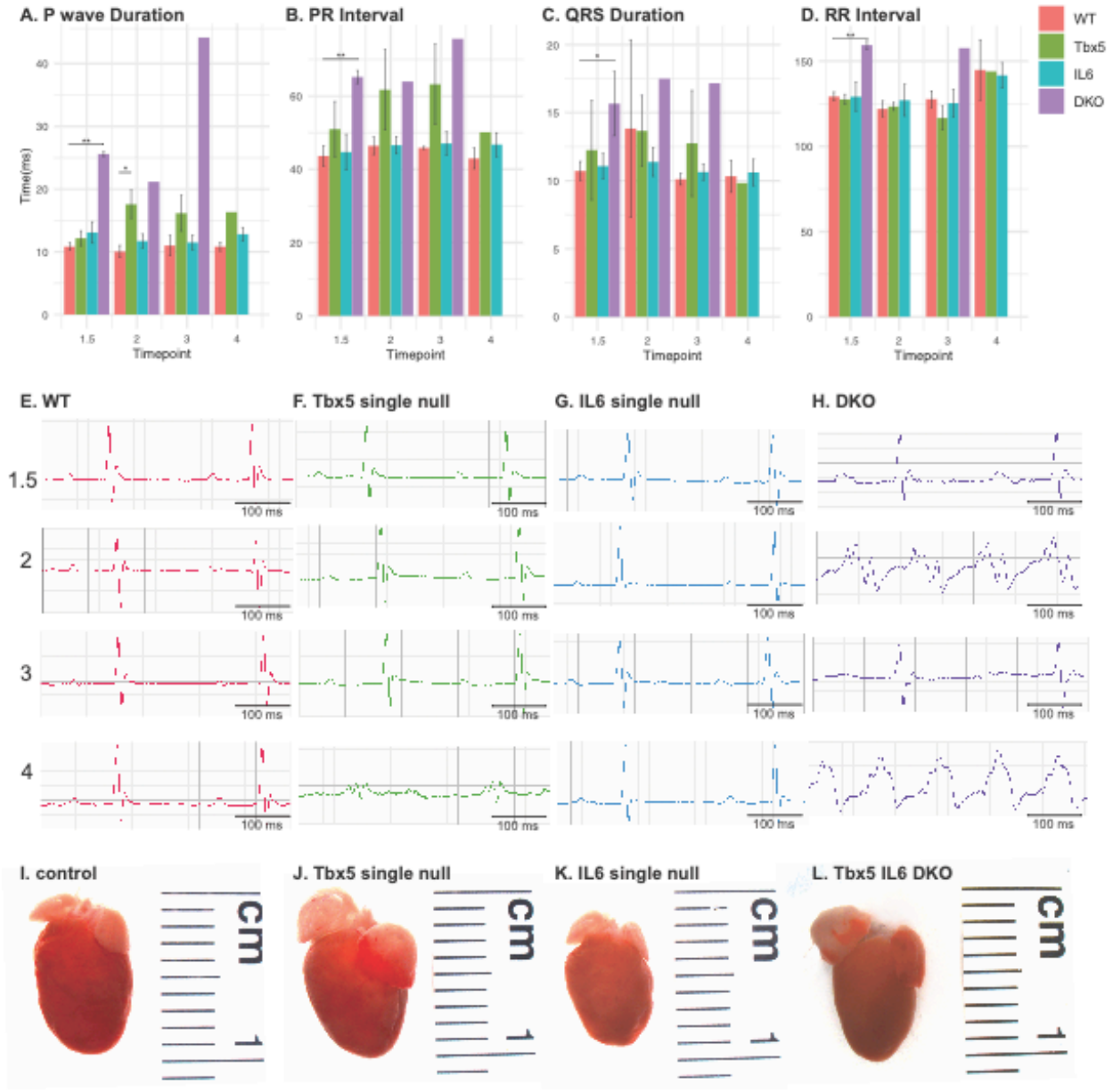


Figure 4.5: Legend

A. P-wave duration in ms for control (3), Tbx5 single null (2), IL6 single null (6), and DKO (2) animals across timepoints at 1.5 weeks, 2, 3, and 4 weeks post-tamoxifen. Error bars show standard deviation. Significant differences with pvalue < 0.05 marked with *, pvalue < 0.01 marked with **. Significance calculated using two-tailed ttest.
B. PR interval in ms for control (3), Tbx5 single null (2), IL6 single null (6), and DKO (2) animals across timepoints at 1.5 weeks, 2, 3, and 4 weeks post-tamoxifen. Error bars

Figure 4.5: Legend, continued

show standard deviation. Significant differences with pvalue < 0.05 marked with *, pvalue < 0.01 marked with **. Significance calculated using two-tailed ttest.

C. QRS duration in ms for control (3), Tbx5 single null (2), IL6 single null (6), and DKO (2) animals across timepoints at 1.5 weeks, 2, 3, and 4 weeks post-tamoxifen. Error bars show standard deviation. Significant differences with pvalue < 0.05 marked with *, pvalue < 0.01 marked with **. Significance calculated using two-tailed ttest.

D. RR interval in ms for control (3), Tbx5 single null (2), IL6 single null (6), and DKO (2) animals across timepoints at 1.5 weeks, 2, 3, and 4 weeks post-tamoxifen. Error bars show standard deviation. Significant differences with pvalue < 0.05 marked with *, pvalue < 0.01 marked with **. Significance calculated using two-tailed ttest.

E-H. Surface ECGs for each genotype at each time point

I-J. Whole heart images from control (I), Tbx5 single null (J), IL6 single null (K) and DKO (L) animals

4.4 Discussion:

This work begins to elucidate the relationship between AF, inflammation, and fibrosis. We show that in *Tbx5*-dependent AF, rescue of calcium homeostasis through *Pln* deletion also rescues the transcriptional activation of the fibrotic and inflammatory response. Interestingly, this occurs prior to the onset of AF, a much earlier timepoint than canonically understood based on the appearance of physical fibrosis and inflammation in the heart. This finding suggests that while electrical remodeling and structural remodeling are indeed interconnected, rescue of electrical defects have the propensity to inhibit the onset of pathological structural remodeling.

Pln deletion rescues the transcriptional fibrotic and inflammatory response in *Tbx5*-dependent AF

The finding that the *Pln* single deletion does not result in any transcriptional changes in the left atrium underscores the precise nature of how *Pln* deletion rescues calcium handling in the *Tbx5* mutant atrium. Through modulation of SERCA activity alone, *Pln* deletion is able to compensate for the suite of transcriptional changes that occur in response to *Tbx5* deletion, including the downregulation of critical calcium handling components such as the ryanodine receptor, gap junction genes, and other cardiac genes. Additionally, this finding indicates the lack of a transcriptional feedback loop at the level of calcium handling, as in the DKO we still observe the downregulation of these critical cardiac genes.

The high degree of similarity between the *Tbx5* single null and the DKO cements the role of *Tbx5* as a master regulator of cardiac gene expression. Since *Tbx5* is of course deleted in both contexts, the fact that the major transcriptional insult to the system is shared between both models fits with our understanding of the role of *Tbx5*. However, there are also transcriptional

differences between the DKO and the *Tbx5* single null, indicative of genes specific to the AF state or the rescued state, depending on their directionality. The 89 direct DE genes between the DKO and the *Tbx5* single are highly enriched for pathways related to cell cycle and cell cycle control (Figure 4.3B). Since cardiomyocytes are terminal, non-dividing cells, we hypothesize that this signal is driven by proliferative changes in other cell types, namely fibroblasts. Therefore, this enrichment highlights the differential activation of fibroblasts and potentially the proliferation of other cell types, such as immune cells, between the *Tbx5* single null AF state and the rescued DKO state.

