NORTHWESTERN UNIVERSITY

Uncovering modifiers of cardiomyopathy in the noncoding genome

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Biochemistry and Molecular Genetics

By

Anthony M. Gacita

EVANSTON, ILLINOIS

JUNE 2020

ABSTRACT

Heart failure due to genetic cardiomyopathy is associated with a range of phenotypic expression. The studies in this body of work interrogated the role of noncoding variation in modifying cardiomyopathy phenotypes. We used cap analysis of gene expression in heathy and failed left ventricles to define the regulatory environment of heart failure. By combining our data with publicly-available datasets, we identified enhancers regulating the cardiomyopathy genes, *MYH7* and *LMNA*. We conducted functional validation of enhancer regions in induced pluripotent stem cell derived cardiomyocytes. To overcome technical challenges in these cells, we developed a multigene qPCR normalization panel. Our findings implicated a super enhancer in the switch of *MYH6* and *MYH7* expression. Sequence variants within transcription factor binding sites were shown to modify enhancer function. We extended our methodology genomewide using a computational pipeline and identified rs875908, which is 2KB 5' of *MYH7*. This variant disrupts the function of an overlapping an *MYH7* enhancer. rs875908 also correlated with longitudinal clinical features of cardiomyopathy in a biobank with clinical imaging and genetic data. Our findings indicate that noncoding variation is phenotypically relevant and may have clinical utility.

ACKNOWLEDGEMENTS

Thank you to my primary mentor Dr. Elizabeth McNally for her mentorship and support. Thank you to Dr. Marcelo Nobrega for his mentorship and willingness to share integral data. A thanks to all members of the McNally lab, past and present, who made this thesis possible. A special thanks to Dr. Andy Vo for starting me on my computational journey and Dr. Eugene Wyatt for teaching me cell culture skills. Excellent technical support was provided by Patrick Page and Katherine Fallon. Important insights and extensive troubleshooting support were provided by Dr. David Barefield and Dr. Mattia Quattrocelli. Thank you to the Northwestern Medical Scientist in Training Program that has supported me during my PhD. Finally, thank you to the patients who participated in my studies by providing samples.

LIST OF ABBREVIATIONS

ATAC-seq	Assay for Transposase- Accessible Chromatin using sequencing
CAGE-seq	Cap Analysis of Gene Expression using sequencing
ChIP-seq	Chromatin Immunoprecipitation followed by sequencing
CIRSPR	Clustered regularly interspaced short palindromic repeats
CRISPRi	CRISPR- interference
CRISRPa	CRISPR- activation
cTnT	Cardiac troponin T
CTSS	CAGE-transcriptional start site
CV	Coefficient of Variation
CPM	Counts per Million
DCM	Dilated Cardiomyopathy
EHT	Engineered Heart Tissue
EKG	Electrocardiogram
EMV	Enhancer modifying variant
ENCODE	Encyclopedia of DNA Elements
eQTL	Expression quantitative trait locus
eRNA	Enhancer RNA
gnomAD	Genome Aggregation Database
GO	Gene ontology
GRO-seq	Global run-on sequencing
GTEx	Genotype-Tissue Expression project
GWAS	Genome-Wide Association Study
HCM	Hypertrophic Cardiomyopathy
iCM/IPSC-CM	Induced Pluripotent Stem Cell derived cardiomyocyte
iPSC	Induced Pluripotent Stem Cell
IQR	Interquantile range
LV	Left Ventricle
LVIDd	Left Ventricular Internal Diameter during diastole
LVPWd	Left Ventricular Posterior Wall thickness during diastole
MHC	Myosin Heavy Chain
MPRA	Massively Parallel Reporter Assay
nAnT-iCAGE-seq	No-Amplification non-Tagging CAGE libraries for Illumina sequencing
pcHi-C	Promoter-capture Hi-C
STARR-seq	Self-Transcribing Active Regulatory Region Sequencing
TAD	Topologically associated domain
TF	Transcription factor
TPM	Tags per Million
TSS	Transcription Start Site

DEDICATION

To my friends, family, and partner who supported me through the good times and the difficult ones. I would also like to dedicate this work to the patients who voluntarily participated in my studies. My hope is that one day that this work can repay your kindness and improve the care of all patients with cardiomyopathy.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

Chapter 1. Introduction

Chapter 2. Enhancer and promoter usage in the normal and failed heart

Chapter 3. Integrated epigenomic analysis identifies enhancer modifying variants linked to cardiomyopathy genes

Chapter 1.

Introduction

I. Prologue

Mendelian inheritance was first described in pea plants where traits were inherited independently in defined ratios. In humans, inheritance of mendelian disease can follow a more complex pattern. Genetic variants are inherited in defined ratios, but many mutations demonstrate incomplete penetrance or variable expressivity. The phenotypic expression of a pathogenic mutation is driven by complex interactions between an individual's genetic background and environmental exposures. To improve the clinical utility of genetic information, we need a more mechanistic understanding behind mutation expression. In this body of work, I study the effect of genetic background in the expression of cardiomyopathy causing mutations

Heart failure is a clinical syndrome that leads to significant morbidity and mortality. One important cause of heart failure is cardiomyopathy, which refers to a group of disorders where there is an intrinsic insult to the cardiomyocyte. Mutations in the coding sequence of many genes have been shown to cause cardiomyopathy. A challenge to the clinical management of these patients is phenotypic heterogeneity in mutation carriers. Even within families, where the same pathogenic mutation is causing disease, individuals can have mild or severe phenotypes. Heterogeneity prevents using genetic information for risk stratification and more tailored management, which can improve clinical outcomes. This phenotypic variability is the result of modifiers, which includes genetic and environmental effects. One class of genetic modifiers are additional variants within the noncoding region of the genome. Next generation sequencing techniques have allowed for analysis of the noncoding area of the genome and many cardiomyopathy-relevant datasets are publicly available. The studies performed in this body of work focus on using epigenomic analyses and functional assays to identify and test noncoding cardiomyopathy modifier variants.

II. Gene Expression and Organization of the Human Genome

The human genome can be broadly organized into coding and noncoding regions. The coding region, which contains genes that produce protein coding RNAs, makes up < 2% of the entire genome (2). The other 98% is referred to as noncoding DNA. Initially, the function of noncoding DNA was unclear and was famously referred to as “junk DNA” (3). Years of investigation have determined that the noncoding genome is not completely “junk”, but some regions control the expression of coding regions.

Coding DNA Features. The coding DNA contains genes that are transcribed into RNA which is translated into protein. The process of gene transcription is complex and involves hundreds of proteins. This process has been well reviewed elsewhere (4, 5). Not all sequences within a coding gene code for proteins, as intronic regions can have regulatory functions (6, 7).

Regulatory Noncoding DNA Features. The regulatory features of the noncoding DNA include gene promoter regions and cis-acting regulatory regions. Gene promoter regions are directly upstream of genes and serve as docking sites for RNA polymerase. The gene promoter is bound by many regulatory proteins, including transcription factors (TFs). TFs are proteins that bind specific DNA sequences and recruit additional proteins that modify gene expression. The expression of these TF’s is often time and tissue specific, resulting in specific gene expression patterns. Gene promoters also contact additional DNA sequences through three-dimensional folding of the DNA molecule. These additional DNA sequences are referred to as cis-acting regulatory regions because they are on the same DNA molecule as their target promoter. Cis-acting regulatory regions can be inhibitory (silencers) or activating (enhancers). Enhancer regions can also be grouped together, where many individual enhancers are arranged in tandem. These groups of enhancers are often referred to as “super enhancers” (8). Importantly, TFs also bind cis-acting regulatory regions, which implies a model where gene promoters, enhancers, TFs, and other transcriptional machinery form a three-dimensional complex to regulate gene expression.

An important concept driving enhancer-promoter interactions is that interactions can span great distances in linear space, but still interact in three-dimensional space. These interactions are not without limits, however. Studies of chromatin conformation have indicated that genomic regions are organized into structural domains, which are referred to as topologically associated domains (TADs) (9). TADs are large, on the order of megabases, and promoter-enhancer interactions are typically confined to the same

TAD. Therefore, TAD information can help focus enhancer searches and link enhancers to their promoter targets.

III. Detecting and Testing Human Left Ventricle Enhancer Regions

Enhancer regions are cis-regulatory sequences that interact with gene promoters and drive gene expression. Initially, enhancer region identification was limited to testing specific regions around genes (10, 11). With the explosion of next generation sequencing technologies, it has become possible to assay enhancer activity genome wide. Since enhancer function is tissue and developmentally restricted, assays for left ventricle enhancers much be conducted in relevant cells/tissues. This section will focus on methods for detecting enhancers and review publicly available cardiac datasets. We focus on datasets derived from human tissue, but where human datasets are unavailable, we provide alternatives from animal or cellular models. Datasets are organized into **Table 1.1.**

Cardiac Model Systems. The human left ventricle (LV) is the major chamber responsible for systolic function and cardiac output. The field of cardiovascular genetics lacks an ideal model of the human LV. Many studies have used the mouse LV to study regulatory regions, but genomic and physiologic differences between mice and humans can limit broader applicability (12, 13). Mouse models are genetically tractable and allow for precise developmental staging, which can assay developmentally specific regulatory regions. Cardiomyocyte cell lines are limited in nature. *In vivo*, mature cardiomyocytes are withdrawn from cell cycle. Thus, an immortal cell line is inherently dedifferentiated and has lost many essential elements of the mature cardiomyocyte. The HL-1 cardiomyocyte cell line is derived from a mouse atrial tumor and can be cultured indefinitely. These cells maintain minimal contractile properties and express some cardiac genes (14). However, the atrial and neoplastic nature of these cells may influence studies of LV regulatory regions. There are additional cell lines available including immortalized human ventricular Ac16 cells and iAM1 cells, which are derived from rat atrial cells. Both lines fail to express many mature cardiomyocyte markers.

Human LV tissues can also be assayed for enhancer function. Most human hearts available for research are available as discarded material after heart transplantation, and so represent end-stage failed hearts. Studies in the failed LV may not be relevant to the healthy heart because the failure state has been shown to influence enhancer usage (6, 15). Healthy human LV samples are difficult to obtain, but can be obtained during failed transplants or by endocardial biopsy during surgery. Although it is rare that a completely healthy heart is being subjected to cardiac surgery. Fetal heart samples from aborted tissues have also been studied but have limited access due to ethical considerations and prohibitions on using federal funds for fetal tissue research.

Induced pluripotent stem cell (iPSC) technology offers an alternative to assaying human tissues. Somatic human cells from a variety of sources can be reprogramed into iPSCs (16, 17). IPSCs differentiate into cardiomyocyte-like (iCMs) cells that express many cardiac markers (18). This process modulates the WNT signaling pathway in a manner that mimics human cardiac development. iCMs contract spontaneously and can be formed into tissue-like structures called engineered heart tissues (EHTs) (19). iPSCs are also amenable to genome editing, allowing for studies of cardiac regulatory regions on isogenic backgrounds. However, iCMs are transcriptionally immature with gene expression patterns that are more similar to fetal cardiomyocytes, which implies an immature regulatory system (20) . An additional challenge to using iCMs is significant variability in the differentiation process. Studies that utilize these cells must include controls to normalize cell purity and maturation level across samples, which are often missing.

Methods for Detecting Regulatory Function in the Human Left Ventricle. The most widely used method for detecting enhancer function genome wide is chromatin immunoprecipitation followed by sequencing (ChIP-seq). In ChIP-seq experiments, an antibody directed to a transcription factor or other protein marker enriches for DNA fragments that have been crosslinked to these proteins (or protein modifications). The enriched DNA fragments are sequenced and mapped to genomic regions of interest either specifically or genome wide. Two large scale projects, the Roadmap Epigenomics Project and the Encyclopedia of DNA elements (ENCODE), were formed to generate a repository of ChIP-seq data across healthy human tissues and cells (21, 22). A major advantage of these large projects is consistency in library preparation and informatic processing that aids in data comparison across samples.

In the mammalian genome, DNA is wrapped around protein complexes called histones. Histone tails are common sites of post-translational modifications that can influence the dynamics of DNA-histone interactions. The proteins that bind regulatory sequences create specific post-translational modifications on histone tails. Therefore, ChIP-seq directed towards these histone modifications is a common method for detecting regulatory regions. The correlation between certain modifications and regulatory function (“the histone code”) has been reviewed extensively (23, 24). Promoter regions are mainly marked by H3K4me3, while enhancer regions are marked by both H3K4me1 and H3K27ac. The Roadmap and ENCODE projects include ChIP-seq data for these marks from human left ventricle for adult, child, and fetal samples. These samples are useful as reference epigenomic data, but do not capture population-level diversity. The Roadmap data was generated from male adult (34 years) and male child (3 years). The ENCODE data comes from two adult females (51 and 53 years). Roadmap also includes right ventricle data from the same subjects and data from one fetal (91 days) heart sample.

A. Regulatory Function Datasets							
Region Type	**Target**	**Assay**	**Species**	**Sex/Age**	**Tissue(s)**	**Accession(s)**	**Ref**
I. Histone Modifications							
Promoters	H3K4me3	ChIP-seq	Human	Female/51 Y, 53Y. Male/3Y, 34Y, 91D.	LV, Heart	ENCSR181ATL, ENCSR901SIL, ENCSR377KDN, ENCSR487BEW, ENCSR358RVW	(21, 22)
Enhancers	H3K27Ac	ChIP-seq	Human	Female/51 Y, 53Y. Male/3Y, 34Y	LV	ENCSR702OVJ, ENCSR854OX, ENCSR150QXE, ENCSR557DFM	(22)
Enhancers	H3K4me1	ChIP-seq	Human	Female/51 Y, 53Y. Male/3Y, 34Y, 91D, 105D	LV	ENCSR449FRQ, ENCSR438QZN, ENCSR111WGZ, ENCSR230VEM, ENCSR480OF, ENCSR676ZKW	(21, 22)
II. General Transcription Factor Binding							
Promoters/Enhancers /TAD Boundaries	CTCF	ChIP-seq	Human	Female/51 Y, 53Y.	LV	ENCSR718SDR, ENCSR791AYW, ENCSR544APK	(22)
Enhancers	P300	ChIP-seq	Human	Male/16W, 45Y	LV	GSE32587	(25)
Promoters	POL2A	ChIP-seq	Human	Female/51 Y, 53Y.	LV	ENCSR699ZGH, ENCSR336YRS	(22)
III. Tissue Specific Transcription Factor Binding							

Enhancers/Promoters	GATA4, TBX5	ChIP-seq	Human	2 Female. 2 Male	iPSC-CM	GSE85631	(26)
Enhancers/Promoters	NKX2-5	ChIP-seq	Human	5 Female, 2 Male	iPSC-CM	GSE125540	(27)
Enhancers/Promoters	GATA4, TBX5, NKX2-5, SRF, MEF2A. (Tagged)	ChIP-seq	Mouse	-	HL-1	GSE21529	(28)
Enhancers/Promoters	GATA4, TBX5, NKX2-5, MEF2A/C, SRF, TEAD1 (Tagged)	ChIP-seq	Mouse	E12.5,P42	LV	GSE124008	(29)
Enhancers/Promoters	GATA4, TBX3/5, NKX2-5	ChIP-seq	Mouse	Adult	Heart	GSE35151	(30)
IV. Open Chromatin							
Enhancers/Promoters	Open Chromatin	DNase-Seq	Human	Female/ 53Y.Male/27Y, 35Y.	LV, Heart	ENCSR070CMW ,ENCSR000EJQ	(21, 22) *
Enhancers/Promoters	Open Chromatin	ATAC-Seq	Human	Female/51 Y, 53Y.	LV, iPSC-CM	ENCSR117PYB, ENCSR851EBF, GSE85330, E-MTAB-8983	(22, 31-33)
V. eRNA Expression							
Enhancers	eRNA Expression	CAGE-Seq	Human	Female/73 Y,92Y,76Y, 26Y. Male/62Y,47Y,20Y,16 Y,54Y	LV, Heart	FANTOM consortium webpage, GSE147236	(6, 34)
Enhancers	eRNA Expression	GRO-Seq	Mouse	9W	Heart	GSE57926	(35)
B. Regulatory Target Datasets							
Enhancers/Promoters /TAD Boundaries	3D Chromatin Interactions	Hi-C	Human	-	LV, iPSC-CMs	GSE58752, E-MTAB-6014, GSE116862	(33, 36, 37)
Enhancers/Promoters	3D Promoter-Enhancer Interactions	pcHi-C	Human	-	LV, iPSC-CMs	GSE86189, GSE100720, E-MTAB-6014,	(37-39)

Table 1.1 Available epigenomic datasets identifying regulatory regions in cardiomyopathy-relevant systems. * Roadmap contains many fetal DNase-seq samples from multiple ages. LV, left ventricle. IPSC-CM, induced pluripotent stem cell derived cardiomyocyte

ChIP-seq can also target DNA binding proteins that bind specifically to regulatory regions. ENCODE contains ChIP-Seq data for two of these proteins, CTCF and POLR2A. CTCF plays an important role in the three-dimensional structure of chromatin and is enriched at the boundaries of topologically associated domains (TADs) and promoter-enhancer interactions (40). POLR2A is the largest subunit of the RNA polymerase II complex, which binds promoter regions to transcribe mRNAs. The ENCODE project includes human left ventricle ChIP-seq for these two proteins. Another protein, p300, acts as a histone acetyltransferase and binds to enhancer regions. ChIP-seq for p300 from human left ventricle tissue is available from a male fetal (16 weeks) heart and an adult (45 years) failing heart (25).

Tissue specific transcription factors (TFs) also bind to enhancer and promoter regions. There are many TFs critical for cardiac development and maintenance, including *GATA4*, *TBX3/5*, *NKX2-5*, *MEF2*, *HAND*, and *SRF* (41). Large projects like ENCODE and Roadmap that study multiple tissue types usually do not include ChIP-seq datasets targeting tissue specific TFs. Another challenge is the technical difficulty in having high quality, high affinity antibodies that can detect TFs in the nuclei of the myocardium. As a result, there are no datasets available for cardiac TF ChIP-seq derived from human heart tissues. The only human datasets available are ChIP-seq data using antibodies to *GATA4*, *TBX5,* and *NKX2-5* derived from iCMs (26) (27). TF ChIP-seq data is also available in murine model systems, but may only be appropriate for highly conserved sites as signals must be lifted over to the human genome. One study examined *GATA4*, *NKX2-5*, *TBX5*, *SRF* and *MEF2A* binding sites in HL-1 cardiomyocytes by over-expressing epitope tagged versions of TFs (28). A similar approach was used to map the same transcription factors in the fetal and adult mouse ventricles (29). Over-expression systems may have low specificity as TF’s at super-physiologic levels may occupy sites that TFs at endogenous levels would not. Untagged endogenously-expressed *TBX3*, *NKX2-5* and *GATA4* were also assessed in the wildtype mouse heart (30). Intersecting data from these cell line and murine sources can provide evidence of enhancer function, but additional studies in human left ventricle tissue are needed.

An important feature of tissue specific TFs is their affinity for binding a particular DNA sequence motif, referred as a TF binding site. These motifs are usually short (~10bp) and contain both highly conserved and degenerate positions. Genome wide analysis of a TF motif identifies potential binding sites, but these algorithms have low specificity. However, binding motifs that overlap ChIP-seq peaks combine informatic and experimental data are, therefore, more sensitive to implicate a particular DNA sequence required for TF binding at that location. Motifs are determined by analyzing ChIP-seq binding data for overrepresented sequences. HOMER contains the binding motifs for many of the core cardiac TFs (42).

Another mark for regulatory regions stems from the idea that active regulatory regions must be accessible by proteins. Therefore, regulatory regions are euchromatic and “open”. DNase-seq uses a bacterial DNAse to selectively cut open chromatin, which can be isolated and sequenced (43). DNase-seq from heart tissues are available from Roadmap and ENCODE. Roadmap contains 12 human DNase-seq datasets from fetal tissues (both sexes, various ages) and one from a male child (3 years). ENCODE contains two additional fetal datasets and two adult male hearts (27 and 35 years old). A newer method, Transposase-Accessible Chromatin followed by sequencing (ATAC-seq), uses a transposase to insert sequencing adapters into open chromatin. The location of adapters can be identified with sequencing, which gives a genome wide view of open chromatin. ATAC-seq is faster and requires much less input material than DNase-seq (44). ENCODE includes two ATAC-seq datasets from female human left ventricles (51 and 53 years). There are also ATAC-seq datasets available from iCMs (32, 33, 45).

Much of the noncoding genome is transcribed into RNAs. One class of these RNAs originate from enhancer regions and have been termed enhancer RNAs (eRNAs). These eRNAs are short, 5’-methyl capped, not polyadenylated, and transcribed in a bidirectional pattern (46-48). eRNAs are unstable and quickly degraded, which makes them difficult to detect. The function of eRNAs are a major area of investigation. The levels of eRNAs originating from an enhancer region has been shown to correlate with target gene expression level, implying a regulatory function (49, 50). Whatever their function, eRNA expression can be used to map active enhancer regions genome wide. Given their quick turnover, methods that assay active transcription will be most sensitive. Global run-on sequencing (GRO-seq) assays active transcription by selectively sequencing transcripts that incorporated a labeled nucleotide. There is a GRO-seq dataset derived from mouse hearts available (35). There is also GRO-seq data available in a porcine model of cardiac ischemia (45). Additional studies using GRO-seq in human cells and tissue are needed to further our knowledge of eRNAs in the human heart. Other methods to detect eRNAs enrich the steady state RNA population for eRNAs. Cap analysis of gene expression (CAGE-seq) enriches for all RNAs that have a 5’methyl cap using biotinylation and a streptavidin pull down (51). This method enriches for any RNA transcribed by the RNA polymerase II complex, including eRNAs and mRNAs. Therefore, deep sequencing is required to achieve adequate signal from eRNAs. The FANTOM consortium contains CAGE-seq data from many human cells and tissues, including human heart. FANTOM used heart RNA that was commercially available and includes one adult healthy left ventricle (female, 73 years), one post-infarction adult heart (female, 92 years), and two pooled samples of multiple adult and fetal hearts (34). While this data is useful, the major challenge with FANTOM data is the lack of sequencing depth as most of the heart samples have < 10 million mapped tags. As part of this thesis, we generated CAGE-seq data from 3 healthy heart samples and four cardiomyopathic failed hearts with high read depth and this dataset is publicly available (6).

Methods for Determining Left Ventricle Enhancer Targets. The above methods are useful for determining if a genomic region has enhancer activity, but do not provide information about an enhancer's target gene. Some studies simply choose the nearest gene in linear space as an enhancer target, but it is clear that this assumption is not true in many cases (37). Chromatin conformation studies cross link regions of the genome together and allow for direct detection of interactions. Targeted experiments like 3C and 4C require *a priori* knowledge of an interacting fragment and can be useful for targeted questions. Hi-C is unbiased and utilizes next generation sequencing to detect all interactions. There is a single Hi-C dataset available from the human left ventricle (36), and an additional 4 left ventricle samples may be available soon (52). There are also datasets available in iCMs (33, 37). Hi-C data can be used to define TAD boundaries, which can confine an enhancer's target gene to the same TAD.

Hi-C data provides information about all interactions, but we are often most interested in enhancer-promoter interactions. Hi-C libraries can be enriched for interaction fragments that include gene promoter regions in a process called promoter-capture Hi-C (pcHi-C) (53). There is a pcHi-C dataset from human left ventricle available (38) and two datasets from iCMs (37, 39). These datasets are very valuable and can be used to focus enhancer searches to regions that interact with particular gene promoters.