Transcriptional differences between DKO and *Tbx5* single null nominate candidates for disease rescue and progression

There are many specific DE genes between the DKO and the *Tbx5* single that are interesting candidate genes for AF rescue and for AF disease programs. One particularly interesting candidate is *Pnmt*, a gene encoding phenylethanolamine N-methyltransferase. *Pnmt* is typically expressed in cardiac development, and seeing its resurgence of expression specifically in the DKO may represent the known paradigm of cardiomyocytes expressing fetal- and developmental-like gene regulatory programs in response to disease [168], [169]. Furthermore, the finding that *Pnmt* expressing cardiomyocytes are a distinct cell population in the developing heart, and in particular contribute to pacemaker-like cells such as the SA node implies a role for *Pnmt* in establishing cardiomyocyte populations that are capable of autonomous depolarization. One hypothesis for the expression pattern we see for *Pnmt* across the Tbx5 single and DKO is that in the DKO in order to rescue AF, atrial cardiomyocytes are departing from a typical adult cardiomyocyte gene regulatory program to express *Pnmt* and become more capable of autonomous depolarizations to mitigate the ectopic triggers that cause AF. Alternatively, *Pnmt*

has been shown to be required for the normal cardiac stress response, so in the DKO context cardiomyocytes may be activating stress-responsive genes and pathways to account for *Tbx5* deletion in the context of normalized calcium handling [154].

Atf3 is also an interesting candidate gene to be involved in mechanisms of AF and AF rescue. This TF has documented roles in promoting cardiac hypertrophy and fibrosis in both cardiomyocytes and activated fibroblasts. The downregulation of *Atf3* in the DKO may drive the downregulation of cardiac hypertrophy and fibrotic gene pathways, leading to the inhibition of these pathways in AF rescue. And conversely, the upregulation of *Atf3* may contribute to the activation of fibrotic pathways in the *Tbx5* single null. It is of interest that in each individual state compared to WT, *Atf3* does not meet the threshold of statistical significance, but does in the comparison between the DKO and the *Tbx5* single. This highlights the utility of performing transcriptional profiling not just between control and diseased states, but also between diseased and rescued states to unveil novel gene expression patterns. In addition, we see a strong transcriptional activation of *Atf3* specifically in cardiomyocytes and epicardial cells in response to *Tbx5* deletion through single cell sequencing, which highlights the cardiomyocyte specific role of *Atf3* in *Tbx5*-dependent AF. Furthermore, the negative relationship between *Atf3* and IL6 fits with our transcriptional findings from the *Tbx5 Pln* DKO mouse, as well as the functional findings from the *Tbx5 Il6* DKO mouse, suggesting a link between *Tbx5* and IL6 signaling in AF.

The role of IL6 in AF and inflammation

Our data suggests that modulation of inflammation through the deletion of *Il6* is detrimental for *Tbx5*-dependent AF, causing an acceleration of the conduction defects seen in *Tbx5* single null animals. *Il6* deletion does not show this effect on its own, suggesting that this

effect is specific to the cellular state conferred by the double knockout. This is an interesting result, especially due to the breadth of literature citing IL6 upregulation being associated with AF, and that IL6 neutralization and transcriptional downregulation as being beneficial in heart failure[160], [162], [167]. However, the role of IL6 in promoting inflammation is complex, and it is known that acute upregulation of IL6 signaling can be protective. We hypothesize that since our model incorporates a germline IL6 null, *Il6* deletion in this case occurs prior to the onset of AF, leading to the absence of acute cardioprotective IL6 signaling. Whereas other models of IL6 neutralization occur after the onset of cardiac disease, allowing for IL6 to perform its acute anti-inflammatory roles, but inhibiting its chronic pro-inflammatory expression[170]. This aligns with time-course transcriptional profiling in *Tbx5*-dependent AF (unpublished), where IL6 upregulation begins at D0 following Tbx5 deletion, and its expression lowers by D3, and rises again at D7 post-tamoxifen. While these data support this hypothesis, more testing is required in order to differentiate between the acute and chronic IL6 upregulation stages and how they impact *Tbx5*-dependent AF.

4.5 Future Directions:

Mechanisms of AF and rescue

One appealing hypothesis that arises from these results is that in AF rescue a more developmental gene regulatory program is activated in order to compensate for the cellular stress of AF. Our data supports this hypothesis as we see canonically developmentally expressed genes, such as *Pnmt*, upregulated in our DKO but not in our *Tbx5* single null. Additionally, we see the enrichment of developmental TF motifs at the promoters of genes directly DE between the DKO and the Tbx5 single null, such as Eomes and Brachyury, indicating that the transcriptional differences between the DKO and *Tbx5* single null may be explained through the activation of a more fetal-like gene regulatory program. This hypothesis requires further testing in order to be confirmed, such as investigating whether other DE genes between the DKO and *Tbx5* single are canonically expressed during development and not in adult stages.

In order to gain a deeper understanding of the candidate genes nominated in this work, cell-type resolved data would provide critical insights to the cell-type specificity of the gene expression patterns that we see. It is well established that AF pathophysiology involves may cell-types within the heart, including cardiomyocytes, fibroblasts and activated fibroblasts, the resident immune population and disease-induced infiltration of immune cells, and changes to other cell types can all contribute to the electrical and structural remodeling that establish AF. In particular, we see a strong enrichment of cell cycle pathways in the DE genes between the DKO and *Tbx5* single, which we hypothesize to be driven largely by non-cardiomyocyte cell types. Single cell approaches would allow for the determination of where cell cycle expression is activated in both the DKO and *Tbx5* single, as well as differentially between the two.

A central question that arises from these data is what transcription factors drive the expression of activated genes in the *Tbx5* single null and the DKO, especially considering that *Tbx5* is deleted in both contexts. Promoter analysis has identified a strong enrichment of STAT motifs in both shared DE genes between the *Tbx5* single and DKO, as well as differentially expressed genes between both contexts, alongside the enrichment of other cardiac and developmental TFs as described above. Deeper investigation into the upstream regulatory network directly regulating these differentially expressed genes would provide a link between *Tbx5* expression and the expression of these genes. Based on the confluence of STAT motifs, as well as the clear link between IL6, STAT3, and AF; performing ChIPseq for STAT3 in control and *Tbx5*-dependent AF samples may provide additional insight to the direct regulation of genes involved in *Tbx5*-dependent AF and rescue.

Relationship between inflammation, fibrosis, and AF

While these results begin to unravel the relationship between inflammation, fibrosis, and AF; as well as the transcription factors that drive these processes, more experimentation is required to fully disentangle these relationships. In order to understand the role of IL6, more mice must be tested for AF in order to make statistically significant claims about the incidence and timing of AF in *Tbx5 Il6* double knock out animals. In addition, differences in P wave duration occur as early as at the 1.5 week time point, it is of interest to look earlier post *Tbx5* deletion in order to identify when exactly *Tbx5 Il6* DKO animals begin to display AF phenotypes. In addition to performing global measures of AF such as surface ECG on these animals, it would be of interest to investigate more mechanistic measures of AF by examining calcium handling defects more closely through measuring cellular calcium transients and SERCA activity in *Il6* single null and *Tbx5 Il6* double knockout animals.