An additional method for predicting enhancer targets takes advantage of common sequence variation. Using large datasets of genotyped RNA-seq data, noncoding variants can be statistically correlated with gene expression in a process called expression quantitative trait mapping (eQTL). Multiple eQTL datasets in the human heart are available (54, 55). Given the genome-wide nature of these studies, many signals may be lost to failing to meet genome wide significance after multiple testing correction. Therefore, it is important to view associations individually with targeted hypotheses about variants and target gene associations.

Methods for Validating Left Ventricle Enhancer Regions. Once potential regulatory regions have been identified, additional experiments are needed to validate those predictions. As mentioned above, the tissue specificity of enhancer regions requires that validation studies be done in relevant experimental systems. To test human LV enhancers, mouse LVs and iCMs are commonly used model systems.

The simplest test of a putative enhancer region is a reporter assay. In these assays, the candidate enhancer is inserted into plasmid DNA upstream of a minimal promoter driving expression of a reporter gene. Any reporter gene can be used, but firefly luciferase is often used because it is easily detectable and quantitative (56). This plasmid is introduced into a cellular system where the plasmid is transcribed episomally by cell TFs and transcription machinery. DNA sequences with enhancer function will increase expression of the reporter gene when compared to an empty plasmid or other non-enhancer genomic region. Enhancer containing plasmids can also be injected into mouse embryos to asses tissue expression patterns. The VISTA enhancer database contains results from > 3,000 enhancer regions tested with this transgenic mouse strategy (57). However, reporter assays are limited by their low throughput nature, relatively high expense, and technical variability.

A newer class of reporter assays aim to increase enhancer testing throughput. Massively parallel reporter assays (MPRA) can test the activity of thousands of regulatory regions in a single experiment. In MPRA experiments, candidate regulatory regions are synthesized with a unique barcode. This library is cloned into a plasmid and an open reading frame (ex. GFP) is inserted between the candidate enhancer and barcode. When this library is transfected into cells, regulatory regions with enhancer activity will increase expression of the inserted open reading frame and barcode. Next generation sequencing can be used to quantify the relative quantities of barcodes, which is directly proportional to the strength of the candidate regulatory region (58). This process is limited by the size of the synthesized regulatory regions (~150bp) and the ability to adequately deliver the library to the system of interest.

One study used an AAV vector to deliver the library of interest into the mouse heart (29). The study design resembled self-transcribing active regulatory region sequencing (STARR-seq). In STARR-seq, the candidate regulatory regions are cloned downstream of an open reading frame (59). Active enhancer regions will interact with the upstream promoter and drive transcription of the open reading frame and their own sequence. Therefore, the amount of RNA matching the enhancer region sequence is directly proportional the enhancer’s activity. With this technology, barcodes are no longer necessary, which significantly decreases the cloning complexity. The candidate enhancer sequences can also be derived from cells or synthesized. Future MPRA or STARR-seq experiments in iCMs or mouse hearts have great potential to inform our knowledge of human left ventricle enhancers.

While the above methods can provide evidence of enhancer function, they provide no information on that enhancer’s target gene. CRISPR/Cas9 technology offers multiple methods for detecting enhancer target genes. IPSCs are amenable to CRSIPR/Cas9 mediated gene editing (60). The simplest experiment is to delete a candidate enhancer region in IPSCs and measure its effect on gene expression in iCMs. One study used this technique to delete an intronic regulatory region in *PHACTR1* in IPSCs, which is a gene linked to myocardial infarction risk by genome wide association studies (61). At present, iCMs are the best system to test enhancer deletions in a human context, but requires that the target gene be expressed in iCMs. CRISPR/Cas9 technology also offers two related methods for detecting enhancer targets, CRISPR-activation (CRISPRa) and CRISPR-interference (CRISPRi). In these methods, a catalytically inactive cas9 (“dead cas9”) is linked to an activating or inhibitory protein (62, 63). The dead cas9 will recruit the linked protein to specific genomic regions determined by the guide RNA. These methods are complementary and can be done in high throughput formats to test many different guides/candidate enhancers at once. iCMs are well positioned for CRISPRa and CRISPRi studies to define enhancer targets.

IV. Genetic Cardiomyopathy and Phenotypic Heterogeneity

Heart failure is a clinical syndrome caused by a reduction in cardiac output. Cardiomyopathy causes morphologic changes to the heart, which can reduce heart function and cardiac output. Genetic mutations, largely in the coding regions of genes, have been linked to multiple autosomal dominant forms of cardiomyopathy including dilated, hypertrophic, restrictive and arrhythmogenic (64, 65). In dilated cardiomyopathy (DCM), the left ventricular chamber progressively dilates and the ventricular wall thins. A dilated left ventricle is unable to efficiently eject blood during systole. In hypertrophic cardiomyopathy (HCM), the ventricular wall progressively thickens, which decreases the chamber size. Early in disease progression, hypertrophic hearts may be hyperdynamic and able to maintain cardiac output. As the walls get thicker, however, the chamber becomes unable to fill appropriately during diastole and cardiac output falls. Wall thickening can also cause the mitral valve to move during systole and block the ventricular outflow track. In both DCM and HCM, arrhythmias can further alter heart function. Both of these disease processes are associated with significant burden (66, 67).

Genes linked to Cardiomyopathy. With the implementation of next generation sequencing and clinical use of genetic testing, around a hundred genes have been associated with genetic forms of cardiomyopathy, although not all are “high confidence” genes or variants. DCM is a genetically heterogenous disease. Cardiomyopathy causing mutations have been found in proteins that form many structures of the cardiomyocyte, including the sarcomere, plasma and nuclear membrane, the nucleus, and the desmosome. In individuals with a clinical picture consistent with genetic DCM, targeted gene sequencing can identify a casual pathogenic variant in ~30-40% of cases (68). Of those mutations identified, 20% are truncating mutations in titin (*TTN*) (69). *TTN* is the largest human gene and encodes a sarcomeric protein that stretches from the M line to the Z disk and provides stability and elasticity to the sarcomere. ~8% of familial DCM mutations are in the intermediate filament protein, lamin A/C (*LMNA*). This protein forms a lattice on the inner nuclear membrane and has been implicated in a variety of functions and phenotypes (70, 71). Importantly, mutations in *LMNA* are associated with an increased risk of arrhythmias and conduction disease. The remaining DCM-associated mutations are within many genes including the sarcomeric genes, *MYH6* and *MYH7*.

The genetics of HCM are less heterogenous. Broadly, HCM mutations can be divided into sarcomeric and non-sarcomeric causes with clear phenotypic differences between the two (72). 60% of individuals with HCM have mutations in one of eight sarcomeric proteins including *MYH7*, *MYPBC3*, *TNNT2*, *TNNI3*, *TPM1*, *MYL2*, *MYL3*, and *ACTC1* (73). Most mutations are either nonsense *MYBPC3* or missense *MYH7* variants. Non-sarcomeric causes of HCM are rare, but include mutations in *PRKAG2* and *LAMP2*. Loss of function mutations in *MYBPC3* are thought the control the contraction dynamics of the sarcomere and contribute to the hypercontractility seen in HCM (74). Missense mutations in *MYH7* are more difficult to interpret, but some mutations, especially in the myosin head domain, affect the relaxed state of the myosin molecule (75).

Phenotypic Variability in Genetic Cardiomyopathy. Genetic cardiomyopathy is characterized by significant phenotypic heterogeneity. Mutations display an age-related onset, variable penetrance and range of expressivity. The same mutation within families can present with multiple different phenotypes. **Figure 1.1** shows two example family pedigrees where cardiomyopathy mutations exhibit variable phenotypes. DCM-related mutations are also seen in the population at higher rates than would be expected based on disease prevalence, which implies that many mutations have low penetrance (76). A large study of kindreds with identical *MYH7* mutations saw differences in the rate of sudden cardiac death and syncope (77). Mutations within the same gene can also result in different morphological remodeling (i.e. DCM vs HCM) (78, 79) .

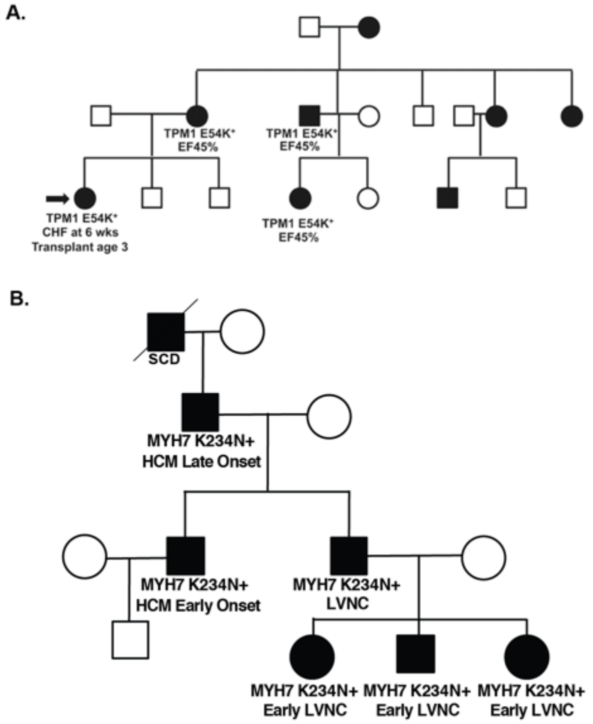


Figure 1.1. Pedigrees demonstrating phenotypic heterogeneity in genetic cardiomyopathy. A. Pedigree from (1) of a family with a tropomyosin I (*TPM1*) mutation demonstrating variable penetrance. **B.** Pedigree of a family with a *MYH7* missense mutation demonstrating variable expressivity and variable remodeling phenotypes. EF, ejection faction. CHF, congestive heart failure. SCD, sudden cardiac death. LVNC, left ventricular non-compaction

The range of phenotypes seen in cardiomyopathy cases is due to modifiers. Environmental exposures and additional genetic mutations may act as modifiers. Modifier coding variants have been identified in the adrenergic receptors, *HSPB7*, *CLCNKA*, *LTBP4, ACE,* and others (80-85). However, the exact mechanism of many of these variants needs further investigation. While multiple coding modifiers of cardiomyopathy phenotypes have been found, noncoding modifiers are just beginning to be investigated.

V. Studies of Noncoding Variation in Cardiac Disease

The noncoding genome refers to intergenic and intronic regions. These regions contain sequences that regulate coding genes. Genome wide association studies (GWAS) have linked common variants within noncoding regions to many cardiac phenotypes (86). In addition, many sub-threshold GWAS hits within noncoding regions may represent true associations (87). This section will provide an overview of studies that have linked noncoding variation to cardiac phenotypes.

Noncoding mutations linked to congenital heart disease. Given the important role that enhancers play in development, it is expected that enhancer variants could result in developmental phenotypes. One study focused on the *TBX5* gene, which encodes a transcription factor with key roles in forelimb and heart development. Many coding mutations in *TBX5* are linked to congenital heart disease and forelimb abnormalities referred to as the Holt-Oram Syndrome. This study identified three enhancers around *TBX5* with cardiac expression patterns. They searched for variation within these enhancers in a large cohort of patients with congenital heart disease. They identified a homozygous variant ~90kb downstream of *TBX5* that was highly conserved across vertebrates and disrupted a transcription factor binding site. Further analysis indicated that this variant reduced enhancer function in the zebrafish heart. Therefore, a noncoding variant was linked to disruption of enhancer function and a resulting congenital defect (88).

Noncoding mutations linked to arrhythmias. Coding mutations in ion channel genes have been linked to inherited arrhythmias. *KCNH2* encodes the voltage-gated potassium channel, and coding mutations result in long QT syndrome. Additionally, GWAS studies have identified common variants outside of *KCNH2*'s coding sequence associated with QT interval. One study used an integrative analysis of epigenomic data to identify candidate enhancers around the *KCNH2* gene (89). They discovered multiple candidates, but a region ~75kb downstream of the *KCNH2* transcriptional start site (TSS) showed the strongest luciferase activity in HL-1 cells, HEK293T cells, and HEPG2, hepatocellular carcinoma cell line. 4C experiments in the murine heart confirmed that this region interacted with the *KCNH2* TSS. Removal of this enhancer region in mice resulted in reduced expression of *KCNH2*. Additionally, a bidirectional RNA was detected originating from this locus. When this RNA was reduced using antisense oligonucleotides in HL-1 cells, *KCNH2* expression was also reduced, indicating that this RNA likely represents a functional eRNA. None of the QT-interval associated GWAS variants directly overlapped this region, but the variants did overlap another candidate region that failed to generate activity in reporter assays, likely due to limitations in the cell culture models used in the study.

Genetic variants near the *HAND1* gene have been implicated in QRS duration (90). Using evolutionary conservation as a guide, a study identified an enhancer ~15kb upstream of the *HAND1* TSS (91). This enhancer showed reporter activity, and when deleted in mice, resulted in reduced *HAND1* expression. Mice with homozygous deletions of this enhancer have electrocardiogram (EKG) abnormalities, including an increased QRS duration. This enhancer also contains two *GATA4* binding sites that are disrupted by GWAS signals. Mice with the minor alleles of both of these variants also have abnormal ventricular conduction systems, matching the phenotype seen in the enhancer deletion (91). This study is an excellent example of using GWAS data to identify functional noncoding variants. However, this method lacks sensitivity as many subthreshold GWAS hits may also be functional or the GWAS signal may be in linkage disequilibrium with the genetic variant(s) responsible for the outcome (87).

Atrial fibrillation is also under significant genetic control. A study that set out to define the regulatory targets of all noncoding GWAS hits associated with atrial fibrillation identified enhancers predicted to regulate *GJA1*, *KCNN3*, and *ZFHX3* (92). The region expected to regulate *GJA4* was 680kb downstream of the *GJA4* TSS. Removing this enhancer in mice showed a significant downregulation of *GJA1*, and the study did not describe the effect any specific variants within this region. A combination of epigenetic datasets was used to predict functional variants within other enhancers, and allele specific activity was detected for 3 of 11 variants assayed in the rat atrial line, iAM1. This study made good use of epigenomic datasets, but the massive number of potentially functional noncoding variants makes it difficult to validate findings on a genome wide scale.

A recent genome wide study assayed GWAS signals associated with EKG traits. Using *NKX2.5* ChIP-seq from iCMs, the authors identified regions of allele specific binding, defined as variants where ChIP-seq reads are biased towards one allele or the other (27). A variant within an intron of the *SSBP3* gene that is associated with P-wave duration was observed to reduce the overlying enhancer's luciferase activity in iCMs. An intronic variant in *CAV1* was described in a GWAS for atrial fibrillation and PR interval (93, 94). This variant reduced the overlying enhancer's luciferase activity in iCMs (27). This study is notable for its use of a human model system, and these findings indicate that key regulatory variants can disrupt transcription factor binding sites. Additionally, this study also highlights the importance of intronic enhancers.

The *SCN5A-SCN10A* locus has been well studied because GWAS studies have multiply associated this locus with EKG traits. *SCN5A* and *SCN10A* encode subunits of voltage-gated sodium channels and *SCN5A* is important for cardiomyocyte depolarization. Coding mutations with *SCN5A* have been linked to the Long QT syndrome and Brugada syndrome (95). In the genome, *SCN5A* and *SCN10A* are organized in tandem. Multiple regulatory regions have been identified in this cluster including one within a *SCN10A* intron that harbors a GWAS variant associated with decreased *SCN5A* expression (96). Multiple enhancer regions have also been identified downstream of *SCN5A*. These enhancers are in close proximity and are likely part of a super enhancer. Deletion of this super enhancer in mice results in not only reduced *SCN5A* expression, but also disrupts the three-dimensional chromatin architecture of the locus. This super enhancer also contains a QRS-duration associated variant that prevents a response to *NKX2-5* and *GATA4* in a cellular reporter assay (97). The organization of the *SCN5A-SCN10A* locus may be a common mechanism for controlling tissue specific genes that are arranged in tandem, including *TBX3*-*TBX5* and *NPPA*-*NPPB* (98, 99).

Noncoding mutations linked to cardiomyopathy. The studies described above focused on congenital heart disease and arrhythmias. Noncoding variation has yet to be substantially linked to heart failure, cardiomyopathy and/or ventricular chamber or wall dimensions. Studies of noncoding variation and arrhythmia phenotypes have been aided by rich GWAS results because EKG measures are easily obtainable quantitative traits with high reproducibility and lower technical variability. GWAS studies of cardiomyopathy phenotypes have mainly relied on echocardiographic measures, but results have been less robust (100). As a result, there are many fewer studies of noncoding variation linked to cardiomyopathy traits. One study assayed a variant upstream of the *MTSS1* gene that has been linked to LV end-diastolic dimensions. The region harboring this variant drove heart expression in a transgenic zebrafish and the minor haplotype had lower activity in a cellular reporter assay. This variant is an eQTL for *MTSS1* and deletion of *MTSS1* in mice lowered ventricular dimensions. Therefore, this variant's GWAS signal likely results from reduced expression of *MTSS1*.

A recent study assayed a GWAS variant for heart failure upstream of *ACTN2* (33). Mutations in *ACTN2* have been linked to various forms of cardiomyopathy (101, 102). Using iCMs, this study showed that open chromatin, H3K27Ac, and H3K4me1 signals arise at the variant site during iCM differentiation. Hi-C data demonstrated an interaction between the overlapping enhancer and the *ACTN2* promoter. Loss of the enhancer region also reduced *ACTN2* expression. Even though this study did not test the effect of the actual enhancer variant, it provides evidence of an enhancer influencing heart failure phenotypes.

VI. Summary

The human genome contains regions that regulate the expression of coding genes. Many techniques exist to detect these regions, predict their targets and validate those predictions. Years of investigation have readily available LV-relevant datasets that are a rich source of information. Cardiomyopathy has a large genetic component and is associated with high levels of phenotypic variation. Variants within the noncoding regions of the genome are thought to play a major role in modifying cardiomyopathy phenotypes. Driven by strong GWAS results, most studies of noncoding variation in the heart have focused on arrhythmia/EKG phenotypes. The impact of noncoding variation on cardiomyopathy phenotypes has been underexplored and is positioned to aid in clinical management of heart failure.

V. THESIS OVERVIEW

This section provides a brief overview of the content in the following chapters.

Chapter 2. Enhancer and promoter usage in the normal and failed human heart.

The failed left ventricle is associated with significant changes in gene expression. I set out to assay the regulatory regions active in both the healthy and failed heart with Cap Analysis of Gene Expression (CAGE-seq). CAGE-seq identified enhancer and promoters active in healthy and failed hearts, providing a left ventricle enhancer and promoter map. Additionally, I identified promoters and enhancers that changed in the failure state, which are attractive therapeutic targets to ameliorate ventricular remodeling and heart failure severity.

Chapter 3. Integrative epigenomic analysis identifies enhancer modifying variants linked to cardiomyopathy genes.

Coding mutations in many gens can result in cardiomyopathy. A well-recognized feature of genetic cardiomyopathy is varying phenotypic expression, which may be due to modifier variants within the noncoding genome. I used over >20 publicly-available heart enhancer datasets to identify enhancer regions regulating the cardiomyopathy genes, *MYH7* and *LMNA*. I validated these enhancer regions with multiple complementary approaches in iCMs. Sequence variants within transcription factor binding sites altered enhancer function. Additional analysis identified a variant upstream of *MYH7* that correlates with *MYH7* expression and a worse cardiomyopathy phenotype over time.

Chapter 4. A transcriptional method for assaying IPSC-derived cardiomyocyte purity and maturity level.

Currently, induced pluripotent stem cell derived cardiomyocytes (iCMs) are the best model for studying human left ventricle enhancer regions. Variable differentiation purity and maturity pose a significant technical challenge when using iCMs. I developed a transcriptional purity/maturity assay that relies on publicly-available iCMs differentiation RNA-seq data. Using this data, I identified a panel of qPCR normalization genes that minimize variability in *MYH6* and *MYH7* expression measurements.

Chapter 2.

Enhancer and promoter usage in the normal and failed human heart

Abstract

The failed heart is characterized by re-expression of a fetal gene program, which contributes to adaptation and maladaptation in heart failure. However, the genomic regulatory changes that contribute to these changes are not well understood. To define genomewide enhancer and promoter use in heart failure, Cap Analysis of Gene Expression (CAGE-seq) was applied to three healthy and four failed human left ventricles to define RNAs associated with both promoters and enhancers. Healthy hearts were derived from donor hearts unused in transplantation and failed hearts were obtained as discarded tissue after transplantation. Integration of CAGE-seq data with RNA sequencing identified a combined ~17,000 promoters and ~1,800 putative enhancers active in healthy and failed human left ventricles. Comparing promoter usage between healthy and failed hearts highlighted promoter shifts which altered amino-terminal protein sequences. Comparing putative enhancer usage between healthy and failed hearts revealed a majority of differentially utilized heart failure enhancers were intronic and primarily localized within the first intron, identifying this position as a common feature associated with tissue-specific gene expression changes in the heart. This dataset defines the dynamic genomic regulatory landscape underlying heart failure and serves as an important resource for understanding genetic contributions to cardiac dysfunction. Additionally, regulatory changes contributing to heart failure are attractive therapeutic targets for controlling ventricular remodeling and clinical progression.

This work is under review at Circulation: Heart Failure:

> **Gacita AM**, Dellefave-Castillo L, Page PGT, Barefield DY, Waserstrom JA, Puckelwartz MJ, Nobrega MA, McNally EM. Enhancer and promoter usage in the normal and failed human heart. *Circulation: Heart Failure* (in review)

Respective Contributions

AMG conducted the analysis and drafted the manuscript. LDC secured patient consent and genotype information. PP assisted with genotyping. JAW provided access to control samples. DYB and MJP provided helpful advice and commentary and assisted with interpretation. MAN and EMM assisted with analysis, writing and editing the manuscript.

Introduction

The failed heart is characterized by reduced function, impaired filling, and altered metabolism, all of which contribute to an inability to meet the body's demands for normal activity. Heart failure is associated with global changes in gene expression and splicing, some of which directly drive pathological and adaptive remodeling.(103) For example, the failed heart shifts its metabolism towards glycolysis, driven in part by gene expression changes.(104) Within the failed human heart, a distinct isoform of myosin heavy chain is expressed,(105, 106) as are alternatively spliced forms of *TNNT2* and *TTN*, encoding troponin T and titin, respectively, and these changes directly modify contractility and compliance.(107-109) In addition, mutations in many of these genes directly lead to cardiomyopathy and heart failure.(65, 110) The regulatory regions responsible for driving normal and pathological gene expression in the heart are incompletely understood. For some genes, specific genetic regulatory regions have been characterized,(29, 111) but comparatively few genomewide analyses have been conducted. Surveys of the cardiac epigenome have been used to infer regulatory regions in the developing mouse heart and embryonic stem cell derived-cardiomyocytes.(112, 113) Using mouse hearts subjected to pressure overload, chromatin conformational state was evaluated to indicate potential regulatory regions active in this setting.(114) However, much less is known about the promoter and enhancer shifts underlying human heart failure.

Transcription factors bind promoters and enhancer sequences, which interact in three-dimensional space to modify gene expression. Estimates of number of active heart enhancers vary from several thousand to tens of thousands depending on the approach used.(25, 115) One assessment of active cardiac enhancers monitored p300/CBP binding sites from one human fetal and one adult failed heart.(25) This analyses evaluated candidate enhancers more than 2.5kb from transcriptional start sites, potentially missing proximal enhancers.(25) Nonetheless, this analysis identified ~5,000 active enhancers in fetal tissue and ~2,000 active enhancers in adult tissue, with approximately half of adult heart enhancers also active in fetal heart, underscoring the importance of developmental enhancers in the adult heart.(25) A similar approach used normal human and mouse hearts integrating p300/CBP binding sites with H3K27ac marks.(115) This integrated approach described more than 80,000 potential heart enhancers.(115) These studies provide a valuable datasets, but do not identify the genomic alterations seen in human heart failure, a condition known to have distinct gene expression.

Gene expression changes can also result from shifts in promoter usage. Alternative promoters are estimated to affect 30-50% of human genes.(116) Alternative promoters may affect the amino-terminal amino acids of proteins and/or the 5'UTR of transcripts, both of which can mediate functional consequences. Alternative promoters can also influence the effect of genetic variants on protein function and thus are vital for accurate variant effect predictions.(117) Despite the potential of broad proteome differences due to alternative promoter usage, a genomewide view of promoter and enhancer shifts in human heart failure is lacking.

Next generation sequencing technologies including Cap-Analysis of Gene Expression (CAGE) make it possible to assay transcriptome usage by determining RNA transcriptional start sites at single base pair resolution.(51) Enhancer regions are transcribed into low-abundance enhancer RNAs (eRNAs) in a bidirectional pattern,(47, 48) contrasting with the unidirectional transcriptional expression seen near gene promoters, which produce high-abundance signals. Because of the precision with which these RNAs can be mapped, it is possible to accurately map enhancer and promoter signals at high resolution. To define alternative promoter and enhancer use in heart failure, we generated CAGE sequence datasets from healthy and failed human left ventricles. CAGE sequencing information was integrated with RNA sequencing from these same samples, to improve sensitivity in detecting low-abundance eRNAs and rarely used gene promoters. We relied on a no-amplification non-tagging CAGE sequencing protocol, which allows for more robust and less biased detection of transcriptional start sites.(51) These data identified unique signatures of housekeeping and tissue specific promoters, as well as a pattern of enhancers within first introns that regulate tissue specific expression. In addition to identifying differential candidate enhancer use in heart failure, we cataloged 129 genes with differential promoter usage in heart failure. These alternative promoters have the potential to encode proteins with unique amino-termini, highlighting potential protein composition shifts in the failed heart.

Methods

Materials, Code and Data Availability. All scripts/code used in this analysis are available upon request. Sequence data has been uploaded to the NCBI-GEO under accession number GSE147236.

RNA-Extraction, Library Preparation, and Sequencing. Healthy and failed left ventricle samples were obtained from failed transplants or as discarded tissue, respectively. Living subjects provided consent. Healthy left ventricular samples were obtained from hearts provided by the Gift of Hope of Illinois and were found to be unsuitable for transplant due to age or prior cardiac surgeries. All patients were declared to be brain dead as the result of cerebral hemorrhage and familial consent was obtained for organ use in research. Tissues were snap-frozen in liquid nitrogen and stored at -80ºC until use. Approximately 50mg of frozen tissue was ground into a fine powder using a mortar and pestle under liquid nitrogen. Ground powder was added to 1mL TRIzol (Invitrogen) containing 250ul of silica zirconium beads. Samples were placed in a bead homogenizer for 1 minute, allowed to cool on wet ice, and centrifuged at 12,000xg to remove any unhomogenized tissue pieces followed by chloroform extraction. Phases were separated by centrifugation and the upper aqueous layer was added to fresh 70% ethanol. The RNA-ethanol mix was used as in input to the Aurum Total RNA Mini Kit (BIORAD). RNA was isolated (including on-column DNAse digestion) following manufacturer’s instructions. Concentration was measured using a NanoDrop spectrophotometer and quality was assessed using an Aligent Bio analyzer. Only RNA extractions with RIN values ≥7 were used. If necessary, RNA extraction was repeated until ~10μg of RNA was obtained.

Custom nAnT-iCAGE-seq (no-Amplification-no-Tagging Illumina Cap Analysis of Gene Expression libraries) libraries were prepared by DNAFORM (Japan) following a previously-described protocol (51). Briefly, 5μg of RNA was reverse transcribed using random primers. 5’-methyl-caps were biotinylated and enriched using streptavidin beads. cDNA was released and sequencing adapters were added using blunt-ended ligation. Following second strand synthesis, the libraries were quantified using qPCR. ~50pM of pooled libraries were loaded into an entire run on the NextSeq 500 (Illumnia) to yield ~400 million total 75bp single end reads (**Table 2.1**).

	CAGE-Seq			RNA-Seq	
Sample	**pM**	**Total reads (75bp SE)**	**Uniquely Aligned Reads (%)**	**Total Reads (150bp PE)**	**Uniquely Aligned Reads (%)**
Healthy 1	8.10	59,865,536	45,191,501 (75.49)	53,274,910	41,113,982 (77.17)
Healthy 2	7.34	56,226,409	44,062,633 (78.37)	58,609,578	44,742,094 (76.34)
Healthy 3	5.07	32,763,023	25,314,481 (77.27)	49,620,253	38,276,453 (77.14)
Heart Failure 1	5.30	39,295,214	30,906,318 (78.65)	49,587,645	38,077,307 (76.79)
Heart Failure 2	8.25	45,584,689	36,344,157 (79.09)	43,147,599	32,813,912 (76.05)
Heart Failure 3	8.03	46,398,662	35,734,514 (77.02)	47,529,437	35,808,836 (75.34)
Heart Failure 4	4.53	22,668,805	18,101,857 (79.85)	40,737,687	31,579,044 (77.52)

Table 2.1. Sequencing Read Yields and Mapping Rates for CAGE-seq and RNA-seq Libraries. *pM*, picomolar. *SE*, single end. *PE*, paired end.

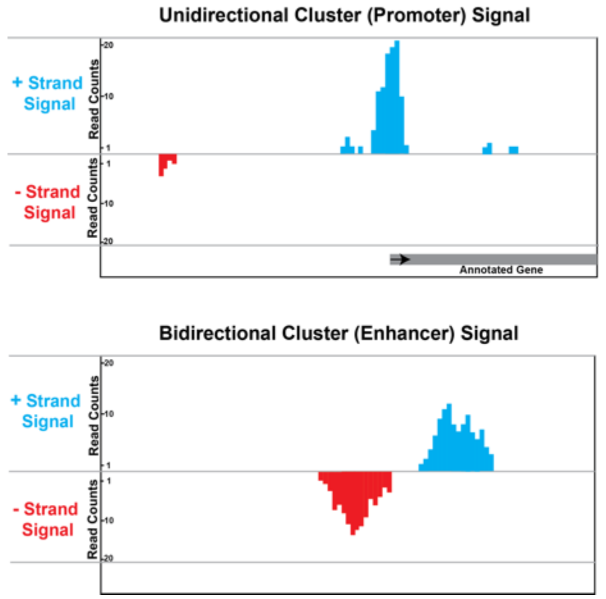


Figure 2.1. Example schematic of unidirectional (top) and bidirectional (bottom) CAGE clusters representing promoters and enhancer regions, respectively. Positive (sense) strand signals are shown in blue and minus (antisense) signals are shown in red.

RNA-seq libraries were prepared using the TruSeq mRNA-seq library preparation kit (Illumina). Libraries were pooled in equimolar amounts and loaded on the HiSeq 4000 (Illumina) to generate ~40 million 150bp paired-end reads/sample.

CAGE-Seq Alignment and Clustering. Raw CAGE-seq reads were checked with fastQC(v0.11.5) and aligned to the human genome (hg19) using STAR (v2.5.2) with default settings.(118) Uniquely aligning reads were inputted into CAGEr and converted into

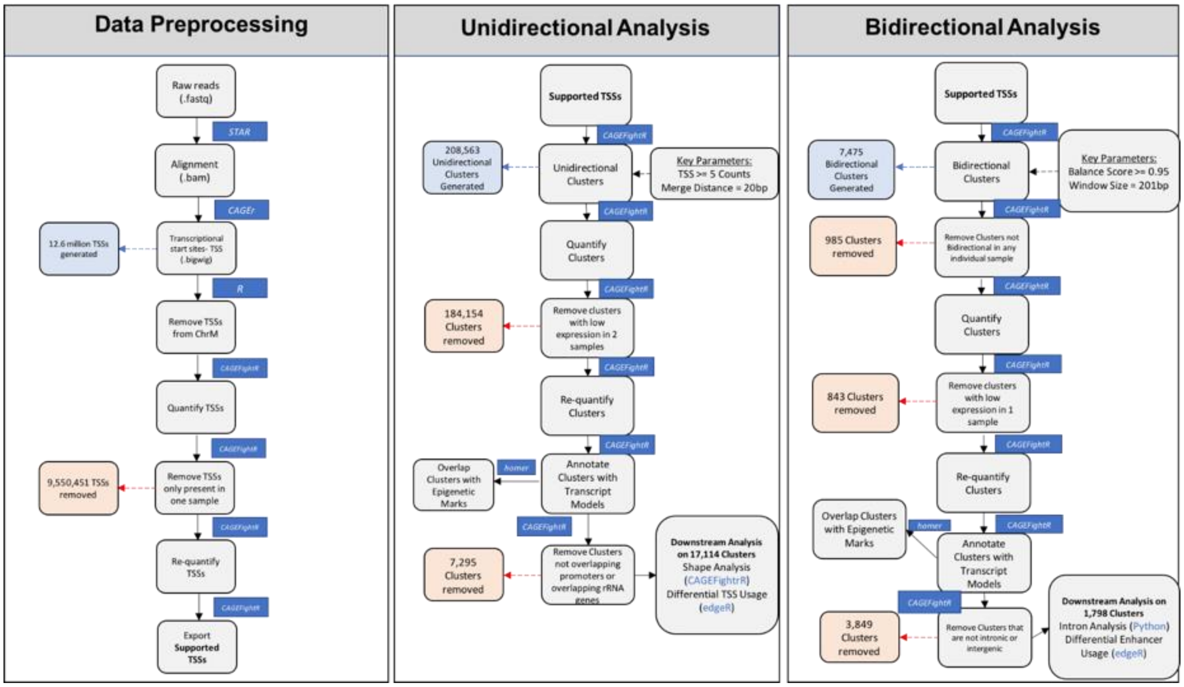

Data Preprocessing
Raw reads (.fastq)
STAR
Alignment (.bam)
CAGEr
12.6 million TSSs generated
Transcriptional start sites- TSS (.bigwig)
R
Remove TSSs from ChrM
CAGEFightR
Quantify TSSs
CAGEFightR
9,550,451 TSSs removed
Remove TSSs only present in one sample
CAGEFightR
Re-quantify TSSs
CAGEFightR
Export Supported TSSs
Unidirectional Analysis
Supported TSSs
CAGEFightR
208,563 Unidirectional Clusters Generated
Unidirectional Clusters
Key Parameters:
TSS >= 5 Counts
Merge Distance = 20bp
CAGEFightR
Quantify Clusters
CAGEFightR
184,154 Clusters removed
Remove clusters with low expression in 2 samples
CAGEFightR
Re-quantify Clusters
CAGEFightR
Overlap Clusters with Epigenetic Marks
homer
Annotate Clusters with Transcript Models
CAGEFightR
7,295 Clusters removed
Remove Clusters not overlapping promoters or overlapping rRNA genes
Downstream Analysis on 17,114 Clusters
Shape Analysis (CAGEFightrR)
Differential TSS Usage (edgeR)
Bidirectional Analysis
Supported TSSs
CAGEFightR
7,475 Bidirectional Clusters Generated
Bidirectional Clusters
Key Parameters:
Balance Score >= 0.95
Window Size = 201bp
CAGEFightR
985 Clusters removed
Remove Clusters not Bidirectional in any individual sample
CAGEFightR
Quantify Clusters
CAGEFightR
843 Clusters removed
Remove clusters with low expression in 1 sample
CAGEFightR
Re-quantify Clusters
Overlap Clusters with Epigenetic Marks
homer
CAGEFightR
Annotate Clusters with Transcript Models
CAGEFightR
3,849 Clusters removed
Remove Clusters that are not intronic or intergenic
Downstream Analysis on 1,798 Clusters
Intron Analysis (Python)
Differential Enhancer Usage (edgeR)

Figure 2.2. Data analysis pipeline for identifying and analyzing promoters and enhancers from CAGE-seq data. *TSS*, transcriptional start sites.

quantified CAGE transcriptional start site (CTSS) coordinates with removal of first G nucleotide mismatches.(119) CTSS coordinates and counts were outputted as bigwig files for input to CAGEfightR.(120) CTSSs from mitochondrial chromosomes and CTSSs only present in a single sample were removed. We clustered CTSSs from all samples into unidirectional and bidirectional clusters (**Figure 2.1**). CTSS's were required to have ≥5 pooled counts be included in a unidirectional cluster and all CTSSs within 20bp were merged into the same cluster. Unidirectional clusters also were required to have >1 TPM (tag per million) in at least 2 samples. Bidirectional clusters were required to have a balance score ≥0.95 and a 200bp window on either side of the midpoint was used to quantify each cluster, as described in (50). Bidirectional clusters were also required to be bidirectional in at least one sample and have ≥ 2 counts in at least one sample. These clusters were annotated with Ensembl GTF file version 87 annotations (downloaded May 2016), which includes known coding and noncoding RNA transcripts (**Figure 2.2**). Unidirectional clusters overlapping known rRNA genes were also removed.

CAGE Cluster Epigenetic and Transcription Factor Overlaps. Epigenetic datasets of interest were downloaded from their respective locations (**Table 2.2**). Bam files were converted into tag directories using HOMER.(42) HOMER annotatePeaks.pl was used to determine the read depth of each epigenetic dataset (normalized for cluster length and number) for each cluster annotation type (±1000bp of the cluster midpoint). HOMER findMotifsGenome.pl was used to check for enrichment of known transcription factors for each annotation type. For unidirectional clusters, the cluster midpoint ±200bp was used as input. For bidirectional clusters, the cluster boundaries were used and no additional nucleotides were added. Homer generated background sequences generated from a masked genome were used.

Target	Assay Type	Tissue Source	Accession #
Open Chromatin	ATAC-Seq	Female adult (51 years) LV Tissue	ENCSR117PYB
Open Chromatin	DNAse-Seq	Female adult (53 years) LV Tissue	ENCFF702IJE
H3K4me1	ChIP-Seq	Male child (3 years) LV Tissue	ENCFF901JPP
H3K4me3	ChIP-Seq	Male Adult (34 years) LV Tissue	ENCFF527LGE
H3K27Ac	ChIP-Seq	Female Adult (51 years) LV Tissue	ENCFF625XET
CTCF	ChIP-Seq	Female Adult (53 years) LV tissue	ENCFF738KRH
POL2A	ChIP-Seq	Female Adult (53 years) LV tissue	ENCFF318MWF

Table 2.2. Epigenetic Datasets used for CAGE-Cluster Functional Annotation. *ATAC*, assay for transposase-accessible chromatin. *ChIP*, chromatin immunoprecipitation.

Promoter Width Analysis. CAGEfightR was used to calculate the 0.1 to 0.9 inter-quantile range (IQR) for unidirectional clusters overlapping known promoter regions. A cutoff of 10bp was used to define a sharp and broad populations of promoters. The genomic coordinates of each promoter’s predominant TSS with 100bp added upstream and 50bp added downstream were inputted into bedtools getfasta to obtain genomic sequences.(121) These sequences were inputted into the WebLogo tool to generate visual representations of nucleotide enrichment at each position relative to the predominant TSS.(122) The genes of each promoter type were also inputted into the PantherGO online tool to check for enrichment of gene ontology terms. (123) The R package ggplot2 was used to generate violin plots of sharp and broad promoter pooled expression levels and basepair width. To compare promoter IQR values across failed and healthy hearts, we first filtered out any promoters that were not present in all hearts. We used CAGEfightR with sample-specific scores to calculate the IQR of each promoter in each sample. We compared the average IQR across all promoters in the three healthy samples to the average IQR in the four failed samples using the nonparametric Wilcoxon rank sum test in R.

Intronic Enhancer Analysis. A custom script was written to generate an annotation file of first intron locations for all transcripts present in the Ensembl GTF file version 87 (Downloaded May 2016). We used a custom Python script to evaluate if bidirectional intronic completely overlapped the first intron of any transcript. ggPlot2 was used to generate violin plots of enhancer eRNA expression levels, base pair width, and bidirectionality scores. Genes with first intron and other intron enhancers were inputted into the PantherGO online software tool to check for enrichment of gene ontology terms.(123)

RNA-Seq Data Analysis and Comparison with CAGE-Seq Data. Raw RNA-seq reads were trimmed with trimmomatic (v0.36) and aligned to the human genome (hg19) using STAR with default settings (118). Uniquely aligned reads were assigned to genes using htseq-count using the Ensembl GTF file version 87 (Downloaded May 2016) as annotations.(124) Raw count matrices were inputted into EdgeR for normalization, dispersion estimation, and glm-model approach measures of differential expression between healthy and failed hearts.(125) Genes were required to have at least 1 count per million in at least 3 different samples. We defined differentially expressed genes as any gene with an FDR-corrected p-value of < 0.05. The read counts of CAGE-seq unidirectional clusters overlapping gene promoters were used to quantify overall gene expression. Expression values from multiple promoters of the same gene were collapsed into a gene-level value. These count values were inputted into EdgeR and analyzed similar to the RNA-seq data above. ggPlot2 was used to graph the log-normalized and depth-normalized gene expression values generated by CAGE-seq and RNA-seq. The R package corrplot was used on the normalized count matrix to generate a correlation matrix across all samples.

Significantly downregulated and upregulated genes were separated based on the sign of their log fold-change value. Ensembl gene IDs were inputted into the PantherGO online tool to check for enrichment of gene ontology terms.(123)

Differential Enhancer Analysis. Raw count values representing eRNA expression levels for bidirectional enhancers annotated as intragenic or intronic were exported as a counts table. This counts table was imported into EdgeR for differential expression analysis.(125) Counts were normalized to library size, dispersion estimated, and differential enhancer usage called using glm-models. EdgeR was also used to generate an MDS plot of normalized enhancer counts. Due to the low count numbers associated with eRNA expression, expression values are subject to high levels of variation. EdgeR, like other RNA-seq analysis tools, detects this increased variation and reports higher p-values for calling differential expression. After multiple testing correction, there are too few enhancers surviving for downstream analysis. Therefore, raw p-value cutoffs were used. Enhancers with raw p-values ≤ 0.05 were used as input to HOMER findmotifsGenomewide.pl to find *de novo* motif enrichments in differential enhancers (enhancers with raw p-values > 0.05 were used as background sequences).(42)

Alternative Promoter Usage Analysis. Unidirectional CAGE clusters overlapping annotated promoters were used as our promoter set. We required that an individual promoter make up at least 1% of total gene counts in all samples to be included in our analysis. A python script was written to count the number of promoters per gene. A python script was also written to calculate the percent usage of each promoter. The percent usage was averaged for the 3 healthy hearts and 4 failed hearts and the difference was calculated. To assess the alternative promoters' effect on gene protein sequence, a custom annotation of all transcripts' start codons was generated using the Ensembl GTF file version 87 (Downloaded May 2016).

Enhancer Validation with Other Methods. Enhancer files from published sources were downloaded. To determine overlap with the Vista enhancer browser, enhancers with heart signal, enhancers with any positive signal, and enhancers with no signal were downloaded.(57) For Dickel et al. 2016, the "Putative human heart enhancers identified by integrative analysis" table was used as enhancer predictions.(115) For Spurrell et al. 2019, the "Predicted Enhancers in any 2 Samples" file was used as enhancer predictions. For FANTOM data, the "Ubiquitous enhancers organs" file for enhancer regions active in all organs tested was used. The FANTOM left ventricle and cerebellum enhancer sets were determined by requiring that the enhancer have non-zero expression in each tissue.(50) As a negative control, genomic coordinates of CAGE-defined enhancer regions were scrambled ~500 times, avoiding placement in repeat or gap sequences using bedtools shuffle, which keeps region sizes consistent. For each scramble, we calculated the number of overlaps with downloaded predictions using bedtools intersect and requiring at least 1bp overlap (121). Significance was determined by comparing the CAGE-defined enhancers overlap value with the normally distributed shuffled control values using R's pnorm function.

Results

Identifying genetic regulatory regions of the left ventricle. CAGE sequencing identifies promoters and putative enhancer regions. Since gene expression patterns differ between normal and failing hearts, we generated CAGE sequence datasets from left ventricle (LV) from three healthy and four failed hearts. Healthy LV samples were those acquired but not used for transplant due to age or other incompatibility. Failed hearts were obtained at the time of transplant from patients with a range of ages (**Table 2.3**). The small sample size did not allow the consideration of age and primary gene mutation in the analysis. Each

Sample	Primary phenotype	Additional phenotypes	Sex	Race	Age	Primary gene mutation(s)
Healthy 1	Healthy	-	M	Caucasian	62	N/A
Healthy 2	Healthy	-	M	Caucasian	47	N/A
Healthy 3	Healthy	-	F	Caucasian	76	N/A
Heart Failure 1	Cardiomyopathy	Ventricular Tachycardia	M	Caucasian	20	*TPM1* D230N
Heart Failure 2	Cardiomyopathy	-	M	Hispanic	16	*TTN* c.42521-5 C>G, *TNNT2* K210del
Heart Failure 3	Cardiomyopathy	Becker Muscular Dystrophy	M	Caucasian	54	*DMD* IVS +1 G>T
Heart Failure 4	Cardiomyopathy	Limb Girdle Muscular Dystrophy	F	Caucasian	26	*LMNA* c.1142-1157+1del17

Table 2.3. Left Ventricle Tissue Source Demographics, Phenotypes, and Mutations

library was sequenced to high depth with comparable alignment rates (**Table 2.1**). To generate a comprehensive list of all potential promoters and enhancers, the initial analysis combined the results from healthy and failed hearts. CAGE sequence analysis identified 23,676 promoter regions, defined as unidirectional sequence clusters, and 5,647 enhancers, defined as bidirectional sequence clusters.

Unidirectional CAGE sequence clusters (promoters) were mapped relative to annotated genes. 70.1% of clusters mapped near transcriptional start sites (±100bp), consistent with their putative role as promoters (**Figure 2.3A**). An additional 8.1% of these unidirectional clusters, mapped between 100 and 1000bp upstream of transcriptional start sites. The remaining 21.8% of sequence clusters mapped to untranslated regions, exons, noncoding RNAs, introns or intergenic regions.

We next analyzed clusters for the presence of transcription factor binding motifs. The 70.1% of clusters mapping within 100bp of transcriptional start sites were highly enriched for GFY-Staf, Sp1, and Elk/ETS binding motifs, which are transcription factors known to bind promoters.(126) Clusters mapping into other regions showed minimal enrichment of these motifs (**Figure 2.3B**). To provide additional support for the promoter-enriched sequence clusters, ATAC sequencing and H3K4me3 ChIP-seq datasets were compared since these data indicate open chromatin and promoter function; **Table 2.2** provides source information for these datasets. The promoter clusters overlapped considerably with ATAC-seq and H3K4me3 ChIP-seq signals, indicative of open chromatin and active promoter regions (**Figure 2.3C**). These clusters also showed high CTCF and Pol2A binding, as well as a reduction of H3K4me1 histone modifications, which supports their role as promoters and not enhancers (**Figure 2.4**). The bimodal shape of histone methylation patterns is consistent with open chromatin signals being flanked by promoter histone marks. Taken together, the unidirectional CAGE sequence clusters bear the genomic signatures of active promoters.