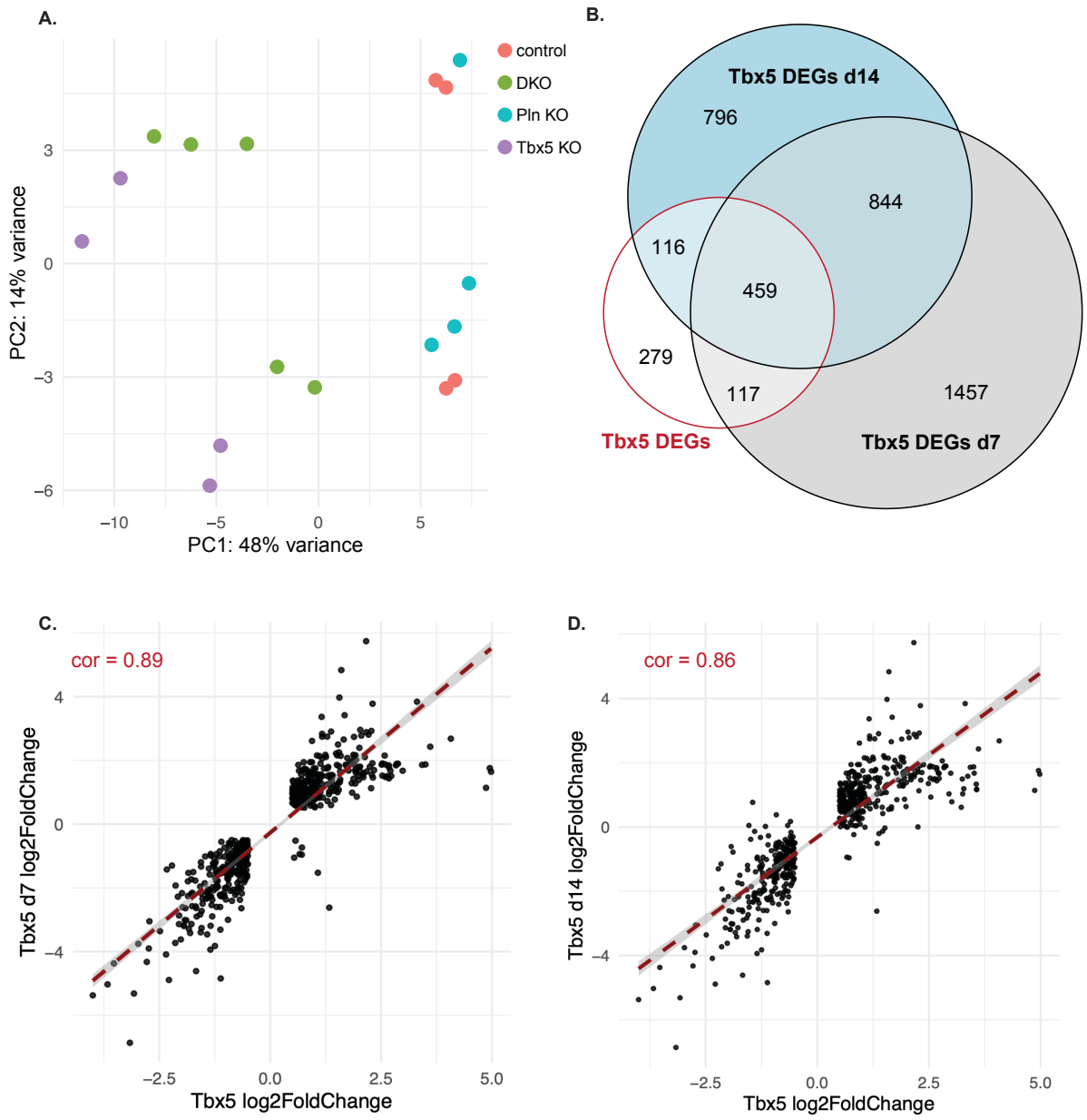
Deletion of IL6 is a suboptimal approach to investigate the role of inflammation in AF. In order to fully investigate the relationship between inflammation and AF, it would be of interest to use other methods to modulate inflammation, such as treating mice with dexamethasone, a repressor of inflammation that has been shown to inhibit cardiac macrophage expansion[171], [172]. Comparing the phenotypes of dexamethasone treated WT and *Tbx5* null animals to the results from our IL6 investigation would allow us to understand how IL6 may be functioning in concert with, or even separate from, the inflammation associated with AF. Additionally, in order to test the hypothesis that IL6 provides anti-inflammatory protection in the acute phase, but pro-inflammatory roles in the chronic phase, developing a conditional knockout of IL6 would allow for the modulation of IL6 deletion in a time controlled manner.

While additional experiments are necessary to confirm and interrogate the exact nature of the role of IL6 signaling in AF, this work begins to disentangle the relationship between AF, fibrosis, and inflammation; in addition to illuminating key transcription factors and pathways involved in AF and AF rescue. The application of transcriptional profiling to the rescued DKO state in addition to the diseased and control state provides enhanced insight to the mechanisms of AF and nominates candidate genes involved in rescuing AF disease processes beyond calcium handling. The finding that rescue of calcium handling through *Pln* deletion also rescues the transcriptional fibrotic and inflammatory response prior to the onset of AF, which translates to a rescue of physical atrial hypertrophy, begins to delineate the relationships between electrical remodeling and structural remodeling in AF. In addition, this work highlights the functionality and specificity of calcium handling rescue through *Pln* deletion. Finally, while preliminary, the finding that *Il6* deletion exacerbates and accelerates AF provides critical insight to the role of IL6 signaling in AF, and highlights the complex relationship between inflammation and AF.

4.6: Acknowledgement of Work Performed:

I would like to acknowledge the individuals who contributed to this work. This project was first conceptualized due to the work of Wenli Dai, Brigitte LaForest, Christopher Weber, and Ivan Moskowitz. Zhezhen Wang assisted with alignment and QC for RNAseq processing. All other analysis and experimentation presented in this chapter were performed by Linsin Smith.

Supplemental Figure 4.1



Supplemental Figure 4.1: Legend

A. Principle component analysis on control, Tbx5 single null, Pln single null, and DKO samples. Observe clear separation by genotype (PC1) and sex (PC2).
B. Overlap between identified significant DEGs between Tbx5 and WT (outlined in red) and independently generated Tbx5 vs WT RNAseq experiments at 7 days post tamoxifen and 14 days post tamoxifen respectively (outlined in black, data unpublished).

Supplemental Figure 4.1: Legend, continued

C. Scatterplot shows log2FoldChange for shared DEGs between Tbx5 vs WT (x-axis) and independent Tbx5 vs WT at 7 days post tamoxifen (y-axis), Pearson's correlation = 0.89
D. Scatterplot shows log2FoldChange for shared DEGs between Tbx5 vs WT (x-axis) and independent Tbx5 vs WT at 14 days post tamoxifen (y-axis), Pearson's correlation = 0.86

Bibliography

[1] I. Dunham *et al.*, "An Integrated Encyclopedia of DNA Elements in the Human Genome," *Nature*, vol. 489, no. 7414, p. 57, Sep. 2012.

[2] T. I. Lee and R. A. Young, "Transcriptional Regulation and its Misregulation in Disease," *Cell*, vol. 152, no. 6, p. 1237, Mar. 2013.

[3] S. L. Edwards, J. Beesley, J. D. French, and M. Dunning, "Beyond GWASs: Illuminating the dark road from association to function," *American Journal of Human Genetics*, vol. 93, no. 5. Cell Press, pp. 779–797, 07-Nov-2013.

[4] D. Fatkin, C. F. Santiago, I. G. Huttner, S. A. Lubitz, and P. T. Ellinor, "Genetics of Atrial Fibrillation: State of the Art in 2017," *Heart. Lung Circ.*, vol. 26, no. 9, pp. 894–901, Sep. 2017.

[5] M. S. Olesen, M. W. Nielsen, S. Haunsø, and J. H. Svendsen, "Atrial fibrillation: the role of common and rare genetic variants," *Eur. J. Hum. Genet.*, vol. 22, no. 3, pp. 297–306, Mar. 2014.

[6] Y.-S. Ang *et al.*, "Disease Model of GATA4 Mutation Reveals Transcription Factor Cooperativity in Human Cardiogenesis Graphical Abstract Highlights d Systems-level approach reveals GATA4 roles in human cardiac development and function," 2016.