Bidirectional eRNA clusters indicate likely enhancer regions. We similarly annotated bidirectional CAGE clusters for the position relative to genes. Only 44.5% of bidirectional clusters mapped ±100bp within transcriptional start sites. In contrast to unidirectional clusters,

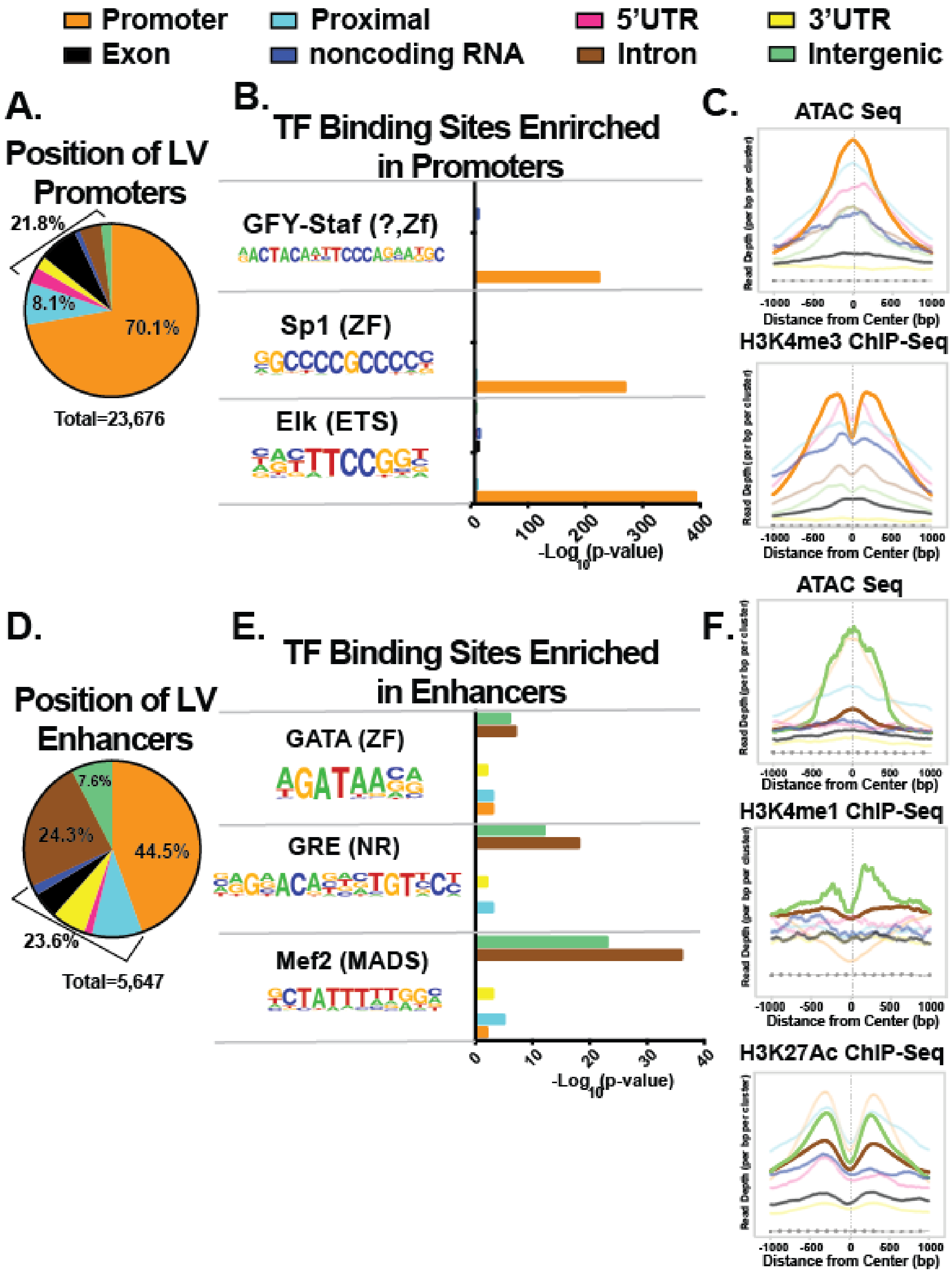


Figure 2.3. Promoters and enhancers of human left ventricles. CAGE sequencing was used to identify putative regulatory regions used in the LV. **A.** The majority of promoter regions (70.1%) mapped near ±100bp to the transcriptional start sites. **B.** Promoter regions were enriched for three transcription factor (TF) binding motifs. **C.** Left ventricle signals of open chromatin (ATAC-seq) and a promoter-associated histone mark (H3K4me3) across these predicted promoters, consistent with their role as active promoters. **D.** Only 44.5% of enhancers mapped in promoter regions and 24.3% mapped in introns. **E.** Enhancer regions were enriched for cardiac transcription factor binding sites, and these sites mapped preferentially to introns and intergenic regions. **F.** These enhancer regions also had signals of active enhancer function, seen as ATAC-seq and H3K4me1 and H3K27Ac histone marks. Dashed lines in **C** and **F** represent negative control signals from genomic regions created by scrambling the location of unidirectional and bidirectional clusters, respectively.

24.3% of clusters mapped to gene introns and 7.6% were intergenic (**Figure 2.3D**). Intergenic and intronic clusters showed enrichment of GATA, GRE, and MEF2 transcription factor binding motifs, and each of these transcription factors are essential for cardiomyocyte specification and maintenance (**Figure 2.3E**).(127) Intergenic bidirectional clusters showed enrichment of open chromatin signals (ATAC-seq), H3K4me1, and H3K27Ac histone modifications in human left ventricles. Intronic clusters also showed a similar pattern, but with lower magnitude (**Figure 2.3F**). The intergenic and intronic bidirectional clusters showed enrichment of CTCF and Pol2A binding as well as reduced H3K4me3 modifications (**Figure 2.4**). These patterns are highly consistent with the role of bidirectional CAGE sequence clusters as being enhancers, rather than promoters. Furthermore, these patterns represent multiple, independently-derived sources of evidence that bidirectional eRNA transcription signify functional enhancers.

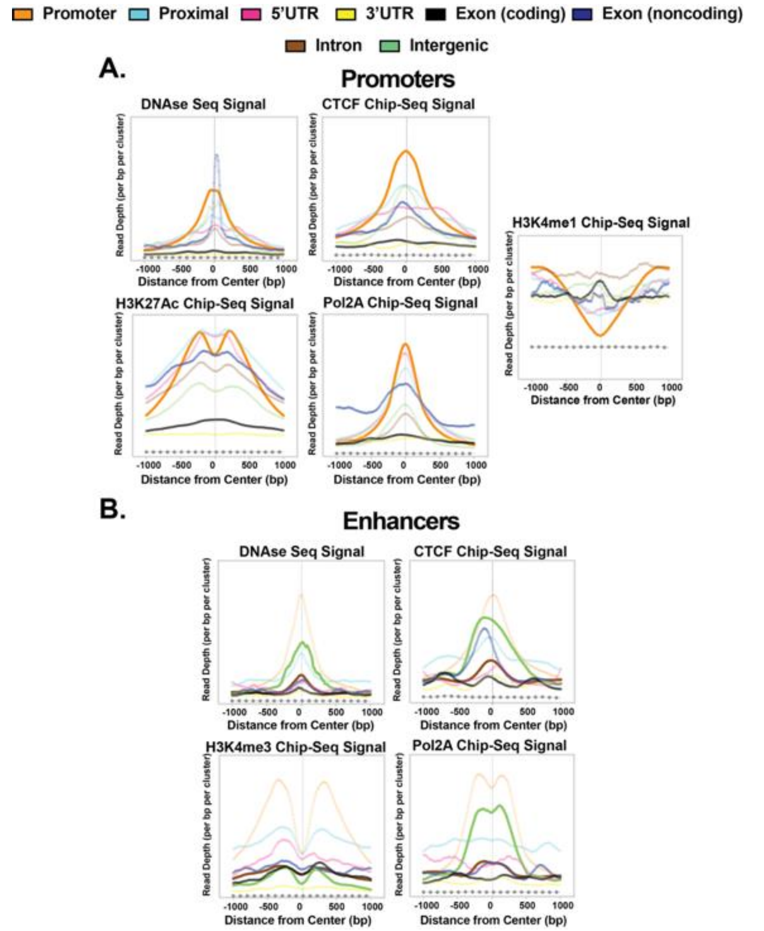


Figure 2.4. Additional left ventricle epigenetic signals of unidirectional and bidirectional CAGE clusters. **A.** Left ventricle open chromatin (DNAse-seq), protein binding (CTCF, Pol2A), and enhancer histone marks (H3K27Ac and H3K4me1) signals for unidirectional CAGE clusters of all annotation classes. **B.** Left ventricle open chromatin (DNAse-seq), protein binding (CTCF, Pol2A), and promoter histone marks (H3K4me3) signals for bidirectional CAGE clusters. Dashed lines in A and B represent signals from genomic regions created by scrambling the location of unidirectional and bidirectional clusters, respectively.

CAGE sequence-predicted promoters show shape divergence. Mammalian promoters can initiate transcription across broad or narrow genomic regions, and these promoter shapes, broad or narrow (sharp), correlate with distinct transcriptional regulatory mechanisms.(128) We evaluated cardiac promoters predicted from CAGE clusters for these two major types of promoters by calculating the interquantile range (IQR) of promoter CAGE clusters by determining the base pair distance between 10% and 90% of a promoter's total signal. We observed the expected two distinct populations, defined as sharp (IQR < 10bp) and broad promoters (IQR ≥ 10bp) (**Figure 2.5A**). Broad promoters were those associated many different cellular functions, including housekeeping genes. In contrast, genes linked to

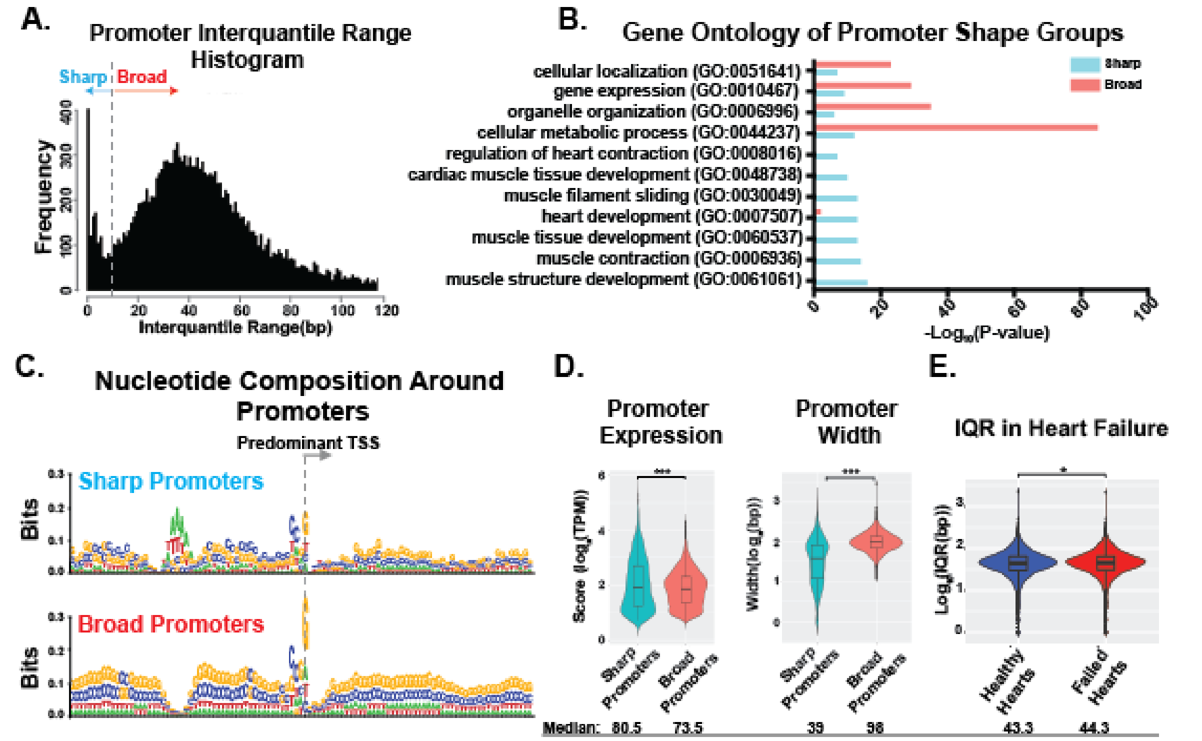


Figure 2.5. Cardiac promoters have sharp or broad transcriptional start sites. A. Histogram of interquantile range (IQR) to identify sharp (narrow) and broad promoters. IQR was defined as the number of basepairs between 10% and 90% of a total signal from a given promoter. **B.** Genes driven by broad promoters had housekeeping functions while sharp promoters were found across all gene ontology categories including tissue specific genes like muscle filament and muscle contraction genes. **C.** Nucleotide compositions of the upstream and downstream sequences from the transcriptional start site for sharp and broad promoters identified that sharp promoters used TATA regulatory motifs, while broad promoters are defined by CpG islands. **D.** Violin plots comparing the expression level and width of sharp and broad promoters showed that sharp promoters were expressed higher. **E.** Violin plot comparing the interquantile range of promoters in healthy and failed hearts. Significance determined by two-tailed nonparametric Mann Whitney Test ($p \leq 0.05$(*), ≤ 0.0005 (***)). *TSS*, transcriptional start site. *IQR*, interquantile range. *bp*, basepair.

the sharp (narrow) promoters encoded genes important for tissue specific functions seen by the presence of muscle and cardiac gene ontology terms (**Figure 2.5B**). Thus, tissue specific genes important for left ventricular specification and function were more likely to have sharp promoters.

Sharp and broad promoters also displayed differential enrichment of upstream sequence DNA-binding motifs. Sharp promoters had TATA motifs at positions 30-33 upstream of the predominant transcriptional start site, representing canonical TATA boxes (**Figure 2.5C**). Broad promoters were devoid of TATA motifs, but did show enrichment of GC nucleotides consistent with CpG-islands.(129) Sharp, tissue-specific promoters were also more highly expressed compared to broad promoters, and this observation was driven by a smaller population of very highly expressed sharp promoters, for example those encoding sarcomere genes (**Figure 2.5D**). We compared promoter shape between healthy and failed hearts and found a small but significant genomewide increase in promoter IQR in failed hearts, consistent with a slight widening of transcriptional start sites in heart failure (**Figure 2.5E**).

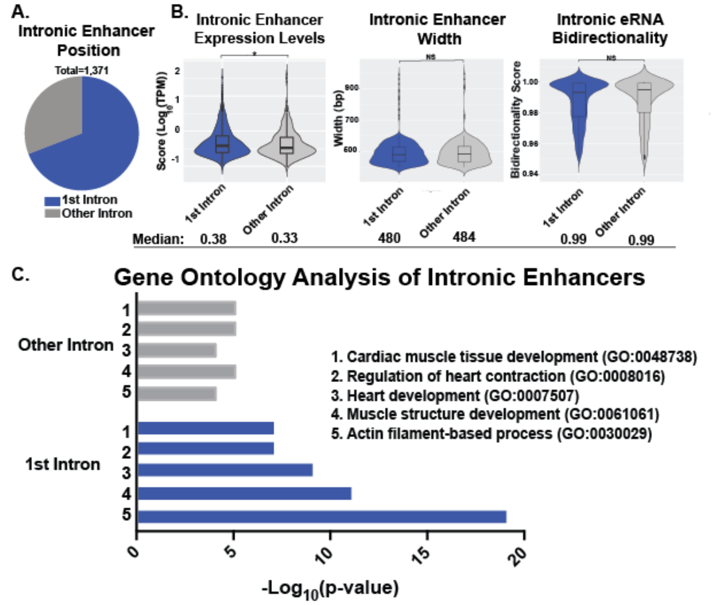
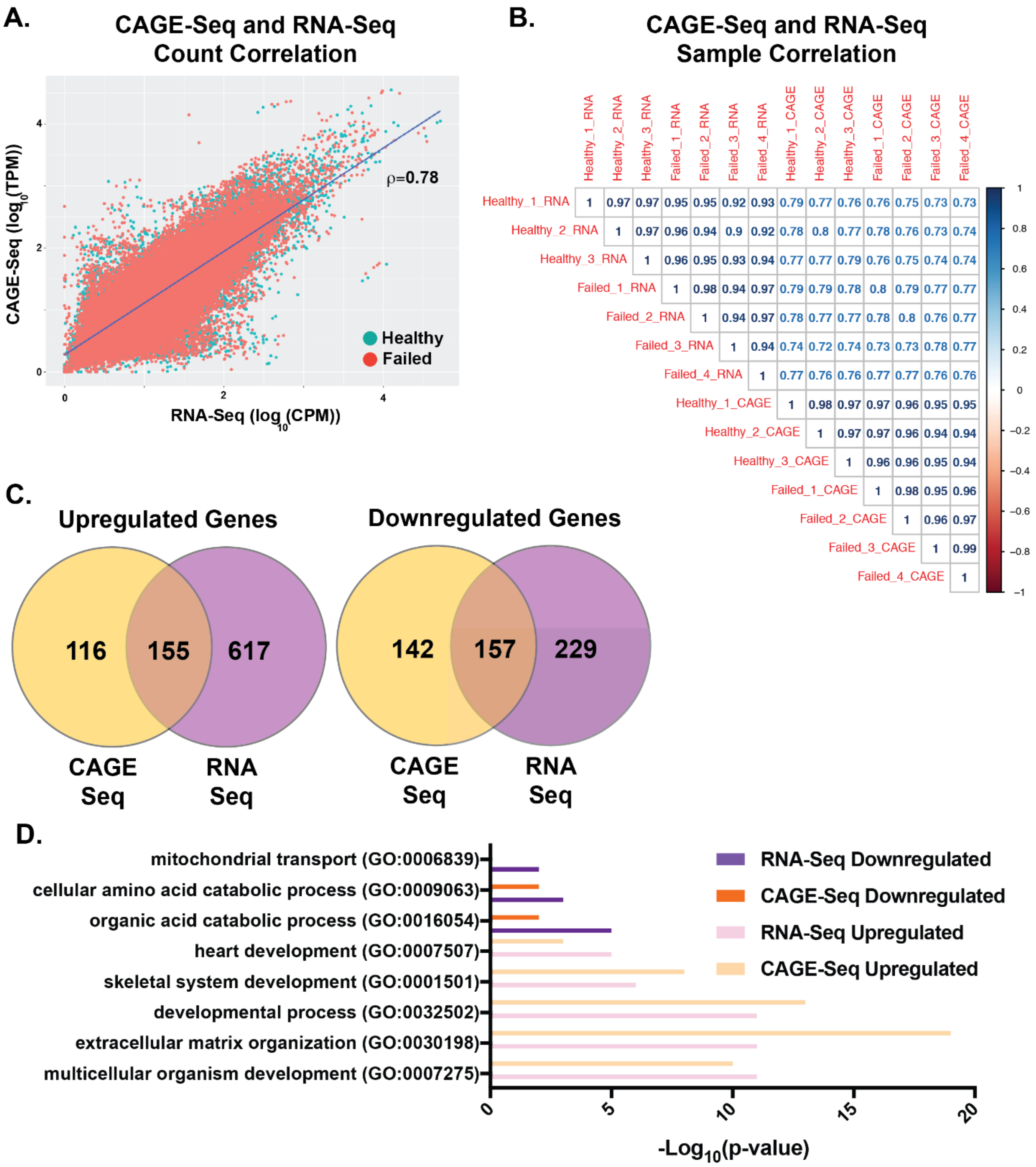


Figure 2.6. The majority of intronic cardiac enhancers localize to the first intron. A. Approximately 70% of intronic enhancers were within the first intron of an overlapping transcript. **B.** Violin plots comparing enhancer expression levels, enhancer width, and enhancer bidirectionality score between enhancers in the first intron and enhancers in other introns indicating that first intronic enhancers were similar to other enhancers except that there were expressed more highly. **C.** Gene ontology analysis of genes with first intron enhancers and genes with other intron enhancers indicates tissue specific genes are more likely to have first intron enhancers. Significance determined by two-tailed nonparametric Mann Whitney Test ($p \leq 0.05$(*)).

Predicted enhancers map within the first Intron. A large proportion of the predicted enhancers mapped to introns. These intronic clusters shared transcription factors and epigenetic marks with intergenic CAGE clusters, consistent with their roles as enhancers (**Figure 2.3E**). We observed that the majority (69%) of intronic enhancers in this dataset mapped to the first intron (**Figure 2.6A**). First introns are more conserved than other introns and correlate with higher levels of gene expression.(130) In the LV CAGE sequence data, these first intron enhancers generated more eRNA than enhancers in other introns, but were not wider and did not differ in their balance of bidirectionality (**Figure 2.6B**). Notably, these intronic enhancers mapped within genes enriched for cardiac and muscle gene ontology terms (**Figure 2.6C**).

Correlation of CAGE-sequencing and RNA-sequencing. RNA sequencing was carried out on the same LVs and compared to CAGE sequencing. Since CAGE sequencing quantifies promoter expression, it reflects overall gene expression. Consistent with this, there was a tight correlation between CAGE sequencing and RNA sequencing results (**Figure 2.7A**). Additionally, we assessed correlations between pairs of samples. In general, healthy hearts correlated best with other healthy hearts, and failed hearts compared best with failed hearts. The RNA sequence and CAGE sequence expression estimates were most correlated for matched samples except for Failed Heart 4, which likely reflects the lower CAGE sequence read depth in this sample (**Figure 2.7B**). We next compared gene expression differences between failed and nonfailed hearts using both CAGE and RNA sequence datasets. RNA-seq identified more upregulated and downregulated genes, and approximately half of the genes identified by CAGE-seq were also identified by RNA-seq (**Figure 2.7C**). Gene ontology analyses on differentially expressed genes were similar in both sequence datasets. Genes associated with developmental pathways and extracellular matrix organization were upregulated in heart failure while genes associated with catabolism were downregulated in heart failure (**Figure 2.7D**), consistent with prior gene expression profiling of failed hearts.(103, 104)



	Healthy_1_RNA	Healthy_2_RNA	Healthy_3_RNA	Failed_1_RNA	Failed_2_RNA	Failed_3_RNA	Failed_4_RNA	Healthy_1_CAGE	Healthy_2_CAGE	Healthy_3_CAGE	Failed_1_CAGE	Failed_2_CAGE	Failed_3_CAGE	Failed_4_CAGE
Healthy_1_RNA	1	0.97	0.97	0.95	0.95	0.92	0.93	0.79	0.77	0.76	0.76	0.75	0.73	0.73
Healthy_2_RNA		1	0.97	0.96	0.94	0.9	0.92	0.78	0.8	0.77	0.78	0.76	0.73	0.74
Healthy_3_RNA			1	0.96	0.95	0.93	0.94	0.77	0.77	0.79	0.76	0.75	0.74	0.74
Failed_1_RNA				1	0.98	0.94	0.97	0.79	0.79	0.78	0.8	0.79	0.77	0.77
Failed_2_RNA					1	0.94	0.97	0.78	0.77	0.77	0.78	0.8	0.76	0.77
Failed_3_RNA						1	0.94	0.74	0.72	0.74	0.73	0.73	0.78	0.77
Failed_4_RNA							1	0.77	0.76	0.76	0.77	0.77	0.76	0.76
Healthy_1_CAGE								1	0.98	0.97	0.97	0.96	0.95	0.95
Healthy_2_CAGE									1	0.97	0.97	0.96	0.94	0.94
Healthy_3_CAGE										1	0.96	0.96	0.95	0.94
Failed_1_CAGE											1	0.98	0.95	0.96
Failed_2_CAGE												1	0.96	0.97
Failed_3_CAGE													1	0.99
Failed_4_CAGE														1

Figure 2.7. Comparison of CAGE-seq and RNA-seq gene expression levels. A. Scatter plot of CAGE-seq gene expression values versus RNA-seq expression values for healthy (teal) and failed (red) samples demonstrating tight correlation. **B.** Sample level correlation matrix of Spearman's correlation coefficient of genomewide gene expression levels. **C**. Venn diagrams displaying the number of differentially upregulated and downregulated genes determined by CAGE-seq and RNA-seq. **D.** Gene ontology analysis of genes identified as differentially upregulated or downregulated by CAGE-seq and RNA seq.