[7] R. D. Nadadur *et al.*, "Pitx2 modulates a Tbx5-dependent gene regulatory network to maintain atrial rhythm," *Sci Transl Med*, vol. 8, no. 354, p. 354ra115, Aug. 2016.

[8] E. P. Consortium, "An integrated encyclopedia of DNA elements in the human genome," *Nature*, vol. 489, pp. 57–74, 2012.

[9] M. T. Y. Lam, W. Li, M. G. Rosenfeld, and C. K. Glass, "Enhancer RNAs and regulated transcriptional programs," *Trends Biochem. Sci.*, vol. 39, no. 4, pp. 170–182, Apr. 2014.

[10] M. S. Werner *et al.*, "Chromatin-enriched lncRNAs can act as cell-type specific activators of proximal gene transcription," *Nat. Struct. Mol. Biol.*, vol. 24, no. 7, pp. 596–603, Jun. 2017.

[11] C. Davidovich and T. R. Cech, "The recruitment of chromatin modifiers by long noncoding RNAs: lessons from PRC2," *RNA*, vol. 21, pp. 2007–2022, 2015.

[12] W. Li, D. Notani, and M. G. Rosenfeld, "Enhancers as non-coding RNA transcription units: recent insights and future perspectives," *Nat. Rev. Genet.*, vol. 17, no. 4, pp. 207–223, Apr. 2016.

[13] X. H. Yang *et al.*, "Transcription-factor-dependent enhancer transcription defines a gene regulatory network for cardiac rhythm," *Elife*, vol. 6, Dec. 2017.

[14] G. Natoli and J.-C. Andrau, "Noncoding transcription at enhancers: general principles and functional models.," *Annu. Rev. Genet.*, vol. 46, pp. 1–19, 2012.

[15] H. Wu *et al.*, "Tissue-Specific RNA Expression Marks Distant-Acting Developmental Enhancers," *PLoS Genet.*, vol. 10, no. 9, p. e1004610, Sep. 2014.

[16] C. Perez-Cervantes *et al.*, " Enhancer transcription identifies cis -regulatory elements for photoreceptor cell types ," *Development*, vol. 147, no. 3, p. dev.184432, Jan. 2020.

[17] S. P. Simonett *et al.*, "Identification of direct transcriptional targets of NFATC2 that promote β-cell proliferation," *J. Clin. Invest.*, Sep. 2021.

[18] J. G. Azofeifa, M. A. Allen, J. R. Hendrix, T. Read, J. D. Rubin, and R. D. Dowell, "Enhancer RNA profiling predicts transcription factor activity.," *Genome Res.*, vol. 28, no. 3, pp. 334–344, Feb. 2018.

[19] D. Wang *et al.*, "Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA," *Nature*, vol. 474, no. 7351, pp. 390–394, Jun. 2011.

[20] O. Mikhaylichenko *et al.*, "The degree of enhancer or promoter activity is reflected by the levels and directionality of eRNA transcription.," *Genes & Dev.*, vol. 32, pp. 42–57, 2018.

[21] S. Bonn *et al.*, "Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development," *Nat Genet*, vol. 44, pp. 148–156, 2012.

[22] E. C. Mouse *et al.*, "An encyclopedia of mouse DNA elements (Mouse ENCODE)," *Genome Biol*, vol. 13, p. 418, 2012.

[23] M. P. Creyghton *et al.*, "Histone H3K27ac separates active from poised enhancers and predicts developmental state," *Proc Natl Acad Sci U S A*, vol. 107, pp. 21931–21936, 2010.

[24] N. Dogan *et al.*, "Occupancy by key transcription factors is a more accurate predictor of enhancer activity than histone modifications or chromatin accessibility," *Epigenetics Chromatin*, vol. 8, no. 1, p. 16, Dec. 2015.

[25] A. Visel *et al.*, "ChIP-seq accurately predicts tissue-specific activity of enhancers," *Nature*, vol. 457, pp. 854–858, 2009.

[26] M. S. Werner and A. J. Ruthenburg, "Nuclear Fractionation Reveals Thousands of Chromatin-Tethered Noncoding RNAs Adjacent to Active Genes," *Cell Rep.*, vol. 12, no. 7, pp. 1089–1098, Aug. 2015.

[27] D. Hartl, A. R. Krebs, J. Jüttner, B. Roska, and D. Schübeler, "Cis-regulatory landscapes of four cell types of the retina," *Nucleic Acids Res.*, vol. 45, no. 20, pp. 11607–11621, Nov. 2017.

[28] M. A. White, C. A. Myers, J. C. Corbo, and B. A. Cohen, "Massively parallel in vivo enhancer assay reveals that highly local features determine the cis-regulatory function of ChIP-seq peaks," *Proc. Natl. Acad. Sci.*, vol. 110, no. 29, pp. 11952–11957, Jul. 2013.

[29] M. A. White, J. C. Kwasnieski, C. A. Myers, S. Q. Shen, J. C. Corbo, and B. A. Cohen Correspondence, "A Simple Grammar Defines Activating and Repressing cis-Regulatory Elements in Photoreceptors," *Cell Rep.*, vol. 17, no. 5, pp. 1247–1254, Oct. 2016.

[30] L. Xiong *et al.*, "Genome-wide Identification and Characterization of Enhancers Across

10 Human Tissues," *Int. J. Biol. Sci.*, vol. 14, pp. 1321–1332, 2018.

[31] J. M. Engreitz *et al.*, "Local regulation of gene expression by lncRNA promoters, transcription and splicing," *Nature*, vol. 539, pp. 452–455, 2016.

[32] H. Wu *et al.*, "Tissue-Specific RNA Expression Marks Distant-Acting Developmental Enhancers," *PLoS Genet.*, vol. 10, no. 9, p. e1004610, Sep. 2014.

[33] S. Blackshaw *et al.*, "Genomic Analysis of Mouse Retinal Development," *PLoS Biol.*, vol. 2, no. 9, p. e247, Jun. 2004.

[34] A. La Torre, S. Georgi, and T. A. Reh, "Conserved microRNA pathway regulates developmental timing of retinal neurogenesis," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, pp. E2362–E2370, 2013.

[35] L. Zelinger *et al.*, "Regulation of Noncoding Transcriptome in Developing Photoreceptors by Rod Differentiation Factor NRL," *Investig. Ophthalmol. & Vis. Sci.*, vol. 58, pp. 4422–4435, 2017.

[36] C. L. Cepko, "The Determination of Rod and Cone Photoreceptor Fate," *Annu. Rev. Vis. Sci.*, vol. 1, pp. 211–234, 2015.

[37] C. Koike *et al.*, "Functional roles of Otx2 transcription factor in postnatal mouse retinal development," *Mol. Cell. Biol.*, vol. 27, pp. 8318–8329, 2007.

[38] A. Nishida *et al.*, "Otx2 homeobox gene controls retinal photoreceptor cell fate and pineal gland development," *Nat Neurosci*, vol. 6, pp. 1255–1263, 2003.

[39] D. Altshuler and L. Lillien, "Control of photoreceptor development," *Curr Opin Neurobiol*, vol. 2, pp. 16–22, 1992.

[40] M. M. Emerson, N. Surzenko, J. J. Goetz, J. Trimarchi, and C. L. Cepko, "Otx2 and Onecut1 promote the fates of cone photoreceptors and horizontal cells and repress rod photoreceptors," *Dev Cell*, vol. 26, pp. 59–72, 2013.

[41] A. Gonzalez-Cordero *et al.*, "Recapitulation of Human Retinal Development from Human Pluripotent Stem Cells Generates Transplantable Populations of Cone Photoreceptors," *Stem Cell Reports*, vol. 9, pp. 820–837, 2017.

[42] B. P. Hafler *et al.*, "Transcription factor Olig2 defines subpopulations of retinal progenitor cells biased toward specific cell fates," *Proc Natl Acad Sci U S A*, vol. 109, pp. 7882–7887, 2012.

[43] J. Chen, "The rod photoreceptor-specific nuclear receptor Nr2e3 represses transcription of multiple cone-specific genes. - PubMed - NCBI," *J. Neurosci.*, vol. 25, pp. 118–129, 2005.

[44] A. J. Mears *et al.*, "Nrl is required for rod photoreceptor development," *Nat Genet*, vol. 29, pp. 447–452, 2001.

[45] E. C. Oh, N. Khan, E. Novelli, H. Khanna, E. Strettoi, and A. Swaroop, "Transformation of cone precursors to functional rod photoreceptors by bZIP transcription factor NRL," *Proc Natl Acad Sci U S A*, vol. 104, pp. 1679–1684, 2007.

[46] P. K. Swain *et al.*, “Multiple phosphorylated isoforms of NRL are expressed in rod photoreceptors.,” *J. Biol. Chem.*, vol. 276, pp. 36824–36830, 2001.

[47] C.-J. Jeon, E. Strettoi, and R. H. Masland, “The Major Cell Populations of the Mouse Retina,” *J. Neurosci.*, vol. 18, pp. 8936–8946, 1998.

[48] A. E. Hughes, J. M. Enright, C. A. Myers, S. Q. Shen, and J. C. Corbo, “Cell Type-Specific Epigenomic Analysis Reveals a Uniquely Closed Chromatin Architecture in Mouse Rod Photoreceptors,” *Sci Rep*, vol. 7, p. 43184, 2017.

[49] A. Swaroop, D. Kim, and D. Forrest, “Transcriptional regulation of photoreceptor development and homeostasis in the mammalian retina,” *Nat. Rev. Neurosci.*, vol. 11, no. 8, pp. 563–576, Aug. 2010.

[50] A. E. O. Hughes, C. A. Myers, and J. C. Corbo, “A massively parallel reporter assay reveals context-dependent activity of homeodomain binding sites in vivo.,” *Genome Res.*, vol. 28, pp. 1520–1531, 2018.

[51] A. Mo *et al.*, “Epigenomic landscapes of retinal rods and cones,” *Elife*, vol. 5, p. e11613, 2016.

[52] M. M. Andzelm *et al.*, “MEF2D Drives Photoreceptor Development through a Genome-wide Competition for Tissue-Specific Enhancers,” *Neuron*, vol. 86, pp. 247–263, 2015.

[53] S. Chen *et al.*, “Crx, a Novel Otx-like Paired-Homeodomain Protein, Binds to and Transactivates Photoreceptor Cell-Specific Genes,” *Neuron*, vol. 19, pp. 1017–1030, 1997.

[54] T. Furukawa, E. M. Morrow, and C. L. Cepko, “Crx, a Novel otx-like Homeobox Gene, Shows Photoreceptor-Specific Expression and Regulates Photoreceptor Differentiation,” *Cell*, vol. 91, pp. 531–541, 1997.

[55] A. K. Hennig, G.-H. Peng, and S. Chen, “Regulation of photoreceptor gene expression by Crx-associated transcription factor network.,” *Brain Res.*, vol. 1192, pp. 114–133, 2008.

[56] T. H.-C. Hsiau, C. Diaconu, C. A. Myers, J. Lee, C. L. Cepko, and J. C. Corbo, “The Cis-regulatory Logic of the Mammalian Photoreceptor Transcriptional Network,” *PLoS One*, vol. 2, no. 7, p. e643, Jul. 2007.

[57] F. J. Livesey, T. Furukawa, M. A. Steffen, G. M. Church, and C. L. Cepko, “Microarray analysis of the transcriptional network controlled by the photoreceptor homeobox gene Crx,” *Curr Biol*, vol. 10, pp. 301–310, 2000.

[58] P. A. Ruzycki, X. Zhang, and S. Chen, “CRX directs photoreceptor differentiation by accelerating chromatin remodeling at specific target sites.,” *Epigenetics & chromatin*, vol. 11, p. 42, 2018.

[59] J. C. Corbo, C. A. Myers, K. A. Lawrence, A. P. Jadhav, and C. L. Cepko, “A typology of photoreceptor gene expression patterns in the mouse,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 104, no. 29, pp. 12069–74, Jul. 2007.

[60] J. C. Corbo *et al.*, “CRX ChIP-seq reveals the cis-regulatory architecture of mouse

photoreceptors," *Genome Res.*, vol. 20, no. 11, pp. 1512–1525, Nov. 2010.

[61] H. Hao *et al.*, "Transcriptional Regulation of Rod Photoreceptor Homeostasis Revealed by In Vivo NRL Targetome Analysis," *PLoS Genet.*, vol. 8, p. e1002649, 2012.

[62] K. P. Mitton, P. K. Swain, S. Chen, S. Xu, D. J. Zack, and A. Swaroop, "The Leucine Zipper of NRL Interacts with the CRX Homeodomain A POSSIBLE MECHANISM OF TRANSCRIPTIONAL SYNERGY IN RHODOPSIN REGULATION," *J. Biol. Chem.*, vol. 275, pp. 29794–29799, 2000.

[63] G. Oh *et al.*, "DNA modification study of major depressive disorder: beyond locus-by-locus comparisons.," *Biol. Psychiatry*, vol. 77, no. 3, pp. 246–55, Feb. 2015.

[64] C. L. Montana, A. V Kolesnikov, S. Q. Shen, C. A. Myers, V. J. Kefalov, and J. C. Corbo, "Reprogramming of adult rod photoreceptors prevents retinal degeneration," *Proc Natl Acad Sci U S A*, vol. 110, pp. 1732–1737, 2013.

[65] S. Anders, P. T. Pyl, and W. Huber, "HTSeq--a Python framework to work with high-throughput sequencing data," *Bioinformatics*, vol. 31, pp. 166–169, 2015.

[66] M. I. Love, W. Huber, and S. Anders, "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2," *Genome Biol*, vol. 15, p. 550, 2014.

[67] A. Samuel, M. Housset, B. Fant, and T. Lamonerie, "Otx2 ChIP-seq Reveals Unique and Redundant Functions in the Mature Mouse Retina," *PLoS One*, vol. 9, no. 2, p. e89110, Feb. 2014.

[68] R. M. Layer, B. S. Pedersen, T. DiSera, G. T. Marth, J. Gertz, and A. R. Quinlan, "GIGGLE: a search engine for large-scale integrated genome analysis.," *Nat. Methods*, vol. 15, no. 2, pp. 123–126, 2018.

[69] R. Kolde, "pheatmap: Pretty Heatmaps," *R Packag. version 1.0.12*, 2019.

[70] Y. Shen *et al.*, "A map of the cis-regulatory sequences in the mouse genome," *Nature*, vol. 488, no. 7409, pp. 116–120, Aug. 2012.

[71] Q. Szabo *et al.*, "TADs are 3D structural units of higher-order chromosome organization in *Drosophila*," *Sci. Adv.*, vol. 4, no. 2, p. eaar8082, Feb. 2018.

[72] G. Yu, L.-G. Wang, Y. Han, and Q.-Y. He, "clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters," *Omi. A J. Integr. Biol.*, vol. 16, no. 5, pp. 284–287, May 2012.

[73] M. Martin, "Cutadapt removes adapter sequences from high-throughput sequencing reads," *2011*, vol. 17, p. 3.

[74] B. Langmead and S. L. Salzberg, "Fast gapped-read alignment with Bowtie 2," *Nat Methods*, vol. 9, pp. 357–359, 2012.