CAGE sequencing-defined enhancer regions are validated by other enhancer datasets. Of the ~1,800 candidate enhancer regions identified by CAGE sequencing, data was available from 45 of these in the Vista Enhancer Browser, a list of enhancers tested in an *in vivo* reporter assay using transgenic mouse embryos.(57) Of the 45 present in Vista, 31 (70%) demonstrated enhancer activity in the developing mouse heart (**Figure 2.8A**). We also compared CAGE sequence-predicted enhancers to those predicted by H3K27Ac and p300 ChIP sequencing from developing and adult human and mouse tissues.(115) CAGE-sequence-defined enhancers showed significantly higher overlap to H3K27Ac/p300 ChIP regions compared to length-matched scrambled control regions (**Figure 2.8B**). One study with H3K27Ac ChiP-seq data from healthy and failed human hearts was similarly compared and showed significant overlap (**Figure 2.8C**).(15) Finally, we compared CAGE sequence-defined enhancer predictions from the FANTOM consortium, which used CAGE sequencing across many non-diseased tissues to define enhancers.(50) We observed significant overlap with FANTOM predictions (**Figure 2.8D**), but we found many additional enhancers beyond the FANTOM predictions because we used a higher depth of sequencing (**Table 2.1**). The intersection of these

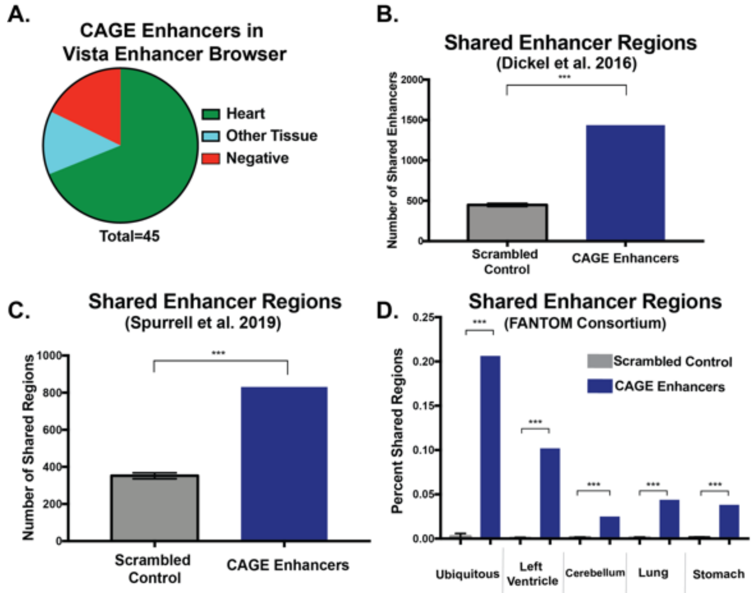


Figure 2.8. CAGE-defined enhancers overlapped enhancers determined by independent methods. **A.** Pie chart of Vista Enhancer Browser data displaying the results of functional testing of 45 CAGE enhancers. **B & C**. Bar charts showing the number of overlapping enhancers when comparing CAGE and two independent methods. **D**. Bar charts indicating the percentage of FANTOM enhancers that overlap CAGE enhancers for five different groups of FANTOM enhancers. Scrambled controls represent the number of overlaps obtained when randomly shuffling the genomic location of CAGE enhancers. Significance determined by pnorm function in R ($p \leq 0.0005$ (***)).

orthogonal datasets corroborates the cardiac enhancers now identified by CAGE sequence in both healthy hearts and failed hearts.

Alternative promoter usage in heart failure. In the LV, 3,032 (23%) expressed genes had evidence for more than one promoter (**Figure 2.9A**). For these multi-promoter genes, we compared the average percent-usage of each promoter in healthy and failed hearts and found 609 promoters in 325 genes with a shift ≥10% (**Figure 2.9B**). Of these, 149 promoters in 124 genes occurred after the exon containing the start codon, indicating the potential to alter the amino-terminal amino acid sequence of the resulting protein (**Figure 2.9C**). Of the 124 genes identified as having alternative promoters that occur after start codons in heart failure, many are associated with sarcomere regulation or muscle structure development, including *TNNT*, *MYOT*, and *SPEG.* This indicates the heart failure can result in alternative proteins due to promoter shifts. We annotated a significant promoter switch in *PRKAG2*, a gene linked to hypertrophic cardiomyopathy and critical to heart metabolism.(131) Three major *PRKAG2* promoters were identified, encoding three different isoforms- γ2a, γ2-3b, and γ2b **(Figure 2.9D)**. In healthy hearts, the relative expression of these three transcripts is 53% γ2b, 28% γ2-3b, and 17% γ2a. In heart failure, these percentages significantly shift with 29% γ2b, 59% γ2-3b, and 10% γ2a isoform (**Figure 2.9E**). Notably, the γ2-3b isoform encodes a unique 32 amino acid sequence at the amino-terminus (**Figure 2.9F**). Total expression of *PRKAG2* was not different between healthy and failed hearts. We interrogated the 30kb upstream of the γ2b and γ2-3b isoforms for transcription binding motifs and found an enrichment of Smad and GRE motifs upstream of the γ2-3b isoform, suggesting a role for these transcription factors (**Table 2.4**).

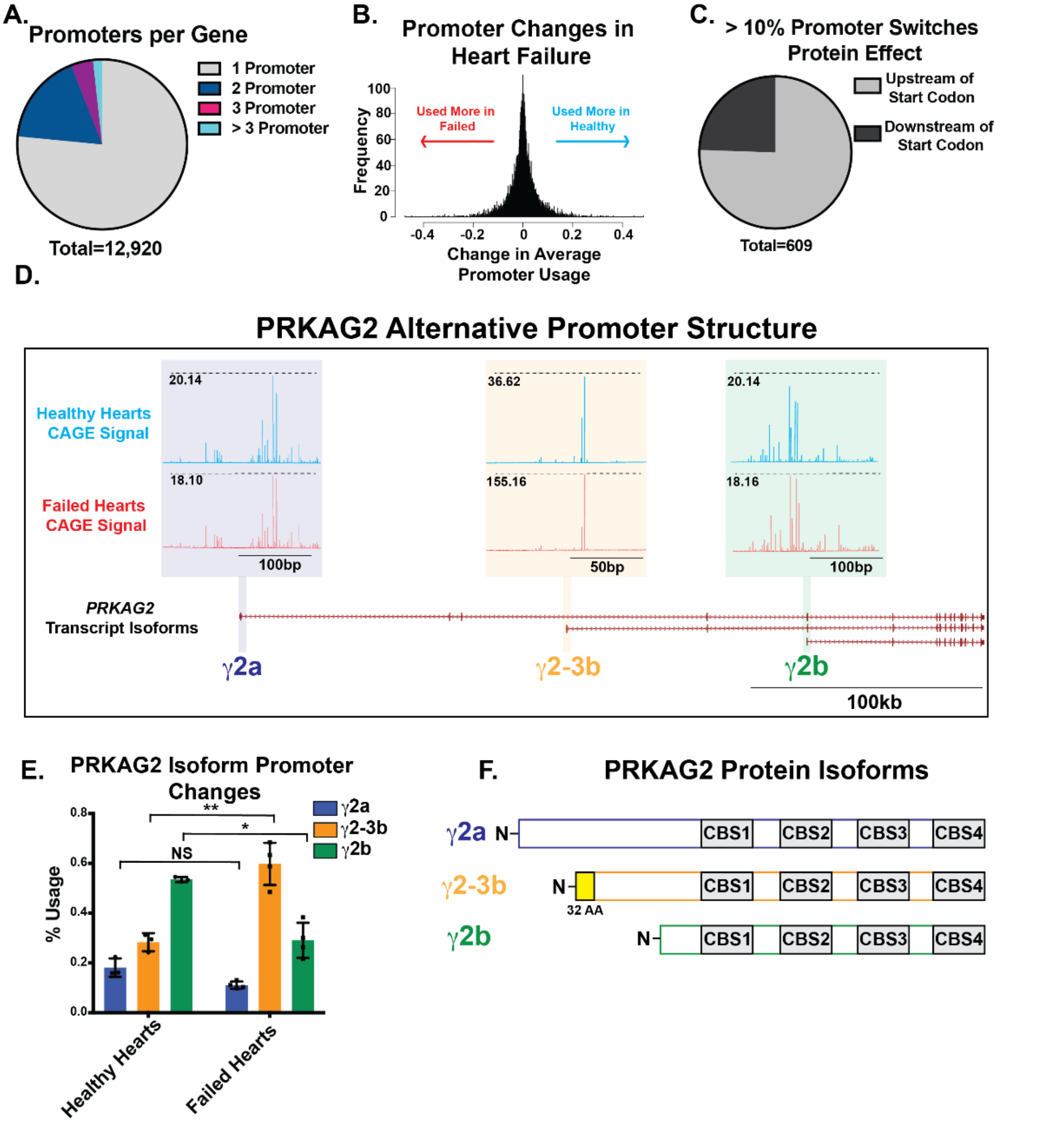


Figure 2.9. Heart failure was associated with significant changes in LV promoter usage. A. The fraction of genes using more than one promoter is indicated. **B.** Histogram displaying the distribution of average promoter percent usage changes in heart failure. The x-axis represents the difference between a promoter's average percent usage in three healthy left ventricles and four failed left ventricles. Left-shifted promoters make up a greater percentage of gene expression in failed ventricles and right-shifted promoters are a greater percentage in healthy ventricles. **C.** Pie chart of the relationship to an overlapping transcript's start codon for promoters that undergo a ≥ 10% shift in heart failure. **D**. Genome browser representation of the alternative promoter structure of the *PRKAG2* gene. Three known isoforms of *PRKAG2* are represented at the bottom. Above the promoter of each transcript, the CAGE-seq signal for healthy (blue) and failed (red) hearts is shown and the scales of each representation are indicated in black. **E.** Quantification of the CAGE-seq signals shown in D indicating the promoter percent usage of each isoform in healthy and failed hearts. Significance determined by a two-tailed Student's t-test. **F.** Schematic of the predicted amino acid sequences translated from each *PRKAG2* isoform. ($p \leq 0.05$(*), ≤ 0.005(**)) *PRKAG2,* Protein Kinase AMP-Activated Non-Catalytic Subunit Gamma.

Motif	Upstream γ2b Motif Counts	Upstream γ2-3b Motif Counts	γ2-3b Enrichment
GRE(NR)	1	8	8.00
KLF1(Zf)	4	22	5.50
Smad2(MAD)	11	31	2.81
Smad3(MAD)	40	54	1.35

Table 2.4. Selected transcription factor motifs upstream of *PRKAG2* isoforms

Enhancer usage shifts in heart failure. The CAGE sequence analysis identified ~1,800 enhancer regions actively transcribed in human LV (**Figure 2.3A**). Multidimensional scaling of normalized expression levels showed an overall similar profile of enhancer usage across the healthy LVs, but disparate enhancer usage across the four failed LVs (**Figure 2.10A**). Comparing enhancer usage across heathy and failed LV revealed 264 enhancers that changed significantly in heart failure (raw p-value ≤ 0.05). To assess whether differential enhancer transcription was associated with differential transcription factor binding site profiles, we compared transcription factor motif instances across enhancers in healthy and failed LVs. We found SMAD2, NFIX, NFAT, TCF7L2, ZNF740, and AR motifs enriched in enhancers that changed in heart failure. SMAD2, NFIX, TCF7L2, and AR motifs were found more in downregulated enhancers. While NFAT and ZNF740 motifs were found more in upregulated enhancers. RNA-sequencing demonstrated that *SMAD2* and *NFAT5* were significantly upregulated in heart failure (**Figure 2.10C**).

Figure 2.10D illustrates alternative enhancer use within the first intron of *TRPM7*, which encodes the transient receptor potential cation channel subfamily M member, a gene implicated in ischemic cardiomyopathy and cardiac rhythm.(132, 133) This intronic enhancer showed

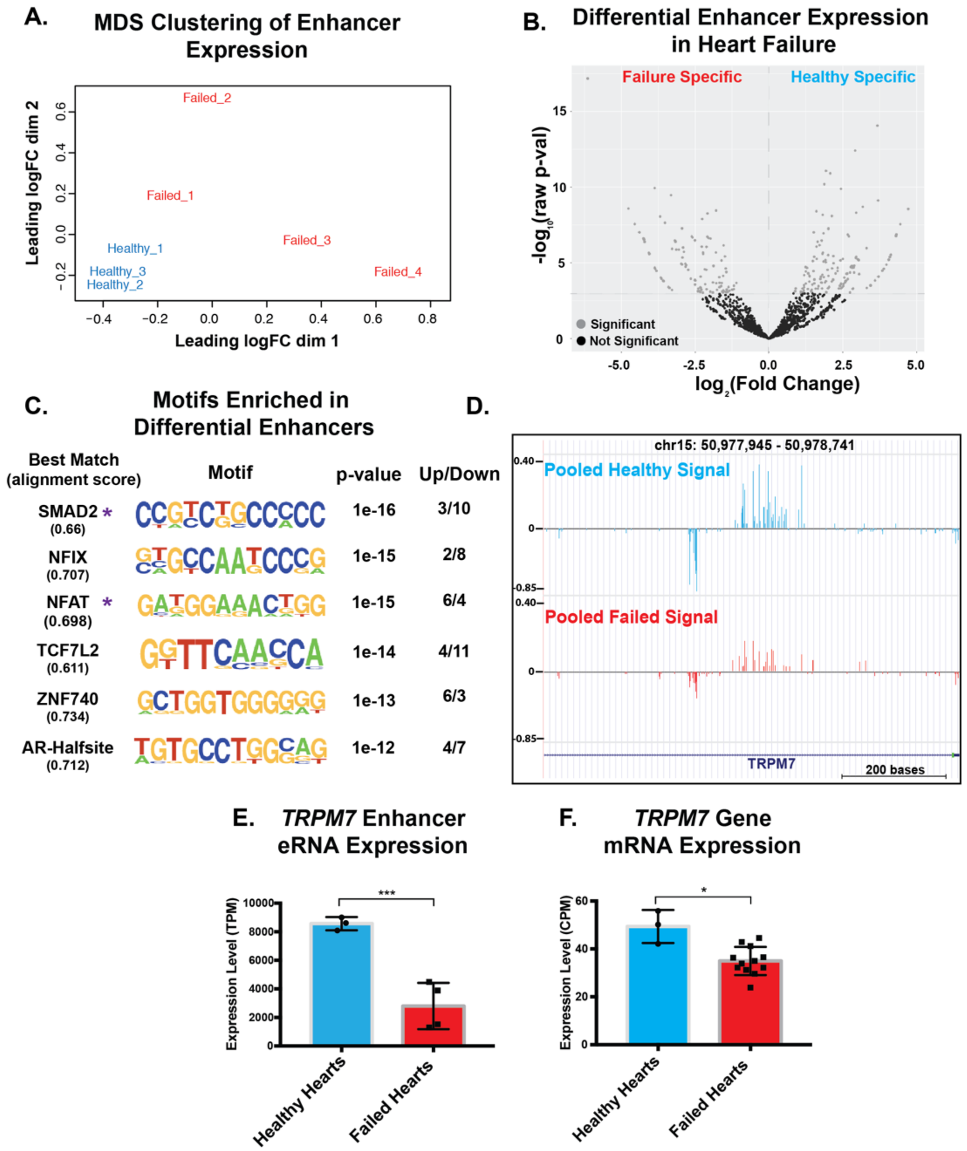


Figure 2.10. Differential enhancer usage in heart failure. **A.** Multidimensional scaling plot of enhancer expression levels showing tighter clustering of healthy LVs than failed LVs. **B.** Differentially used enhancers are shown in light gray. Left shifted enhancers are expressed higher in failed hearts and right shifted enhancers are expressed higher in healthy hearts. **C.** *De novo* transcription factor motif enrichment analysis comparing differentially changed enhancers to unchanged enhancers. The best match of enriched motifs is listed to the left. A purple asterisk indicates that the matching transcription factor was differentially expressed by RNA-seq. Up/down indicates the instances of the identified motif in upregulated and downregulated enhancers, respectively. **D**. Genome browser representation of a differentially expressed enhancer within the first intron of the *TRPM7* gene. The gene annotation is at the bottom and the healthy and failed CAGE-seq signals are graphed above on the same scale. **E**. Quantification of the healthy and failed CAGE-seq signals for the intronic *TRPM7* enhancer in D. **F**. Quantification of *TRPM7* overall gene expression by RNA-seq. Additional failed hearts were added to increase power. Significance determined by EdgeR using a generalized linear model approach. (p ≤ 0.05(*), ≤ 0.0005 (***)). *FC*, fold change. *TRPM7*, transient receptor potential cation channel subfamily M member 7.

significantly lower eRNA expression in heart failure (**Figure 2.10E**), concomitant with a significantly lower expression of *TRPM7* in failed hearts (**Figure 2.10F**). We also identified an enhancer cluster upstream of the *NPPA/NPPB* gene loci (**Figure 2.11**). *NPPA/NPPB* encode natriuretic peptides, which serve as important clinical biomarkers of volume overload in heart failure. One enhancer of this cluster was detected in all samples and showed higher expression of eRNA in heart failure (**Figure 2.11C**), consistent with the regulatory region responsible for the upregulation of natriuretic proteins in heart failure.

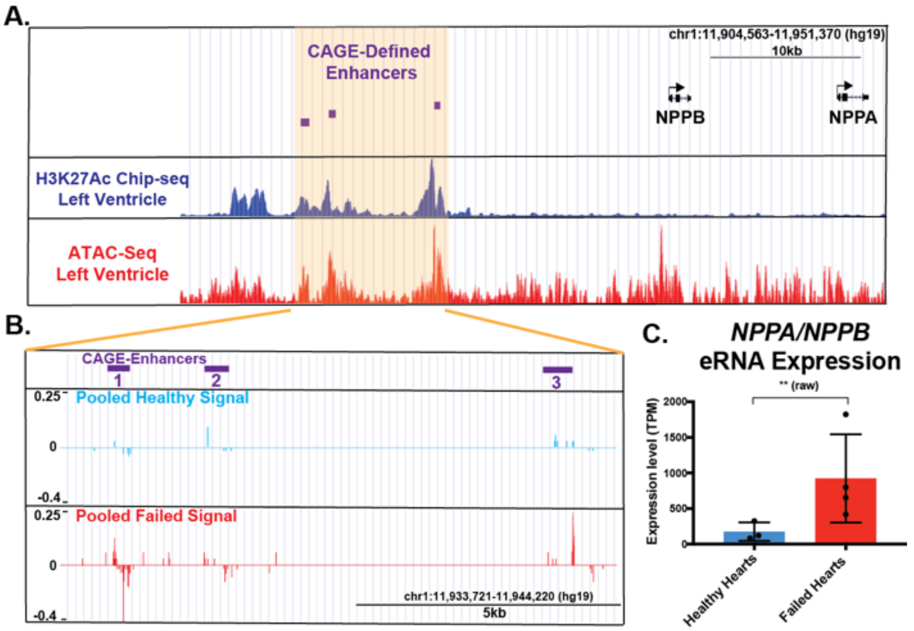


Figure 2.11. An enhancer upstream of *NPPA/NPPB* carries the signature of activation in heart failure. A. Genome browser representation of the *NPPA/NPPB* locus, which encodes ANP and BNP and is known to be upregulated in heart failure, shows three upstream CAGE-defined enhancers (purple) overlapping H3K27Ac ChIP-seq (blue) and ATAC-seq (red) signals. **B.** Depiction of the CAGE-defined enhancers in **A** with pooled healthy hearts (blue) and pooled failed hearts (red) CAGE-seq transcriptional start sites on the same scale. **C.** eRNA expression quantification from enhancer 1 in B for healthy and failed hearts. ($p \leq 0.005$(**))

Discussion

Defining the promoterome of the human LV in health and disease. Heart failure is known to be accompanied by shifts in gene expression, including a re-expression of developmental genes. This analysis provides a detailed depiction of genomewide promoter usage in left ventricle, and further highlights how promoter usage shifts in the failed heart. In total, we report ~17,000 high likelihood promoters active in the adult human heart. We observed two major promoter types, the sharp TATA-box-associated and the broad CpG island-associated.(129) Sharp promoters had single or a few transcriptional start sites and were linked to highly-expressed, tissue-specific genes like *MYH7*, *TTN*, and *MYL2*. Broad promoters had a wider distribution of transcriptional start sites and included both housekeeping and some tissue specific genes. We observed an increase in genomewide promoter width in failed left ventricles, suggesting a loss of tight regulation of transcriptional start sites. This widening of promoters may reflect epigenetic modifications or transcriptional factor profile differences.

Alternative promoter usage in heart failure. The data indicate that ~20% of genes active in the human left ventricle have multiple active promoters, correlating well with previous estimates from different cell types.(134, 135) Promoter switches that alter the noncoding regions can affect translational efficiency, imparting developmental and tissue specificity.(136) Some promoter switches can directly shift the amino-terminus of the resulting protein. We provide as an example a failure-linked promoter shift in the *PRKAG2* gene. Mutations in *PRKAG2* cause hypertrophic cardiomyopathy and arrhythmias.(131) In healthy left ventricles, we found that ~55% of transcripts originated from the γ2b promoter and ~35% originate from the γ2-3b promoter. In failed left ventricles ~60% of the *PRKAG2* transcripts represent the γ2b-3b isoform. The γ2b-3b isoform includes a unique 32 amino-acids that may affect the ability of the *PRKAG2/AMPK* complex to interact with troponin I and regulate contraction dynamics.(137) In the UK Biobank, a polymorphism, rs10224210, in the first intron of the y2-3b isoform links to cardiovascular disease. This specific sequence could alter γ2b-3b isoform expression through a first intron enhancer or, alternatively, may be in linkage disequilibrium with other coding or promoter sequences. This signal provides additional evidence that the γ2b-3b isoform may be an important mediator of heart failure. Upregulation of the γ2b-3b isoform in failed hearts may also influence how mutations in *PRKAG2* are expressed. GTEx expression data indicates that the γ2b isoform is expressed in healthy left ventricle and in cultured fibroblasts, indicating the effect of isoform shifts could alter multiple cell types in the heart.