[75] H. Li *et al.*, "The Sequence Alignment/Map format and SAMtools," *Bioinformatics*, vol. 25, pp. 2078–2079, 2009.

[76] Y. Zhang *et al.*, "Model-based analysis of ChIP-Seq (MACS)," *Genome Biol*, vol. 9, p.

R137, 2008.

[77] M. Lawrence *et al.*, "Software for Computing and Annotating Genomic Ranges," *PLoS Comput. Biol.*, vol. 9, no. 8, p. e1003118, Aug. 2013.

[78] Z. Li, M. H. Schulz, T. Look, M. Begemann, M. Zenke, and I. G. Costa, "Identification of transcription factor binding sites using ATAC-seq," *Genome Biol.*, vol. 20, no. 1, p. 45, Dec. 2019.

[79] J. W. Kim *et al.*, "NRL-Regulated Transcriptome Dynamics of Developing Rod Photoreceptors," *Cell Rep*, vol. 17, pp. 2460–2473, 2016.

[80] M. J. Brooks, H. K. Rajasimha, J. E. Roger, and A. Swaroop, "Next-generation sequencing facilitates quantitative analysis of wild-type and Nrl(-/-) retinal transcriptomes," *Mol Vis*, vol. 17, pp. 3034–3054, 2011.

[81] J.-W. Kim *et al.*, "Recruitment of Rod Photoreceptors from Short-Wavelength-Sensitive Cones during the Evolution of Nocturnal Vision in Mammals," *Dev. Cell*, vol. 37, no. 6, pp. 520–532, Jun. 2016.

[82] W. Li, D. Notani, and M. G. Rosenfeld, "Enhancers as non-coding RNA transcription units: recent insights and future perspectives," *Nat. Rev. Genet.*, vol. 17, no. 4, pp. 207–223, Apr. 2016.

[83] I. Brunet *et al.*, "The transcription factor Engrailed-2 guides retinal axons," *Nature*, vol. 438, no. 7064, pp. 94–98, Nov. 2005.

[84] X. Li, V. Perissi, F. Liu, D. W. Rose, and M. G. Rosenfeld, "Tissue-specific regulation of retinal and pituitary precursor cell proliferation," *Science (80-.).*, vol. 297, pp. 1180–1183, 2002.

[85] D. Sapkota, H. Chintala, F. Wu, S. J. Fliesler, Z. Hu, and X. Mu, "Onecut1 and Onecut2 redundantly regulate early retinal cell fates during development," *Proc. Natl. Acad. Sci.*, vol. 111, no. 39, pp. E4086–E4095, Sep. 2014.

[86] M. Trieu, C. Lu, C. Severin, and D. W. Sretavan, "Direct autoregulation and gene dosage compensation by POU-domain transcription factor Brn3a," *Development*, vol. 130, no. 1, pp. 111–121, Jan. 2003.

[87] N. A. Billings, M. M. Emerson, and C. L. Cepko, "Analysis of thyroid response element activity during retinal development," *PLoS One*, vol. 5, p. e13739, 2010.

[88] M. M. Emerson and C. L. Cepko, "Identification of a retina-specific Otx2 enhancer element active in immature developing photoreceptors," *Dev Biol*, vol. 360, pp. 241–255, 2011.

[89] R. W. Young, "Cell differentiation in the retina of the mouse," *Anat. Rec.*, vol. 212, no. 2, pp. 199–205, Jun. 1985.

[90] Y. Fei, "Development of the cone photoreceptor mosaic in the mouse retina revealed by fluorescent cones in transgenic mice.," *Mol. Vis.*, vol. 9, pp. 31–42, Feb. 2003.

[91] M. R. Roberts, A. Hendrickson, C. R. McGuire, and T. A. Reh, "Retinoid X receptor

(gamma) is necessary to establish the S-opsin gradient in cone photoreceptors of the developing mouse retina," *Invest Ophthalmol Vis Sci*, vol. 46, pp. 2897–2904, 2005.

[92] J. Oswald and P. Baranov, "Regenerative medicine in the retina: from stem cells to cell replacement therapy," *Ther Adv Ophthalmol*, vol. 10, p. 2515841418774433, 2018.

[93] P. A. Campochiaro and T. A. Mir, "The mechanism of cone cell death in Retinitis Pigmentosa," *Prog. Retin. Eye Res.*, vol. 62, pp. 24–37, Jan. 2018.

[94] E. J. Shelley, M. C. Madigan, R. Natoli, P. L. Penfold, and J. M. Provis, "Cone Degeneration in Aging and Age-Related Macular Degeneration," *Arch. Ophthalmol.*, vol. 127, no. 4, p. 483, Apr. 2009.

[95] G. D. Aguirre, "Concepts and Strategies in Retinal Gene Therapy," *Invest Ophthalmol Vis Sci*, vol. 58, pp. 5399–5411, 2017.

[96] J. Juettner *et al.*, "Targeting neuronal and glial cell types with synthetic promoter AAVs in mice, non-human primates, and humans," *bioRxiv*, 2018.

[97] D. Sinha, J. Phillips, M. Joseph Phillips, and D. M. Gamm, "Mimicking Retinal Development and Disease With Human Pluripotent Stem Cells," *Invest Ophthalmol Vis Sci*, vol. 57, pp. ORSFf1-9, 2016.

[98] T. Henriques *et al.*, "Widespread transcriptional pausing and elongation control at enhancers," *Genes & Dev.*, vol. 32, pp. 26–41, 2018.

[99] F. Wu, D. Sapkota, R. Li, and X. Mu, "Onecut 1 and Onecut 2 are potential regulators of mouse retinal development.," *J. Comp. Neurol.*, vol. 520, pp. 952–969, 2012.

[100] G. C. Hon *et al.*, "Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues," *Nat. Genet.*, vol. 45, pp. 1198–1206, 2013.

[101] U. Jadhav *et al.*, "Extensive Recovery of Embryonic Enhancer and Gene Memory Stored in Hypomethylated Enhancer DNA," *Mol. Cell*, vol. 74, pp. 542-554.e5, 2019.

[102] B. Chang *et al.*, "Cone photoreceptor function loss-3, a novel mouse model of achromatopsia due to a mutation in Gnat2," *Invest Ophthalmol Vis Sci*, vol. 47, pp. 5017–5021, 2006.

[103] J. de Melo, G. H. Peng, S. Chen, and S. Blackshaw, "The Spalt family transcription factor Sall3 regulates the development of cone photoreceptors and retinal horizontal interneurons," *Development*, vol. 138, pp. 2325–2336, 2011.

[104] H. M. Rodgers, M. Belcastro, M. Sokolov, and P. H. Mathers, "Embryonic markers of cone differentiation," *Mol Vis*, vol. 22, pp. 1455–1467, 2016.

[105] S. Siegert *et al.*, "Transcriptional code and disease map for adult retinal cell types," *Nat. Neurosci.*, vol. 15, no. 3, pp. 487–495, Mar. 2012.

[106] S. Yoshida *et al.*, "Expression profiling of the developing and mature Nrl-/- mouse retina: identification of retinal disease candidates and transcriptional regulatory targets of Nrl," *Hum Mol Genet*, vol. 13, pp. 1487–1503, 2004.

[107] Y. Ogawa *et al.*, “Six6 and Six7 coordinately regulate expression of middle-wavelength opsins in zebrafish,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 116, pp. 4651–4660, 2019.

[108] G.-H. Peng and S. Chen, “Chromatin immunoprecipitation identifies photoreceptor transcription factor targets in mouse models of retinal degeneration: new findings and challenges.,” *Vis. Neurosci.*, vol. 22, pp. 575–586, 2005.

[109] “IDF Diabetes Atlas | Tenth Edition.” [Online]. Available: https://diabetesatlas.org/. [Accessed: 11-Jun-2022].

[110] J. M. Baena-Díez *et al.*, “Risk of Cause-Specific Death in Individuals With Diabetes: A Competing Risks Analysis,” *Diabetes Care*, vol. 39, no. 11, pp. 1987–1995, Nov. 2016.

[111] M. Mambiya *et al.*, “The Play of Genes and Non-genetic Factors on Type 2 Diabetes,” *Front. Public Heal.*, vol. 7, p. 349, Nov. 2019.

[112] R. B. Prasad and L. Groop, “Genetics of Type 2 Diabetes—Pitfalls and Possibilities,” *Genes (Basel).*, vol. 6, no. 1, p. 87, Dec. 2015.

[113] J. H. Cho, J. W. Kim, J. A. Shin, J. Shin, and K. H. Yoon, “β-cell mass in people with type 2 diabetes,” *J. Diabetes Investig.*, vol. 2, no. 1, p. 6, Jan. 2011.

[114] Y. J. Park and M. Woo, “Pancreatic β cells: Gatekeepers of type 2 diabetes,” *J. Cell Biol.*, vol. 218, no. 4, pp. 1094–1095, Apr. 2019.

[115] E. Ahlqvist, T. S. Ahluwalia, and L. Groop, “Genetics of Type 2 Diabetes,” *Clin. Chem.*, vol. 57, no. 2, pp. 241–254, Feb. 2011.

[116] A. Mahajan *et al.*, “Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps,” *Nat. Genet. 2018 5011*, vol. 50, no. 11, pp. 1505–1513, Oct. 2018.

[117] M. P. Keller *et al.*, “The Transcription Factor Nfatc2 Regulates β-Cell Proliferation and Genes Associated with Type 2 Diabetes in Mouse and Human Islets,” *PLOS Genet.*, vol. 12, no. 12, p. e1006466, Dec. 2016.

[118] P. Machanick and T. L. Bailey, “MEME-ChIP: motif analysis of large DNA datasets,” *Bioinformatics*, vol. 27, no. 12, pp. 1696–1697, Jun. 2011.

[119] G. Yu, L. G. Wang, and Q. Y. He, “ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization,” *Bioinformatics*, vol. 31, no. 14, pp. 2382–2383, Jul. 2015.

[120] T.-K. Kim *et al.*, “Widespread transcription at neuronal activity-regulated enhancers,” *Nature*, vol. 465, no. 7295, pp. 182–187, May 2010.

[121] J. S. Tessem *et al.*, “Nkx6.1 regulates islet β-cell proliferation via Nr4a1 and Nr4a3 nuclear receptors,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 111, no. 14, pp. 5242–5247, 2014.

[122] Y. Xu *et al.*, “Prolactin-stimulated survivin induction is required for beta cell mass expansion during pregnancy in mice,” *Diabetologia*, vol. 58, no. 9, pp. 2064–2073, Sep. 2015.

[123] X. Wu *et al.*, “Survivin Is Required for Beta-Cell Mass Expansion in the Pancreatic Duct-Ligated Mouse Model,” *PLoS One*, vol. 7, no. 8, p. e41976, Aug. 2012.

[124] X. Wu *et al.*, “Perinatal survivin is essential for the establishment of pancreatic beta cell mass in mice,” *Diabetologia*, vol. 52, no. 10, pp. 2130–2141, Oct. 2009.

[125] S. C. J. Parker *et al.*, “Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, no. 44, pp. 17921–17926, Oct. 2013.

[126] J. Heijman, N. Voigt, S. Nattel, and D. Dobrev, “Cellular and molecular electrophysiology of atrial fibrillation initiation, maintenance, and progression,” *Circ. Res.*, vol. 114, no. 9, pp. 1483–1499, Apr. 2014.

[127] K. Ihara and T. Sasano, “Role of Inflammation in the Pathogenesis of Atrial Fibrillation,” *Front. Physiol.*, vol. 13, p. 654, Apr. 2022.

[128] X. Zhou and S. C. Dudley, “Evidence for Inflammation as a Driver of Atrial Fibrillation,” *Front. Cardiovasc. Med.*, vol. 7, p. 62, Apr. 2020.

[129] K. Nishida and S. Nattel, “Atrial fibrillation compendium: historical context and detailed translational perspective on an important clinical problem,” *Circ. Res.*, vol. 114, no. 9, pp. 1447–1452, Apr. 2014.

[130] L. C. Weng *et al.*, “Genetic Predisposition, Clinical Risk Factor Burden, and Lifetime Risk of Atrial Fibrillation,” *Circulation*, vol. 137, no. 10, pp. 1027–1038, Mar. 2018.

[131] E. Lee *et al.*, “Mortality and causes of death in patients with atrial fibrillation: A nationwide population-based study,” *PLoS One*, vol. 13, no. 12, Dec. 2018.

[132] K. Nishida and S. Nattel, “Atrial fibrillation compendium: Historical context and detailed translational perspective on an important clinical problem,” *Circ. Res.*, vol. 114, no. 9, pp. 1447–1452, Apr. 2014.

[133] I. E. Christophersen *et al.*, “Familial aggregation of atrial fibrillation: A study in danish twins,” *Circ. Arrhythmia Electrophysiol.*, vol. 2, no. 4, pp. 378–383, Aug. 2009.

[134] I. E. Christophersen and P. T. Ellinor, “Genetics of atrial fibrillation: from families to genomes,” *J. Hum. Genet. 2016 611*, vol. 61, no. 1, pp. 61–70, May 2015.

[135] C. Roselli, C. Roselli, M. Rienstra, P. T. Ellinor, and P. T. Ellinor, “Genetics of Atrial Fibrillation in 2020,” *Circ. Res.*, vol. 127, pp. 21–33, Jun. 2020.

[136] I. E. Christophersen *et al.*, “Fifteen Genetic Loci Associated With the Electrocardiographic P WaveCLINICAL PERSPECTIVE,” *Circ. Cardiovasc. Genet.*, vol. 10, no. 4, p. e001667, Aug. 2017.

[137] B. Laforest *et al.*, “Atrial fibrillation risk loci interact to modulate Ca2+-dependent atrial rhythm homeostasis,” *J. Clin. Invest.*, vol. 129, no. 11, pp. 4937–4950, Nov. 2019.

[138] N. R. Tucker and P. T. Ellinor, “Emerging directions in the genetics of atrial fibrillation.,” *Circ. Res.*, vol. 114, no. 9, pp. 1469–82, Apr. 2014.

[139] B. G. Bruneau *et al.*, “A murine model of Holt-Oram syndrome defines roles of the T-box transcription factor Tbx5 in cardiogenesis and disease.,” *Cell*, vol. 106, no. 6, pp. 709–21, Sep. 2001.

[140] B. G. Bruneau *et al.*, “Chamber-specific cardiac expression of Tbx5 and heart defects in Holt-Oram syndrome,” *Dev. Biol.*, vol. 211, no. 1, pp. 100–108, Jul. 1999.

[141] A. V. Postma *et al.*, “A gain-of-function TBX5 mutation is associated with atypical Holt-Oram syndrome and paroxysmal atrial fibrillation,” *Circ. Res.*, vol. 102, no. 11, pp. 1433–1442, Jun. 2008.

[142] W. Dai *et al.*, “A calcium transport mechanism for atrial fibrillation in Tbx5-mutant mice,” *Elife*, vol. 8, Mar. 2019.

[143] W. Luo *et al.*, “Targeted ablation of the phospholamban gene is associated with markedly enhanced myocardial contractility and loss of beta-agonist stimulation.,” *Circ. Res.*, vol. 75, no. 3, pp. 401–409, 1994.