Differential enhancer usage in heart failure. Heart failure is associated with transcriptional changes.(103) We found that predicted enhancer usage was more variable in failed ventricles, which may indicate genomewide dysregulation of gene expression. *SMAD2* binding motifs were enriched in differentially used heart failure enhancers, and this is highly consistent with the known upregulation of TGF-β signaling in failing hearts.(138, 139) The enrichment of this motif in differential enhancers may reflect increased TGF-β/SMAD activation in cardiomyocytes and/or a larger proportion of cardiac fibroblasts in the failed left ventricle tissues. *SMAD* motifs were found more often in downregulated enhancers, implying a repressive role for TGF-β/SMAD signaling in heart failure. We highlighted a specific differential enhancer located within the first intron of the *TRPM7* gene*. TRPM7* encodes kinase domain-containing cation channel. Deletion of *Trpm7* in mice disrupts cardiac automaticity and causes cardiac hypertrophy and fibrosis.(133) In ischemic cardiomyopathy, *TRPM7* was significantly down regulated in the left atria and ventricle.(132) Together, these findings support a reduction in *TRPM7* in the setting of end stage heart failure. Sequences upstream of the *NPPA*/*NPPB* gene loci were also identified as differentially activated in heart failure. The orthologous region in the mouse genome has been shown to regulate expression of these genes, validating its function *in vivo*.(98) As natriuretic peptide elevation serves as a biomarker for heart failure, this enhancer region may be an attractive target to modulate natriuretic factor expression in heart failure.

Study Limitations and Conclusions. This study used CAGE sequencing to define a broad spectrum of predicted cardiac promoters and enhancers, with focus on their differential use in heart failure. Due to the small cohort size, age, sex, and race could not be considered in the analysis. The majority of samples were of European descent, and controlling for age was not possible as age correlates with heart failure status. Therefore, this analysis was uncorrected for these covariates, and thus may limit the broader applicability of this data. The use of bidirectional eRNA transcription as a genomewide mark of enhancer function is relatively new and the exact role of eRNAs is unknown. To address this, we relied on multiple independent sources of data to support enhancer predictions, but even so, these approaches may over or underestimated the true number active enhancer regions. We observed variability in differential promoter and enhancer usage in failed heart, as the normal control hearts showed tighter correlations. This variability may reflect the end stage process of heart failure. While a larger dataset may be more revealing, the diversity of response in the failed hearts mirrors what has been observed when RNA sequencing was used to define transcripts produced for *TTN*, a large gene that has been examined in multiple failed hearts.(140, 141) The wide array of transcripts produced from even this single gene may underscore that a lack of uniform response itself could contribute to heart failure.

Chapter 3.

Integrated epigenomic analysis identifies enhancer modifying variants linked to cardiomyopathy genes.

Abstract

Inherited cardiomyopathy associates with a range of phenotypic expression. We superimposed epigenomic profiling from multiple sources, including promoter-capture chromatin conformation information, and identified candidate enhancer regions for two cardiomyopathy genes, *MYH7* and *LMNA*. Enhancer function was validated in human cardiomyocytes derived from induced pluripotent stem cells and revealed enhancer regions implicated the switch of *MYH6* and *MYH7* expression. By querying human genomic variation, we identified multiple sequence changes that modified enhancer function by creating or interrupting transcription factor binding sites. rs875908, which is 2KB 5' of *MYH7,* associated with longitudinal clinical features of cardiomyopathy in a biobank with clinical imaging and genetic data. This integrated approach identified noncoding modifiers of cardiomyopathy and is broadly applicable to other cardiomyopathy genes.

This work is under review at Nature Genetics:

> **Gacita AM**, Fullenkamp D, Ohiri J, Pottinger T, Puckelwartz MJ, Nobrega MA, McNally EM. Integrative epigenomic analysis identifies enhancer modifying variants linked to cardiomyopathy genes. *Nature Genetics* (*in review)*.

Respective Contributions:

AMG conducted experiments, analyzed data, and drafted the manuscript. DF and JO assisted in conducting experiments. TP provided access to analyzed phenotypic data. MRP provided helpful advice and commentary and assisted with interpretation. MAN and EMM assisted with analysis, writing and editing the manuscript.

Introduction

Mutations in more than 100 genes have been linked to autosomal dominant cardiomyopathy, which leads to heart failure and significant burden (64-66). A well-recognized clinical feature of genetic cardiomyopathy is varying phenotypic expression. Genetic cardiomyopathy demonstrates an age-dependent penetrance, variable expressivity, and variable clinical presentations. Even with identical primary mutations, there is a range of clinical outcomes (1, 85). Genetic variants in protein coding regions have been described as altering the phenotypic expression of a primary cardiomyopathy-causing mutation (85, 142, 143). However, the contribution of noncoding variation has been less well investigated.

Noncoding regions of the genome harbor important regulatory sequences that control the expression of genes through both distal enhancers and proximal gene promoters (144). ChIP-seq, ATAC-seq, and CAGE-seq can mark genomic regions as having regulatory function, but do not provide information on their gene target. Chromatin conformation assays evaluate genomic three-dimensional organization and link enhancers to their target genes. However, as enhancer function is dependent on tissue-specific transcription factors, assays for enhancer function or targets require the context of relevant tissues/cells.

To define the contribution of noncoding variation, we evaluated the regulatory regions for two commonly mutated cardiopathy genes, *MYH7* and *LMNA*. Mutations in *MYH7* are a common cause of hypertrophic cardiomyopathy while mutations in *LMNA* are a common cause of dilated cardiomyopathy with arrhythmias (1, 145). *MYH7* sits in tandem with *MYH6* on human chromosome 14. *MYH7* encodes β-myosin heavy chain (MHC), which is the major left ventricular myosin heavy chain in the adult human. *MYH6* encodes α-MHC and is the major myosin heavy chain in the developing ventricle and adult atrium. In mice this relationship is not conserved; adult murine ventricular myocardium is dominated by α-MHC, further underscoring the importance of studying *MYH7* regulatory regions in human systems.

We used an integrative analysis that relied on >20 publicly-available heart enhancer function and enhancer target datasets to identify *MYH7* and *LMNA* left ventricle enhancer regions. We confirmed the activity of these regions using reporter assays and CRISPr-mediated deletion in human cardiomyocytes derived from induced pluripotent stem cells (iCMs). These regulatory regions contained sequence variants within transcription factor binding sites that altered enhancer function. Extending this strategy genomewide, we identified an enhancer modifying variant upstream of *MYH7*. This common variant correlated with *MYH7* expression in the GTEx eQTL dataset. Finally, we identified this variant also correlated with a more dilated left ventricle over time. These findings link noncoding enhancer variation to cardiomyopathy phenotypes and provide direct evidence of the importance of genetic background.

Methods

Epigenetic Dataset Downloads and Visualization. Epigenetic datasets were identified from the Encode data repository or GEO. For histone ChIP-Seq datasets and ATAC-seq datasets, the “fold change over negative control” bigwig file was downloaded. For transcription factor Chip-seq datasets, peak bed files were downloaded. For Homer computational predictions, a bed file representing the location of the transcription factor motif genome-wide was downloaded. Files were imported into the UCSC genome browser for visualization. When necessary, datasets from mouse cells/tissues or hg38 were overlaid to hg19 using the UCSC liftover tool. For pcHiC data, the CHiCAGO pipeline raw output of three replicates of iCM promoter capture Hi-C data were downloaded (37). Probe-probe interactions were filtered. 1kb was added to both ends of regions interacting with gene promoters. We intersected data from each replicate using bedtools and retained only genomic interactions that were present in at least two replicates (121). Bed files representing pcHi-C interactions were visualized in the UCSC genome browser.

A UCSC genome browser session containing all tracks used for left ventricle enhancer identification is available by searching for “Gacita_et_al_LV_Enhancer_Tracks” in UCSC’s public sessions repository.

Enhancer Region Cloning. Candidate enhancer regions were ligated into luciferase plasmids using a Gateway cloning strategy. Candidate enhancer regions were amplified from human genomic DNA using primers with a 5'-CACC overhang using Phusion High-Fidelity DNA polymerase (NEB). An aliquot of the PCR reaction was separated on a 1% agarose-TBE gel to confirm amplification, and the remaining reaction was purified using a PCR Purification Kit (Qiagen). In cases where PCR failed to generate an adequate product, the enhancer region sequence (matching hg19) was synthesized as a dsDNA gGlock gene fragment (IDT). Approximately 5ng of PCR product or gBlock was ligated into the pENTR/D-TOPO vector following manufacturer's instructions (ThermoFisher). The enhancer region was recombined into pGL4.23-GW (Addgene #60323) using LR Clonase II Enzyme mix (Thermo) with 150ng of each plasmid. EndoFree Maxipreps (Qiagen) were used to prepare DNA. Plasmids were confirmed using Sanger Sequencing.

Luciferase Reporter Assay. HL-1 cardiomyocytes (Millipore Sigma Cat#SCC065) were cultured on fibronectin coated flasks in Claycomb media with 10% HL-1 qualified FBS as previously described. (14) Twenty-four hours before transfection, 140,000 HL-1 cells per well were plated on to a 12-well plate. On the day of transfection, HL-1 cells were transfected using Lipofecamine 3000 (Thermo Fisher) following manufacturer's instructions. Each well was transfected with 6μl of 0.15μM enhancer firefly luciferase plasmid, 50ng of pRL-SV40 (Promega), 2.5μl of Lipofecamine3000, and 6μl of P3000 in 100μl of Opti-MEM. Cells were allowed to incubate for 6-8 hours, following which half the media was replaced with Claycomb media. Forty-eight hours after transfection, the luciferase assay was performed with the Dual-Glo luciferase assay kit (Promega) according to manufacturer's instructions. The firefly luciferase signal from each well was recorded from three separate replicates and internally normalized to Renilla luciferase signal. Each enhancer construct was tested in a minimum of two separate wells on three separate days.

Induced pluripotent stem cell (iPSC)-derived cardiomyocytes (iCMs) were generated according to standard protocols (18). At approximately day 10 of differentiation, cardiomyocytes were re-plated on to white clear-bottom 96-well plates at 40,000 cells per well. The media was changed every two days and cells began to beat as a syncytium day 14-16. On day 18, cardiomyocytes were transfected with

Lipofecamine3000 (Thermo Fisher) according to manufacturer's instructions. Each well was transfected with 0.2μl of 0.15μM enhancer firefly luciferase plasmid, 5ng of pRL-SV40 (Promega), 0.15μl Lipofecamine3000, and 0.2μl of P3000 in 10μl of Opti-MEM. Forty-eight hours after transfection, the luciferase assay was performed with the Dual-Glo luciferase assay kit (Promega) according to manufacturer's instructions. Firefly luciferase signal was read using 96-well plate reader and signals were internally normalized to the same well's Renilla luciferase signal. Each enhancer construct was tested in 8 separate wells on at least three separate cardiomyocyte differentiations.

IPSC Reprogramming, Culturing, and IPSC-CM Differentiation. Human skin fibroblasts were obtained from Coriell (sample name GM03348, 10 year old male) and cultured in DMEM containing 10% FBS. Fibroblasts were re-programmed into induced pluripotent stem cells (IPSCs) via electroporation with pCXLE-hOCT3/4-shp53-F (Addgene plasmid 27077), pCXLE-hSK (Addgene plasmid 27078), and pCXLE-hUL (Addgene plasmid 27080) as described previously (146). IPSCs were maintained on Matrigel-coated 6-well plates with mTeSR-1 (Stem Cell technologies, Cat#85850) and passaged as colonies every 5-7 days using ReLeSR (Stem Cell technologies, Cat#05872).

IPSCs were differentiated into cardiomyocytes (iCMs) using Wnt modulation as previously described (18). Differentiation was conducted in CDM3 (RPMI 1640 with L-glutamine, 213 μg/mL L-asorbic acid 2-phosphate, 500μg/mL recombinant human albumin) (18). Cells were grown to ~95% confluency and treated with 6μM - 10μM CHIR99021 for 24 hours and allowed to recover for 24 hours. Cells were treated with 2μM Wnt-C59 for 48 hours and then media was changed with CDM3 every two days until beating cardiomyocytes were obtained (~day 6-10). In order to prevent cell detachment, beating cardiomyocytes re-plated on to new plates using TrypLE (Thermo Fisher). Media was changed every two days until downstream assays were performed (~day 20).

CRISPr Enhancer Deletion in IPSCs. To delete enhancer regions, guides targeting the 5' and 3' end of enhancer regions were designed using CRISPOR (147). Guides were ligated into pSpCas9(BB)-2A-Puro (Addgene plasmid #62988) after the U6 promoter using either Bbs1 digestion and ligation or Gibson assembly. DNA preparations of plasmid were prepared using an EndoFree plasmid kit (Qiagen), and plasmid sequences were confirmed with Sanger sequencing. IPSCs were nucleofected using the Neon transfection system (Thermo Fisher). Briefly, GM03348 IPSCs were grown to ~70% confluency and treated with mTeSR-1 containing 2μM thiazovivin (TZV) for one hour. Cells were digested with TrypLE, collected and counted. 3.75 million IPSCs per nucleofection were pelleted at 300g for 3min. Cell pellets were resuspended in 125μl of buffer R and added to an Eppendorf tube containing 1.5μg or 2.5μg of each plasmid. Cells were nucleofected in the Neon system in a 100μl tip with the following settings: 1400 V, 20 ms, 2 pulses. Nucleofected cells were expelled into a single well of Matrigel-coated 6-well plate containing mTeSR-1 supplemented with ClonR (Stem Cell Technologies, Cat#05888) and 2μM TZV. For each round, a pSpCas9(BB)-2A-GFP (Addgene plasmid #48138) control was included. Twenty-four hours later, cells were treated with mTeSR-1 containing 0.15μg/mL puromycin. The next day, selection was continued with 0.2μg/mL puromycin until no viable cells were seen in the GFP control (~2-3 days). Cells were switched to mTeSR-1 supplemented with ClonR and 2μM TZV and media was changed daily until colonies appeared (5-7 days). Colonies were picked on to 96-well plates, expanded, and split on to two duplicate plates. The first plate was used for cryopreservation in 50% mTeSR-1/ClonR/2μM TZV and 50% KnockOut Serum replacement/25% DMSO. The second plate was processed for gDNA isolation using the DirectPCR lysis reagent (Viagen, Cat#301-C) following manufacturer's instructions. Colonies were screened for successful enhancer deletion using a 3-primer PCR approach. PCR products were cloned using the TOPO TA cloning kit (Thermo Fisher) and sequenced to determine alleles present. Positive colonies were thawed from the frozen plate, expanded, re-genotyped, and used for differentiation. In cases where no homozygous deletions were obtained, a heterozygous colony was treated with a second round of CRISPr editing.

IPSC Chromosome Analysis and CRIPSr-Off Target Analysis. IPSC Chromosome analysis was conducted using the hPSC genetic analysis kit (Stem Cell Technologies, Cat#07550) following manufacturer's instructions. IPSC lines must show no amplification or deletion in at least 8 of the 9 tested sites to pass our karyotypic quality control standards. We used the output from the CRISPOR(147) guide design tool to identify the most likely off target cut sites. We selected any regions with < 3 mismatches and additional off targets that were within or near genes important for cardiac function. Primers were designed to amplify putative off target sites and regions were amplified from gene edited cell gDNA. PCR products were purified using ExoSAp-IT (Thermo) or Ampure XP beads (Beckman Coulter) and sequenced with sanger sequencing. Primers used are available upon request. Sanger traces from unedited IPSCs were compared to gene edited lines to identify any off-target changes.

IPSC-CM RNA Extraction and qPCR. At ~day 10 of differentiation, 1 million IPSC-derived cardiomyocytes were plated on a well of 12-well plate. At ~day 20, cells were washed with PBS and 400μl of TRIzol (Thermo Fisher) was added directly to the well. Cells were collected into an Eppendorf tube using a cell scraper. Trizol was kept at -80^{o}C until further processing. Six hundred μl of additional TRIzol was added to the cells and the entire sample was added to a tube containing 250μl of silica-zirconium beads. Tubes were placed in a bead beater homogenizer (BioSpec) for 1 minute and immediately cooled on ice. Samples were incubated at room temperature for 5min and then centrifuged at 12,000g for 5min to remove unhomogenized cell aggregates. Supernatant was transferred to a new tube and 200μl of chloroform was added. After vigorous shaking for 30 seconds followed by 10 min incubation with periodic shaking, samples were centrifuged at 12,000g for 15 min. The upper aqueous layer was added to an equal volume of fresh 70% ethanol and used an input to the Aurum Total RNA Mini Kit (Biorad). RNA was processed according to manufacturer's instructions including on-column DNase digestion. RNA was eluted twice with 30μl of warmed water and the concentration was measured using a nanodrop spectrophotometer.

The qScript cDNA SuperMix (Quantabio) was used to generate a 100ng cDNA library. A 1:10 dilution was used as a template in a 3-step SYBR-green qPCR region with a 57^{o}C annealing temperature. We used a panel of primers targeting cardiomyocyte references genes (*TNNT2*, *MYBPC3*, *TNNI3*, *SLC8A1*, *MYOZ2* and *GAPDH*) that passed optimization studies confirming primer specificity and efficiency. For enhancer deletion measurements, changes in *MYH6* and *MYH7* expression were calculated using the delta-delta Cq method using the geometric mean expression of cardiomyocyte reference genes.

SDS-PAGE of Myosin Heavy Chain Isoforms. We prepared a 6.25% acrylamide/bis-acrylamide(99:1) resolving gel by combining 7.5mL of 25% Acrylamide/bis-acrylamide (99:1), 5.65mL of 2M Tris pH 8.8, 16.55mL of ddH20, 300μl of 10% SDS (w/v), 312μl 10% ammonium persulfate, and 12.5μl of TEMED. The resolving gel was allowed to polymerize for 1 hour at room temperature. A 5% acrylamide/bis-acrylamide (99:1) stacking gel was prepared by combining 2mL of 25% Acrylamide/bis-acrylamide(99:1), 2.5mL of 0.5M Tris pH 6.8, 5.325mL of ddH20, 100μl of 10% SDS (w/v), 90μl 10% ammonium persulfate, and 6μl of TEMED. The stacking gel was allowed to polymerize for 8 hours. Lysates of ~day 20 iCMs were prepared and protein concentrations were quantified with the Quick-Start Bradford Protein Assay (Bio-Rad). ~7μg of protein was mixed 1:1 with 2x Laemmli Sample Buffer containing β-mercaptoethanol. Samples were loaded into the SDS-polyacrylamide gel described above and separated at 13mA for 20min, and 15mA for 21 hours. After electrophoresis, gels were fixed with a 7% acetic acid/50% methanol solution for 1 hour at room temperature. Protein was visualized with the Sypro Ruby Protein Gel Stain (Thermo Fisher) following manufacturer's instructions. Quantification of band intensities was done using Fiji(148).

Engineered Heart Tissue Generation and Measurement of Contractile Properties. Engineered heart tissues (EHTs) were generated according to previously published methods (19). iCMs were differentiated as previously described and when beating cells were present (~day 10), cells were washed with PBS and digested with TrypLE (Thermo). One million cells per EHT were centrifuged at 500g for 5min and resuspended in 65μl of EHT media (CDM3(18), containing 10% of heat-inactivated FBS, 2μM thiazovivin, 33μg/mL aprotinin, and 5U/mL penicillin/streptomycin), 25μl of 25mg/mL fibrinogen and 10μl of Matrigel (Corning). 100μl of this EHT mix was added to 3μl of 100U/mL thrombin and mixed. The whole mixture was pipetted between PDMS posts (EHT Technologies) in an EHT mold created from 2% agarose and a Teflon spacer in a 24-well Nunc plate (Thermo Fisher). Fibrin gel was allowed to polymerize for 2 hours and then 200μl of CDM3 was added to the EHT to help detach it from the mold. After 30min, the PDMS posts were lifted from the mold and the EHT was placed into a new 24 well plate containing 1.6 mL of RPMI containing B27 supplement (Thermo Fisher) and 33μg/mL aprotinin. Media was changed every other day until further processing. After 20 days of culture, videos of EHT contraction were taken on a KEYENCE BZ-X microscope at 50fps with 4x4 pixel binning. Videos were imported into Fiji and analyzed with MUSCLEMOTION macro with default settings (149). The contraction parameters for each contraction were averaged to give an EHT level measurement.

Flow Cytometry Analysis of IPSC-CM Purity. At approximately day 20 of differentiation, iCMs were collected using TrypLE (Thermo Fisher). Cells were resuspended in 1mL of PBS and added to 1mL of 8% PFA in PBS for fixing. Cells were fixed at 37^{o}C for 10min with shaking. Cells were collected by centrifugation at 600g for 5min and resuspended in 100μl ice-cold 90% methanol in PBS per 500,000 starting cells. Cells were stored at -20^o C until further processing. On the day of flow, ~1 million cells were aliquoted into two tubes containing 2mL of 0.5mg/mL BSA in PBS and pelleted. One tube was resuspended in 100μl of PBS containing 1:200 dilution of *TNNT2*-Alexa Fluor 694 (BD Pharmingen #565744) and 1:200 *MYBPC3*-Alexa Fluor 488 (Santa Cruz Biotechnology #sc-137180 AF488) and the other tube was suspended in PBS alone. Cells were stained for 1 hour at room temperature. Four mL of 0.5mg/mL BSA in PBS was added to each tube and cells were pelleted. Cells were resuspended in 100μl in PBS and analyzed on a flow cytometer. The percentage of TNNT2-positive cells was determined by using PBS only as a negative control.

Find Regulatory Variants Computational Pipeline. Figure 5 shows a schematic of the Find Regulatory Variants computational pipeline. The pipeline relies on the bedtools tool to sequentially filter the starting variant list for variants that overlap regions with epigenetic evidence of enhancer modifying potential (121). The pipeline finds variants that are predicted to disrupt or create transcription factor binding sites. In order to use find new transcription factor binding sites created by variants, we used the GATK FastaAlternaitveReferenceMaker to insert SNP variants into the reference genome(150). We then used

Homer’s scanMotifGenomeWide.pl to search for *GATA4* and *TBX5* sites in the alternative reference and kept only sites that were new(42). These additional sites were used in the pipeline alongside sites present in the unchanged reference.

Association of Enhancer Variant with Phenotypic Data. Phenotypic measurements of heart function and whole genome sequencing data were accessed as in (32). Individual measures were obtained for left ventricular internal diameter-diastole (LVIDd) and left ventricular posterior wall thickness during diastole (LVPWd) from echocardiogram reports and spanned as much as 14 years of echocardiogram data. The diagnosis of heart failure was determined by ICD9 diagnosis codes 425 and all sub-codes, and ICD10 diagnostic codes I42 and all sub-codes. Trajectory analysis of echo measurements was conducted as in (32). Briefly, we used PROC TRAJ in SAS 9.4, (151) which uses a likelihood function to assign a each individual a phenotypic cluster and probability of belonging to that cluster. An individual’s variant status was regressed against cluster probability and was controlled for genetic race (PC1-3) and sex in R.

Code Availability. All code and scripts used in this manuscript are available upon request.