[144] A. Ventura *et al.*, “Restoration of p53 function leads to tumour regression in vivo,” *Nat. 2006 4457128*, vol. 445, no. 7128, pp. 661–665, Jan. 2007.

[145] M. Kopf *et al.*, “Impaired immune and acute-phase responses in interleukin-6-deficient mice,” *Nature*, vol. 368, no. 6469, pp. 339–342, 1994.

[146] Y. Zhou *et al.*, “Metascape provides a biologist-oriented resource for the analysis of systems-level datasets,” *Nat. Commun.*, vol. 10, no. 1, Dec. 2019.

[147] C. Y. Loh, A. Arya, A. F. Naema, W. F. Wong, G. Sethi, and C. Y. Looi, “Signal transducer and activator of transcription (STATs) proteins in cancer and inflammation: Functions and therapeutic implication,” *Front. Oncol.*, vol. 9, no. FEB, p. 48, 2019.

[148] J. N. Ihle, “The Stat family in cytokine signaling,” *Curr. Opin. Cell Biol.*, vol. 13, no. 2, pp. 211–217, Apr. 2001.

[149] Z. Harhous, G. W. Booz, M. Ovize, G. Bidaux, and M. Kurdi, “An Update on the Multifaceted Roles of STAT3 in the Heart,” *Front. Cardiovasc. Med.*, vol. 6, p. 150, Oct. 2019.

[150] B. Dawn *et al.*, “IL-6 plays an obligatory role in late preconditioning via JAK–STAT signaling and upregulation of iNOS and COX-2,” *Cardiovasc. Res.*, vol. 64, no. 1, p. 61, Oct. 2004.

[151] T. King, R. S. P. Beddington, and N. A. Brown, “The role of the brachyury gene in heart development and left-right specification in the mouse,” *Mech. Dev.*, vol. 79, no. 1–2, pp. 29–37, 1998.

[152] M. J. Pfeiffer *et al.*, “Cardiogenic programming of human pluripotent stem cells by dose-controlled activation of EOMES,” *Nat. Commun. 2018 91*, vol. 9, no. 1, pp. 1–8, Jan. 2018.

[153] O. Kŕianová, L. Mičutková, J. Jeloková, M. Filipenko, E. Sabban, and R. Kvetňanský, “Existence of cardiac PNMT mRNA in adult rats: Elevation by stress in a glucocorticoid-

dependent manner," *Am. J. Physiol. - Hear. Circ. Physiol.*, vol. 281, no. 3 50-3, 2001.

[154] X. Bao *et al.*, "Epinephrine is required for normal cardiovascular responses to stress in the phenylethanolamine N-methyltransferase knockout mouse," *Circulation*, vol. 116, no. 9, pp. 1024–1031, Aug. 2007.

[155] S. N. Ebert and R. P. Thompson, "Embryonic epinephrine synthesis in the rat heart before innervation: association with pacemaking and conduction tissue development," *Circ. Res.*, vol. 88, no. 1, pp. 117–124, Jan. 2001.

[156] S. N. Ebert, Q. Rong, S. Boe, R. P. Thompson, A. Grinberg, and K. Pfeifer, "Targeted insertion of the Cre-recombinase gene at the phenylethanolamine n-methyltransferase locus: A new model for studying the developmental distribution of adrenergic cells," *Dev. Dyn.*, vol. 231, no. 4, pp. 849–858, Dec. 2004.

[157] A. S. Soraya, H. Tali, S. Rona, F. Tom, K. Roy, and A. Ami, "ATF3 expression in cardiomyocytes and myofibroblasts following transverse aortic constriction displays distinct phenotypes," *Int. J. Cardiol. Hear. Vasc.*, vol. 32, Feb. 2021.

[158] Y. Okamoto *et al.*, "Transgenic Mice with Cardiac-Specific Expression of Activating Transcription Factor 3, a Stress-Inducible Gene, Have Conduction Abnormalities and Contractile Dysfunction," *Am. J. Pathol.*, vol. 159, no. 2, pp. 639–650, Aug. 2001.

[159] M. Gilchrist *et al.*, "Systems biology approaches identify ATF3 as a negative regulator of Toll-like receptor 4," *Nature*, vol. 441, no. 7090, pp. 173–178, May 2006.

[160] J. A. Fontes, N. R. Rose, and D. Čiháková, "The varying faces of IL-6: From cardiac protection to cardiac failure," *Cytokine*, vol. 74, no. 1. Academic Press, pp. 62–68, 01-Jul-2015.

[161] A. M. Terrell, P. R. Crisostomo, G. M. Wairiuko, M. Wang, E. D. Morrell, and D. R. Meldrum, "JAK/STAT/SOCS signaling circuits and associated cytokine-mediated inflammation and hypertrophy in the heart," *Shock*, vol. 26, no. 3, pp. 226–234, Sep. 2006.

[162] K. C. Wollert and H. Drexler, "The role of interleukin-6 in the failing heart," *Heart Fail. Rev.*, vol. 6, no. 2, pp. 95–103, 2001.

[163] Y. Feng *et al.*, "The Role of Interleukin-6 Family Members in Cardiovascular Diseases," *Front. Cardiovasc. Med.*, vol. 0, p. 444, Mar. 2022.

[164] R. L. Amdur *et al.*, "Interleukin-6 Is a Risk Factor for Atrial Fibrillation in Chronic Kidney Disease: Findings from the CRIC Study," *PLoS One*, vol. 11, no. 2, Feb. 2016.

[165] G. M. Marcus, M. A. Whooley, D. V. Glidden, L. Pawlikowska, J. G. Zaroff, and J. E. Olgin, "Interleukin-6 and atrial fibrillation in patients with coronary artery disease: data from the Heart and Soul Study," *Am. Heart J.*, vol. 155, no. 2, pp. 303–309, Feb. 2008.

[166] S. N. Psychari, T. S. Apostolou, L. Sinos, E. Hamodraka, G. Liakos, and D. T. Kremastinos, "Relation of elevated C-reactive protein and interleukin-6 levels to left atrial size and duration of episodes in patients with atrial fibrillation," *Am. J. Cardiol.*, vol. 95, no. 6, pp. 764–767, Mar. 2005.

[167] P. E. Lazzerini *et al.*, “Systemic Inflammation Rapidly Induces Reversible Atrial Electrical Remodeling: The Role of Interleukin-6–Mediated Changes in Connexin Expression,” *J. Am. Heart Assoc.*, vol. 8, no. 16, Aug. 2019.

[168] E. Dirkx, P. A. da Costa Martins, and L. J. De Windt, “Regulation of fetal gene expression in heart failure,” *Biochim. Biophys. Acta - Mol. Basis Dis.*, vol. 1832, no. 12, pp. 2414–2424, Dec. 2013.

[169] M. Rajabi, C. Kassiotis, P. Razeghi, and H. Taegtmeyer, “Return to the fetal gene program protects the stressed heart: a strong hypothesis,” *Heart Fail. Rev.*, vol. 12, no. 3–4, pp. 331–343, Dec. 2007.

[170] J. Liao *et al.*, “Interleukin-6-Mediated-Ca2+ Handling Abnormalities Contributes to Atrial Fibrillation in Sterile Pericarditis Rats,” *Front. Immunol.*, vol. 12, p. 5446, Dec. 2021.

[171] Z. Tao *et al.*, “Dexamethasone inhibits regeneration and causes ventricular aneurysm in the neonatal porcine heart after myocardial infarction,” *J. Mol. Cell. Cardiol.*, vol. 144, pp. 15–23, Jul. 2020.

[172] A. Hafezi-Moghadam *et al.*, “Acute cardiovascular protective effects of corticosteroids are mediated by non-transcriptional activation of endothelial nitric oxide synthase,” *Nat. Med.*, vol. 8, no. 5, p. 473, 2002.