Results

Integrated epigenetic analysis identifies candidate enhancer regions for *MYH7* and *LMNA*. To identify enhancer regions active in the human left ventricle, multiple datasets were overlaid including human left ventricle-derived H3K27Ac ChIP-seq and ATAC-seq as well as ChIP-seq data targeting *GATA4*, *TBX3/5*, and *NKX2.5* from multiple cell/tissue sources (complete list shown in **Table 3.1**). Promoter-capture Hi-C data from iCMs was used to identify genomic regions predicted to interact with promoters (37). We focused on regulatory regions of two genes linked to cardiomyopathy, *MYH7* and *LMNA,* since these genes display tissue specific and broad expression, respectively. Intersection of these datasets identified two enhancer clusters for *MYH7* and three for *LMNA* (**Figure 3.1**). *MYH7* cluster

Target	Dataset	Accession Number	Reference
H3K27Ac Histone Modification	Human LV- ChiP-Seq	ENCSR150QXE	Roadmap Epigenomics Consortium, et al. 2015(21)
H3K4me3 Histone Modifications	Human LV- ChiP-Seq	ENCFF045RCM	Roadmap Epigenomics Consortium, et al. 2015(21)
Open Chromatin	Human LV-ATAC-Seq	ENCFF148ZMS	ENCODE Project. 2018(22)
	iCM- ATAC-Seq	GSE85330	Liu, Q. et al. 2017(31)
p300	Human LV- ChiP-Seq	GSE32587	May, D. et al. 2012(25)
CTCF	Human LV- ChiP-Seq	ENCFF482ZNO	ENCODE Project. 2018(22)
Promoter Interactions	iCM Promoter-Capture Hi-C	E-MTAB-6014	Montefiori, L. et al. 2018(37)
TAD Boundaries	Human LV- Hi-C	GSE58752	Leung, D. et al. 2015(36)
GATA4 Binding Sites	iCM-ChIP-Seq	GSM2280004	Ang, Y. et al. 2016(26)
	HL-1- ChIP-Seq	GSM558904	He, A. et al. 2011
	Mouse LV- ChIP-Seq	GSM862697	van den Boogaard, M. et al. 2012(30)
	Computational Predictions	HOMER	Heinz, S. et al. 2010(42)
TBX5/3 Binding Sites	iCM-ChIP-Seq	GSM2280011	Ang, Y. et al. 2016(26)
	HL-1- ChIP-Seq	GSM558908	He, A. et al. 2011(28)
	Mouse LV- ChIP-Seq	GSM862695	van den Boogaard, M. et al. 2012(30)
	Computational Predictions	HOMER	Heinz, S. et al. 2010(42)
NKX2.5 Binding Sites	HL-1- ChIP-Seq	GSM558906	He, A. et al. 2011(28)
	Mouse LV- ChIP-Seq	GSM862698	van den Boogaard, M. et al. 2012(30)
	Computational Predictions	HOMER	Heinz, S. et al. 2010(42)
eRNA Expression	Human LV-CAGE-Seq	GSE147236	Gacita, A. et al. 2020(6)
Experimentally Validated Heart Enhancers	Reporter Expression in Transgenic Mouse Embryos	VISTA Enhancer Browser	Visel, A. Et al. 2007(57)

Table 3.1. Datasets used for epigenomic identification of candidate enhancers

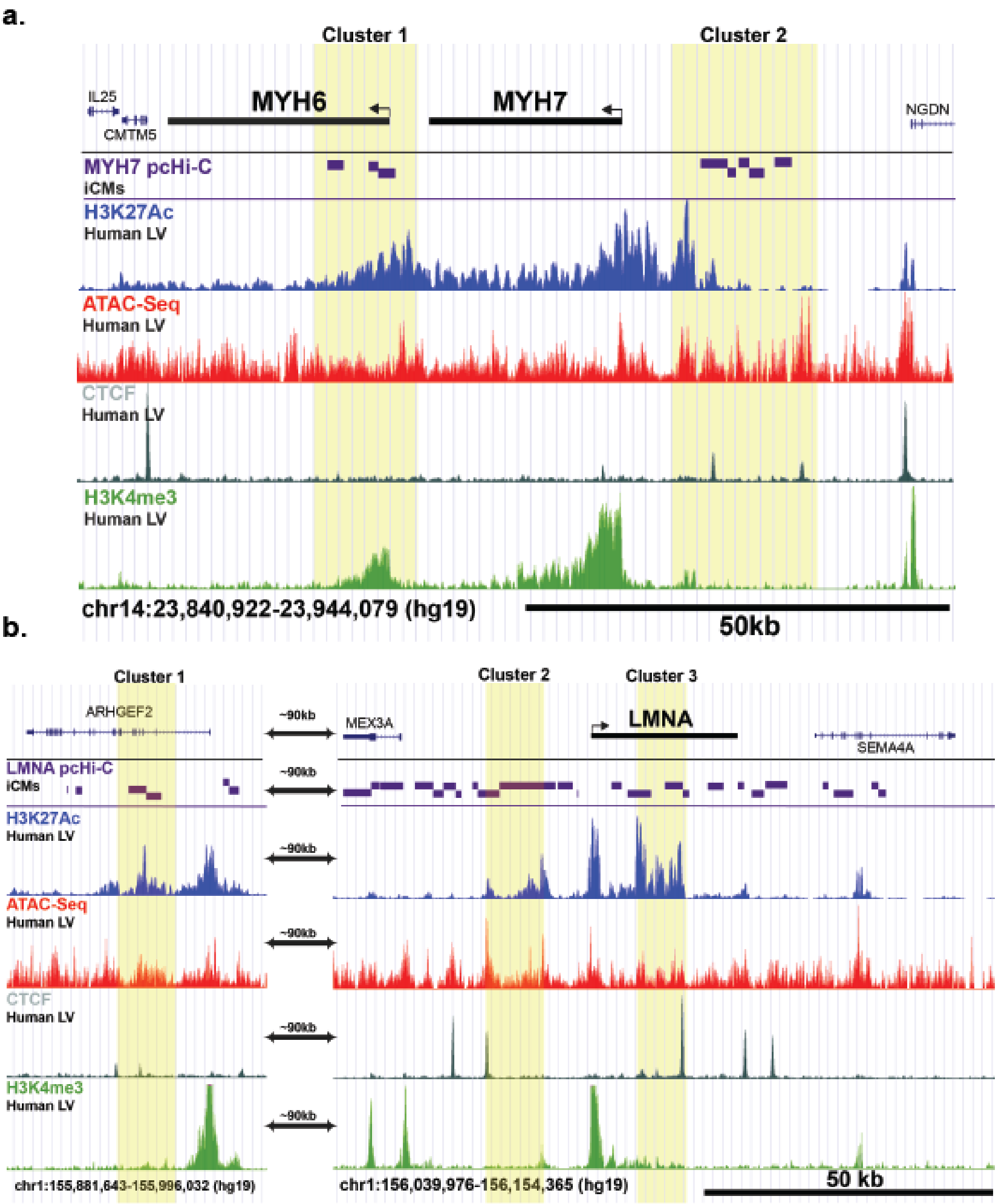


Figure 3.1. Integrated epigenomic analysis identifies candidate regulatory regions for *MYH7* and *LMNA*. *MYH7* encodes β-myosin heavy chain (MHC), the major contractile protein in the human left ventricle; mutations in *MYH7* are a leading cause of inherited cardiomyopathy. Mutations in *LMNA*, which encodes lamin A/C also contribute to inherited cardiomyopathies. **A.** The *MYH6/7* genes are in close proximity with two clusters of candidate enhancers highlighted in yellow boxes. **B.** Integrated epigenomic analysis identified three candidate enhancer clusters at the *LMNA* locus. The labels on the left indicate the data and cell/tissue source (full source listing is found in **Table 3.1).** pcHi-C, promoter capture Hi-C. LV, left ventricle. iCMs, IPSC-derived cardiomyocytes.

1 overlaps the *MYH6* promoter, consistent with their co-regulation in the left ventricle (152). *MYH7* cluster 2 is ~7kb upstream of *MYH7* and is marked by H3K27Ac, CTCF, ATAC signal, transcription factor binding and relatively low H2K4me3 marks. Although many more interactions were identified by promoter capture Hi-C, the integrated analysis highlighted three clusters for *LMNA.* Cluster 1 was located > 100kb from the *LMNA* gene within the *ARHGEF2* gene, while *LMNA* cluster 2 was located directly upstream of *LMNA*. Cluster 3 mapped to the large first intron of *LMNA*, overlapping the second exon. Similar to the *MYH7* sites, the *LMNA* sites showed H3K27Ac and CTCF marks and open chromatin enrichment. The low H3K4me3 signals differentiated these sites from promoter regions. No enhancer clusters crossed TAD boundaries defined by human left ventricle Hi-C (36).

Candidate enhancers display regulatory activity in cardiomyocytes. Individual candidate enhancer regions were assessed for regulatory activity in iCMs using a luciferase reporter assay. We used promoter-capture Hi-C data to define the boundaries of individual enhancers within clusters. Because of size, enhancers were further dissected in some cases. Four of five *MYH7* enhancer regions showed significant activity in iCMs compared to a negative control genomic desert region (**Figure 3.2A** and **Figure 3.4**). *MYH7*-C3, which is ~7kb upstream of *MYH7* had the strongest signal, consistent with its abundant H2K27Ac ChIP-seq marks. *MYH7*-C2, which overlaps the *MYH6* promoter, was active but with lower magnitude. The *MYH7*-C2 region had higher activity in mouse atrial HL-1 cardiomyocytes, consistent with its role in *MYH6* expression (**Figure 3.3**). Both of these regions also showed activity in mouse embryonic hearts in the VISTA browser (57). *MYH7*-C4, located further upstream than *MYH7*-C3 also demonstrated significant activity. For *LMNA*, five of six candidate enhancer regions showed significant activity in iCMs (**Figure 3.2B**). *LMNA* enhancer activity was generally lower than *MYH7* enhancer activity, consistent with lower *LMNA* expression in iCMs. *LMNA* C5, located at the 3’ end of *LMNA*’s large first intron showed the highest activity. This region shows low H3K4me3 signal, consistent with its role as an enhancer and not a promoter. *LMNA* C3 showed modest activity iCMs but was active in mouse embryonic hearts in VISTA (57).

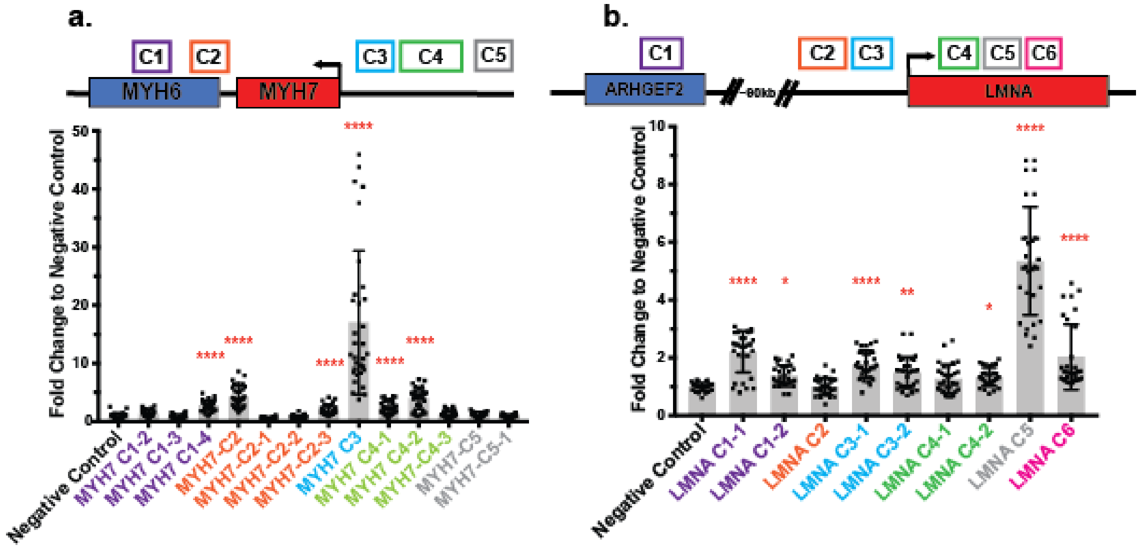


Figure 3.2. Enhancer activity in IPSC-Derived Cardiomyocytes (iCMs). A luciferase reporter assay was used to test for enhancer activity in iCMs. The position of candidate enhancers is shown along the top in colored boxes. The clusters in Figure 1 were evaluated as smaller regions. **A**. Regions from 4 of 5 candidate enhancer regions demonstrated activity in iCMs, with the highest activity for *MYH7* C3. **B.** Five of six candidate enhancer regions for LMNA showed activity in iCMs, with the highest being LMNA C5. Data is displayed as fold change to negative control 500bp genomic desert region with mean ±SD. Significance vs negative control determined by nonparametric one-way ANOVA. *<0.03, **<0.0021, ***<0.0002, ****<0.0001.

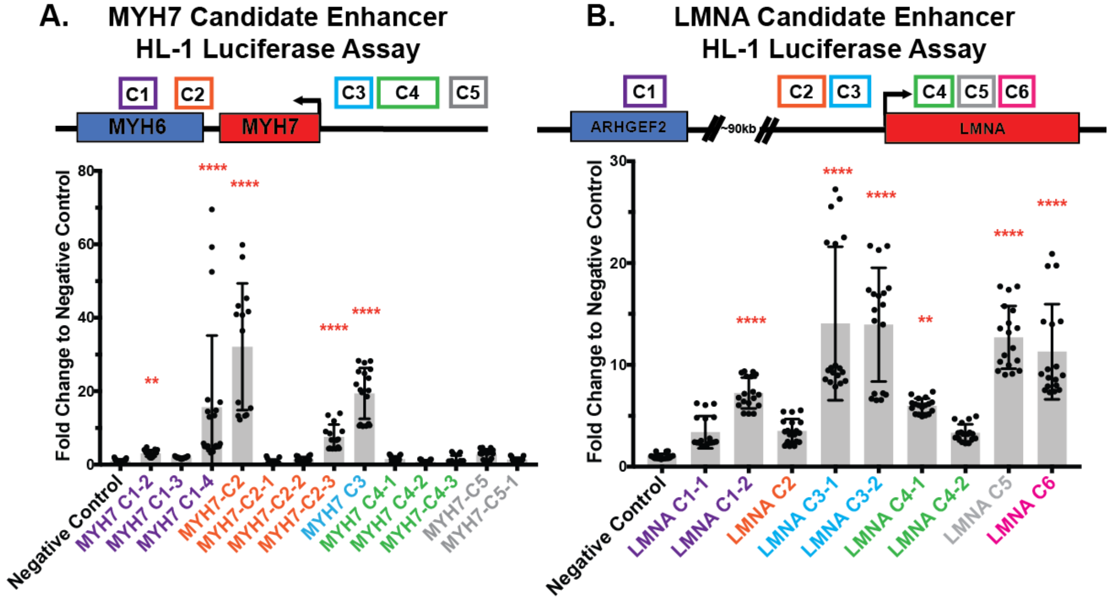


Figure 3.3. Reporter assay for candidate enhancer regions of *MYH7* and *LMNA* in HL-1 cells. A. Above, color-coded schematic of candidate *MYH7* enhancers identified in figure 1. Below, data from luciferase reporter assay in HL-1s for full and partial candidate enhancer regions. **B.** Above, color-coded schematic of candidate *LMNA* enhancers identified in figure 1. Below, data from luciferase reporter assay in HL-1s for full and partial candidate enhancer regions. Data displayed as fold change to negative control genomic 500bp desert region with mean +/- SD. Significance vs negative control determined by nonparametric one-way ANOVA. *<0.03, **<0.0021, ***< 0.0002, ****< 0.0001.

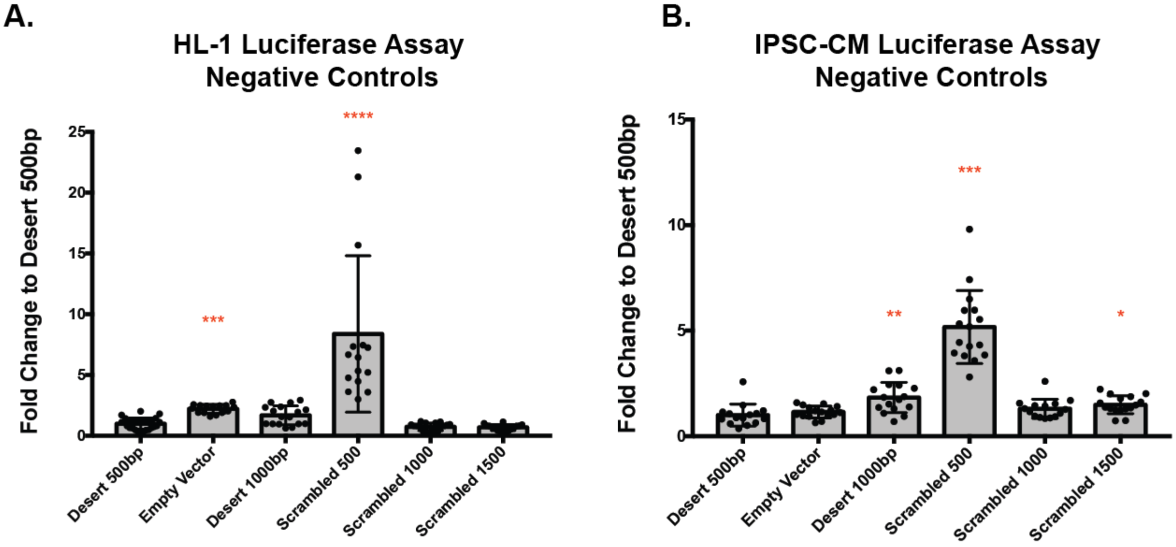


Figure 3.4. Negative control region reporter assay activity in HL-1 cells and iCMs. HL-1 cells are a mouse atrial cell line (14). Expression of *MYH6/7* differs between atrial and ventricles and between mouse and human ventricles (13). **A.** Luciferase assay using HL-1 cells including multiple negative control regions. **B.** Luciferase assay data for multiple negative control regions in iCMs. Desert represents genomic regions with little or no evidence of enhancer function in left ventricle tissue. Scrambled represents randomly selected nucleotides. Significance vs desert 500bp determined by nonparametric one-way ANOVA. *<0.03, **<0.0021, ***<0.0002, ****<0.0001.

Loss of the C3 *MYH7* enhancer shifts from *MYH7* to *MYH6*, altering protein levels and increasing contractile speed in engineered heart tissues. To test if candidate enhancers are required for target gene expression, we deleted regions of interest in IPSCs using gene editing. We focused on *MYH7*-C3 and *MYH7*-C4 due to high activity in reporter assays and intergenic position. *LMNA*-C3 was not evaluated due to low activity and the potential to disrupt *LMNA* splicing. We employed a dual cutting CRISPr-Cas9 strategy to remove the candidate enhancer regions (**Figure 3.6**). PCR genotyping confirmed the expected heterozygous and homozygous deletion in independent lines (**Table 3.4**). All edited cells passed karyotypic and off-target quality control testing (**Figure 3.7, Table 3.5**). We differentiated enhancer-deleted IPSCs into cardiomyocytes and measured *MYH7* and *MYH6* mRNA expression using qPCR. *MYH7*-C3$^{+/-}$ and $^{-/-}$ cells had a significant decrease in *MYH7* expression and increase in *MYH*6 expression, with dose-dependency (**Figure 3.5A**). We evaluated protein expression

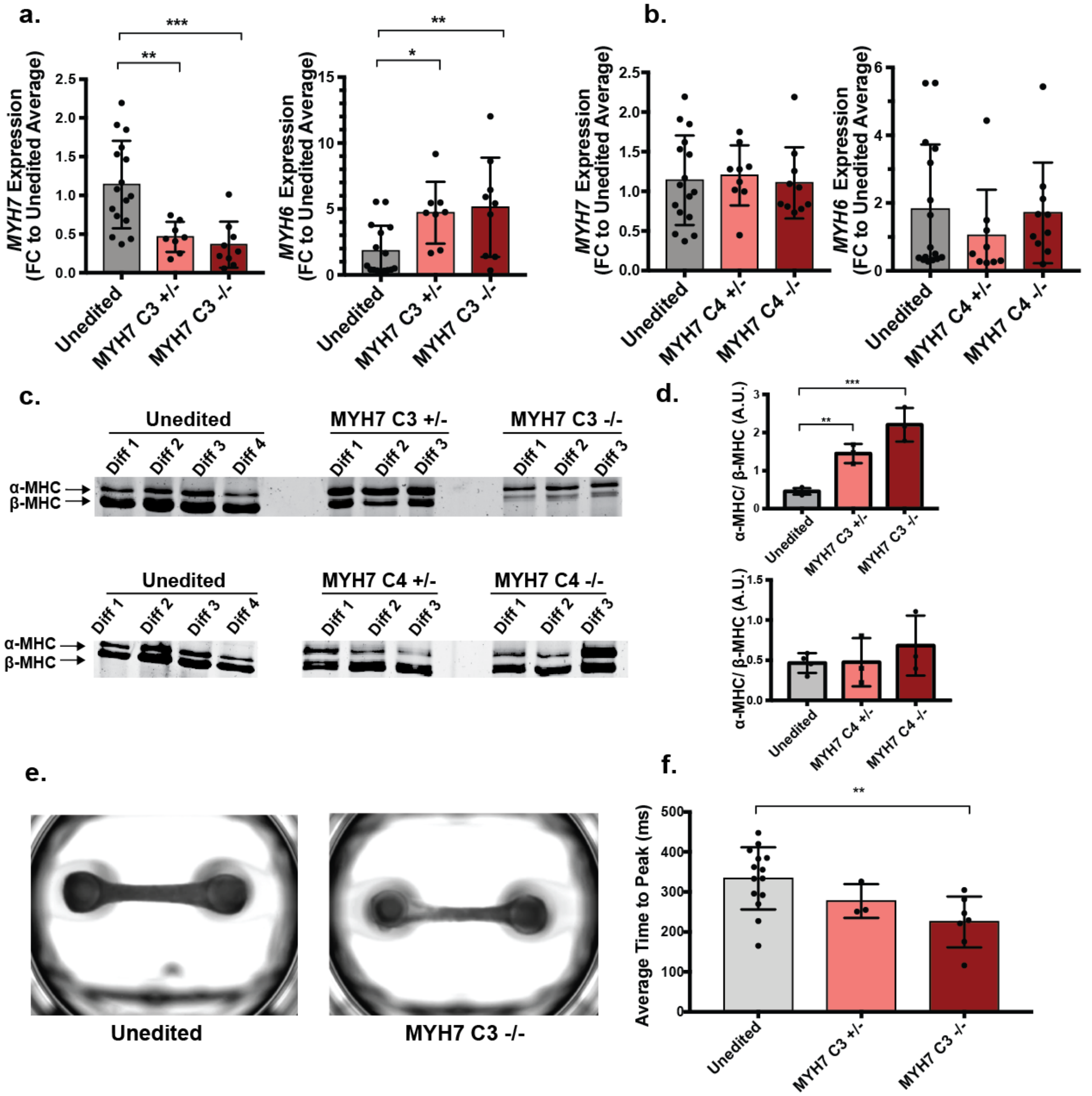


Figure 3.5. Deletion of the *MYH7* C3 enhancer increases *MYH6* and reduces *MYH7* mRNA and protein and produces hyperdynamic function in engineered heart tissues. A. Gene editing was used to delete the *MYH7* C3 enhancer heterozygously ($^{+/-}$) or homozygously ($^{-/-}$). *MYH6* and *MYH7* mRNA expression was assayed by qPCR and showed a dose-dependent increase in *MYH6* expression and reduction in *MYH7* expression. **B.** Deletion of the *MYH7* C4 enhancer had little effect, demonstrating a specificity of these findings to *MYH7* C3. **C.** α-MHC and β-MHC protein ratios were quantified using SDS-PAGE. **D.** Quantification of α-MHC/β-MHC protein ratios in C. **E**. Representative images of engineered heart tissues (EHTs) containing unedited or *MYH7* C3 homozygous deleted iCMs. **F.** Average time to peak measurements of EHT contractions containing unedited or *MYH7* C3 deleted cells showed an increase in time to peak in *MYH7* C3 deleted EHTs, consistent with the shift from *MYH7*/β-MHC to *MYH6*/α-MHC and the known faster ATPase cycle for α-MHC. Each point represents the average time to peak measurement of a single EHT across multiple contractions. All data shown as mean ±SD. * determined by one-way ANOVA. *<0.03, **<0.0021, ***<0.0002, ****<0.0001.

and found *MYH7*-C3$^{+/-}$ and $^{-/-}$ iCMs demonstrated a significant increase in the α-MHC to β-MHC protein ratio (**Figure 3.5E&F**). Loss of *MYH7*-C4 region had no effect on *MYH7* or *MYH6* mRNA or protein levels (**Figure 3.5D&E**). To ensure comparable maturity and purity, *MYH7* and *MYH6* gene expression measurements were normalized using a panel of cardiomyocyte genes in order to control for iCM purity and maturation status. Additionally, there were no significant differences between genotypes in iCM purity as measured by cardiac tropninin T (cTNT) flow cytometry (**Figure 3.7**). α-MHC, encoded by *MYH6*, hydrolyzes ATP at a higher rate than *MYH7*, which leads to a faster rate of force generation (153). We evaluated the contractile properties of engineered heart tissues (EHTs) generated from *MYH7*-C3 deleted cardiomyocytes and unedited controls (**Video 3.1 & Video 3.2**). EHTs deleted for *MYH7*-C3 showed a more rapid time to peak measurement, consistent with an increased rate of force generation (**Figure 3.5F**). Therefore, deletion of *MYH7*-C3 decreases *MYH7* and increases *MYH6,* which translates to a more hyperdynamic tissue.

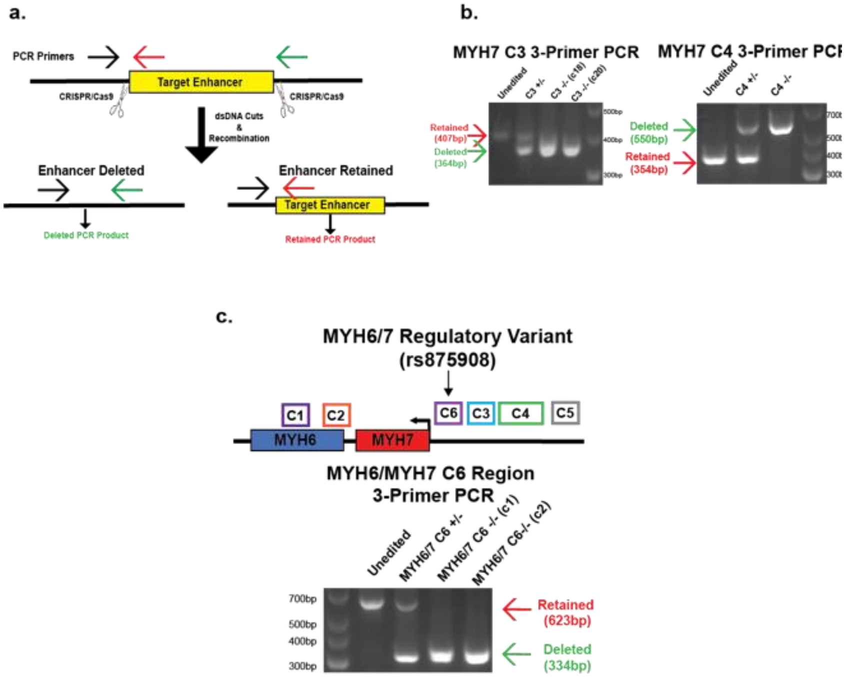


Figure 3.6. CRISPr-Cas9 enhancer deletion strategy successfully removes *MYH7* enhancer regions. A. Schematic of CRISPr-Cas9 deletion strategy and PCR primers used for genotyping. **B.** Agarose gels of 3-pimer PCRs on genomic DNA from IPSCs treated with guides targeting *MYH7* candidate enhancers 3 and 4 demonstrating successful knockout. **C.** Top, schematic representation of the location of the MYH6/7 regulatory variant. Bottom, agarose gel of 3-pimer PCR on genomic DNA from IPSCs treated with guides targeting the region overlapping the MYH6/7 regulatory variant showing successful deletion.

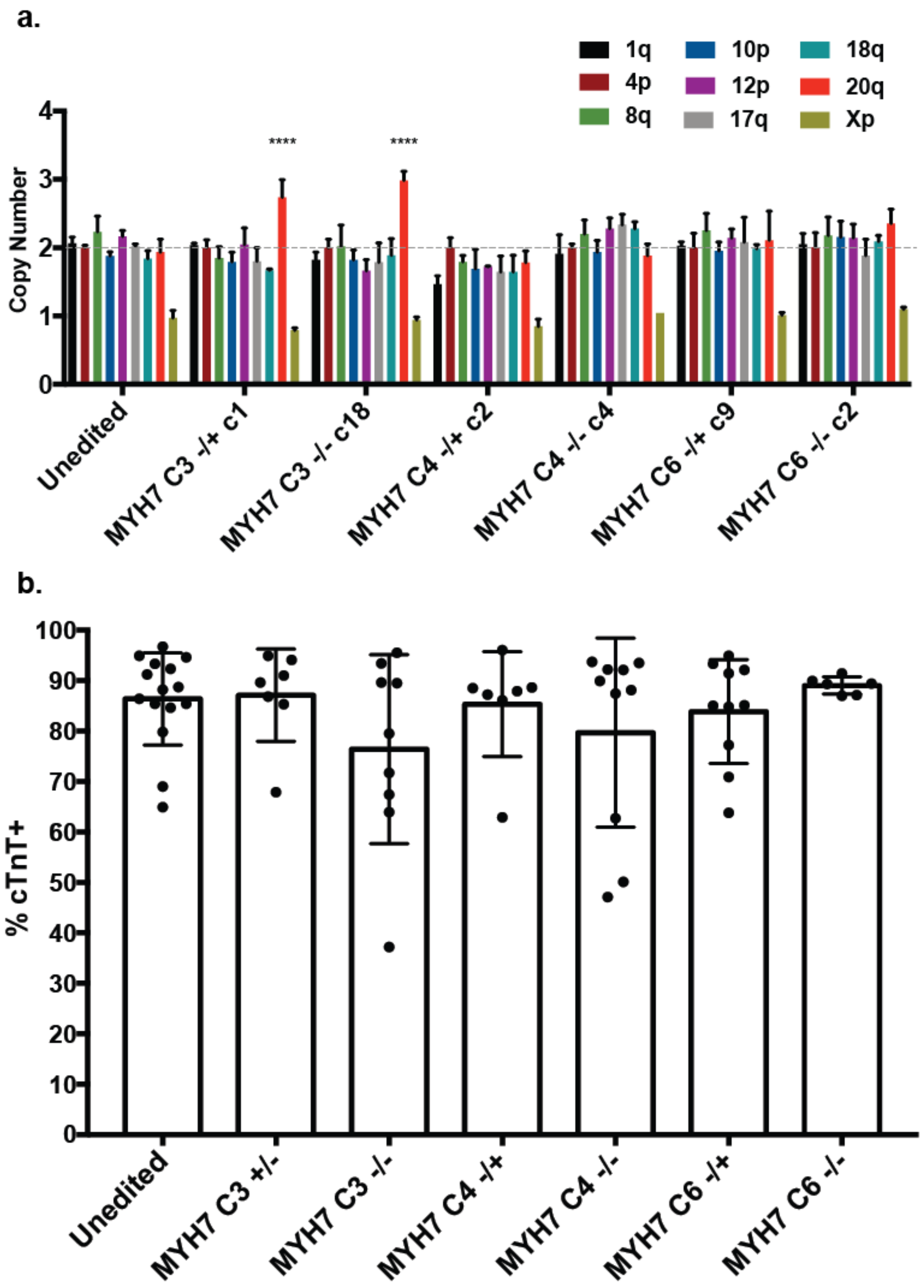


Figure 3.7. Validation of gene edited iPSCs and iCMs. Nonhomologous end joining CRISPr-Cas9 was used to generate guided deletions in iPSCs. Resulting clones were treated isolated analyzed for common chromosomal rearrangements. **A.** Results from the hPSC genetic analysis test kit (Stem Cell Technologies) assaying common chromosomal rearrangements in CRISPr treated IPSCs. **B.** iCM purity measurements evaluating the percent cardiac troponin T (cTNT) cells across different enhancer deletion lines. cTnT, cardiac troponin T. No significant differences were found between unedited and CRISPr treated cells by one-way ANOVA.

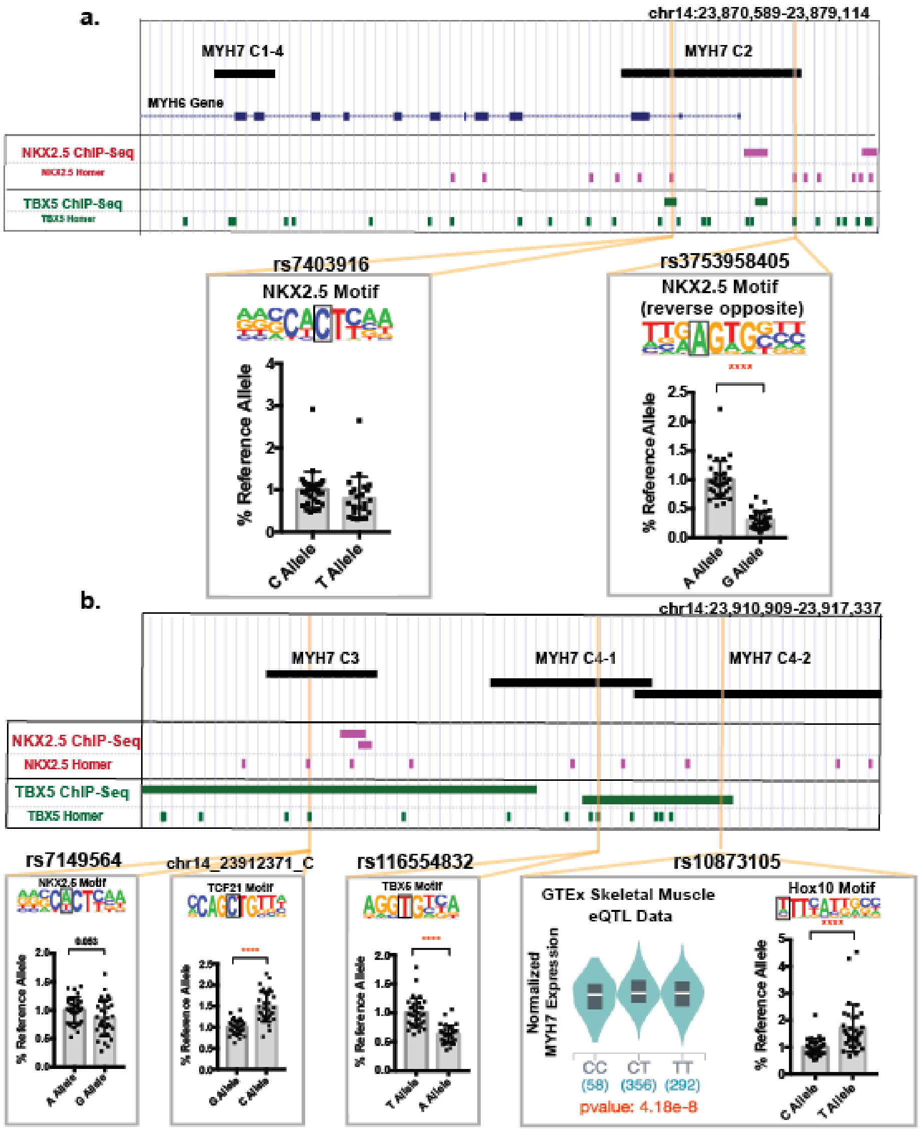


Figure 3.8. Genomic variation in *MYH7* enhancer regions. A. We queried *MYH7* enhancers for naturally occurring sequence variants for those that overlapped cardiac transcription factor binding motifs, and/or were correlated with *MYH7* expression in the GTEx eQTL dataset. rs7403916 and rs373958405 fall within *MYH7* C2 and disrupt *NKX2.5* motifs. These variants were evaluated for reporter activity in iCMs and rs373958405 demonstrates reduced activity compared to the reference allele. **B**. *MYH7* C3 contains rs7149564 and chr14_23912371_C. rs7149564 disrupts an NKX2.5 motif and results in a trending reduction in iCM luciferase signal. chr14_23912371_C generates a TCF21 motif and results in an increased iCM luciferase signal. *MYH7* C4 contains rs116554832 and rs10873105. rs116554832 disrupts a TBX5 motif and results in a reduced iCM luciferase signal. rs10873105 is correlated with *MYH7* expression in GTEx skeletal muscle data and creates a Hox10 motif. This variant results in an increased iCM luciferase signal. The ChIP-seq and homer datasets are listed in **Table 1**. All data shown as mean ± SD. Significance determined by unpaired t-test. *<0.03, **<0.0021, ***<0.0002, ****<0.0001.

Active cardiac enhancers harbor genetic variants in transcription factor binding sites. We queried *MYH7* enhancers in Figure 2 for naturally occurring sequence variants using the gnomAD database and selecting those that overlapped cardiac transcription factor binding motifs, and/or were correlated with *MYH7* expression in the GTEx eQTL dataset (55, 154). We identified six unique variants within *MYH7* enhancers that overlapped transcription factor binding motifs and were within or nearby ChIP-seq peaks showing transcription factor binding in cardiac cells (**Figure 3.8A&B, top**). We compared luciferase signals from plasmids carrying the reference or alternative allele in iCMs. A variant (rs373958405) upstream of *MYH6* disrupts a highly conserved site in the *NKX2.5* binding motif, and plasmids encoding this variant demonstrated significantly reduced signals in iCMs compared to the reference allele (**Figure 3.8A, top**

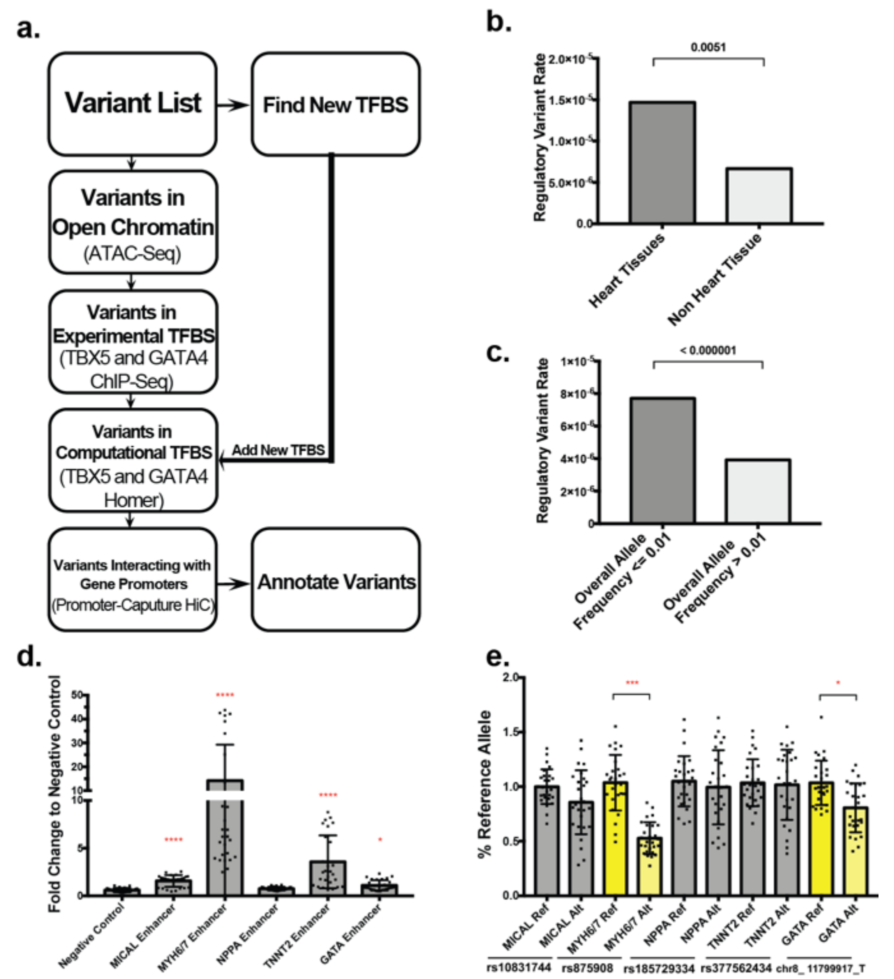


Figure 3.9. Computational pipeline to identify enhancer modifying variants. A. Schematic of pipeline filtering steps to identify enhancer modifying variants (EMVs). All datasets used were generated in iCMs (see Table 1 for more information.) **B.** This strategy disproportionately identified significant GTEx eQTLs from heart tissues versus non-heart tissues, and disproportionately identified rare alleles (**C**). Significance determined in B & C by Fisher's exact test. **D.** Luciferase reporter assay in iCMs for selected regions containing variants of interest identified through this analysis. Significance vs negative control was determined by nonparametric one-way ANOVA. **E.** Luciferase signal for reference and alternative alleles of selected variants identified through this pipeline. Positive enhancer modifying variants highlighted in yellow. Significance determined by unpaired t-test. All data shown as mean ± SD. *<0.03, **<0.0021, ***<0.0002, ****<0.0001.

right). Within *MYH7*-C3, we identified rs7149564 which disrupted a less conserved site in the *NKX2.5* motif, and consequently, showed a modest trending reduction in luciferase signal (**Figure 3.8B, bottom left**). A nearby variant (chr14_23912371_C), also in *MYH7*-C3, creates a *TCF21* motif and correlated with higher luciferase activity, relative to reference. Within *MYH7*-C4, a variant (rs116554832) that overlapped a highly conserved site within a *TBX5* motif resulted in a reduction in signal (**Figure 3.8B, bottom middle**). A second C4 variant (rs10873105) correlated with *MYH7* expression in GTEx skeletal muscle samples. This variant generates a *Hox10* motif and causes an increased signal in reporter assays (**Figure 3.8B, bottom right**). These enhancer modifying variants (EMVs) are positioned to regulate cardiac function.

Genomewide evaluation of enhancer modifying variants identifies variants controlling the activity of cardiac enhancers. Since we identified EMVs within *MYH7* enhancers, we sought variants regulating other cardiac genes by applying this strategy genomewide. We created a computational filtering pipeline to use publicly available data from iCMs to identify variants within enhancer regions that alter transcription factor binding (**Figure 3.9A**). We benchmarked this pipeline using variant sets from GTEx and gnomAD (55, 154). As expected, eQTLs in heart tissues were more likely to be found using this strategy (**Figure 3.9B**). Rare variants were also more likely to survive the filtering steps of this pipeline, consistent with transcription factor binding sites within enhancer regions being under greater constraint (**Figure 3.9C**). From the surviving gnomAD variants, we selected five that were predicted to regulate *MICAL2*, *MYH6*, *NPPA*, *TNNT2*, and *GATA4*, which are proteins important for cardiac function or development (155-158). The genomic regions harboring these variants were tested for activity in iCMs, and four of the five variants overlapped active enhancer regions (**Figure 3.9D**). We tested expression of the reference and alternative alleles in iCMs. The alternative allele of variants predicted to regulate *MYH6* and *GATA4* showed significantly reduced function.

A variant ~2kb upstream of *MYH7* correlates with cardiomyopathic features in longitudinal echocardiographic imaging. rs875908, which was predicted to regulate *MYH6* by the computational

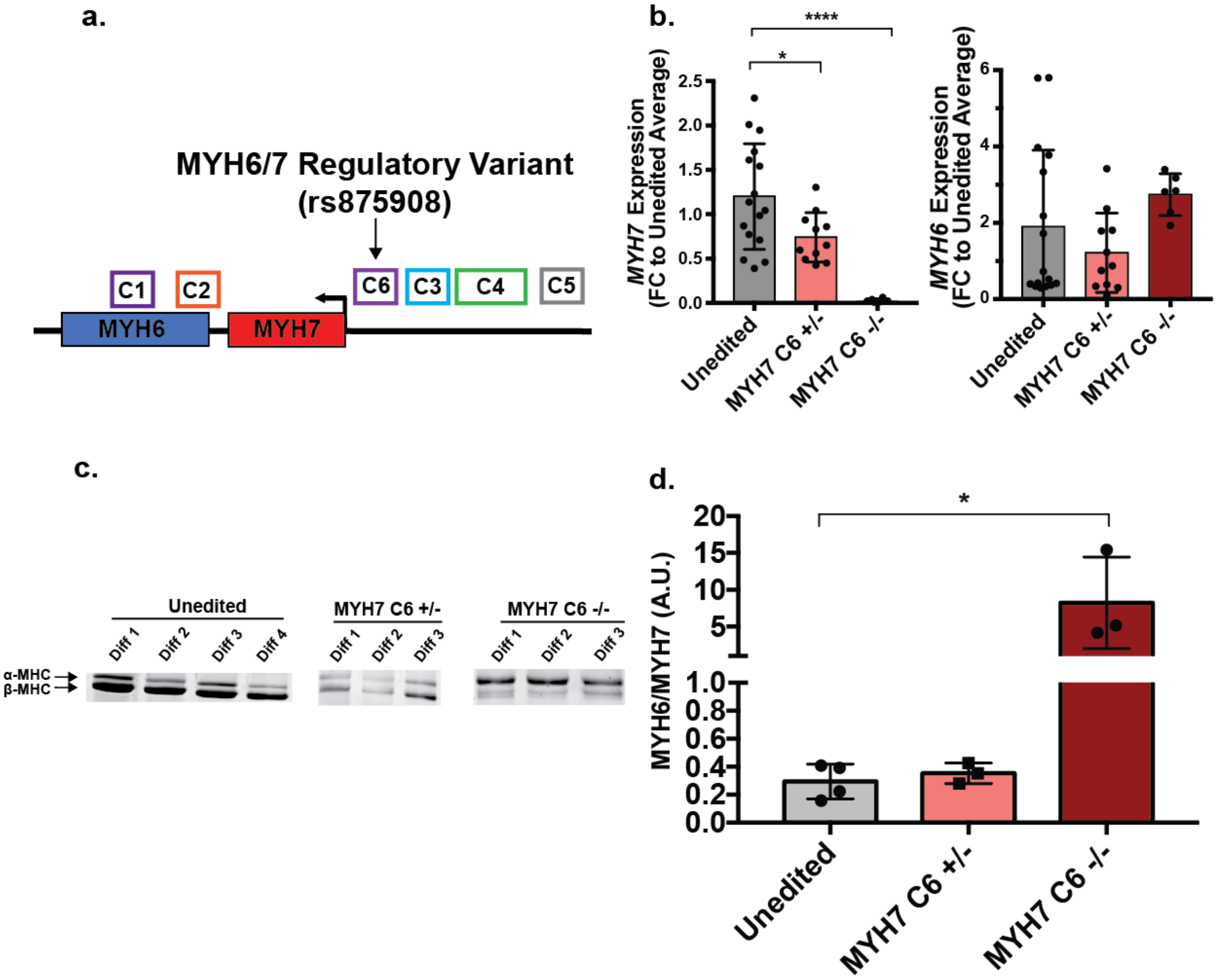


Figure 3.10. Deletion of the C6 enhancer region alters *MYH6/7* expression. A. Schematic demonstrating the location of the *MYH6/7* C6 enhancer region, which contains rs875908. **B.** iCM *MYH6* and *MYH7* expression levels in cells deleted heterozygously or homozygously for the C6 enhancer region containing rs875908. *MYH6*/7 levels were assayed by qPCR, and show a dose-dependent reduction in *MYH7*. **C**. SDS-PAGE analysis of myosin heavy chain protein isoforms in *MYH7* C6 -/+ and -/- cells. **D.** Quantification of α-MHC/β-MHC ratios in C. Significance determined by one-way ANOVA. *<0.03, **<0.0021, ***< 0.0002, ****<0.0001.

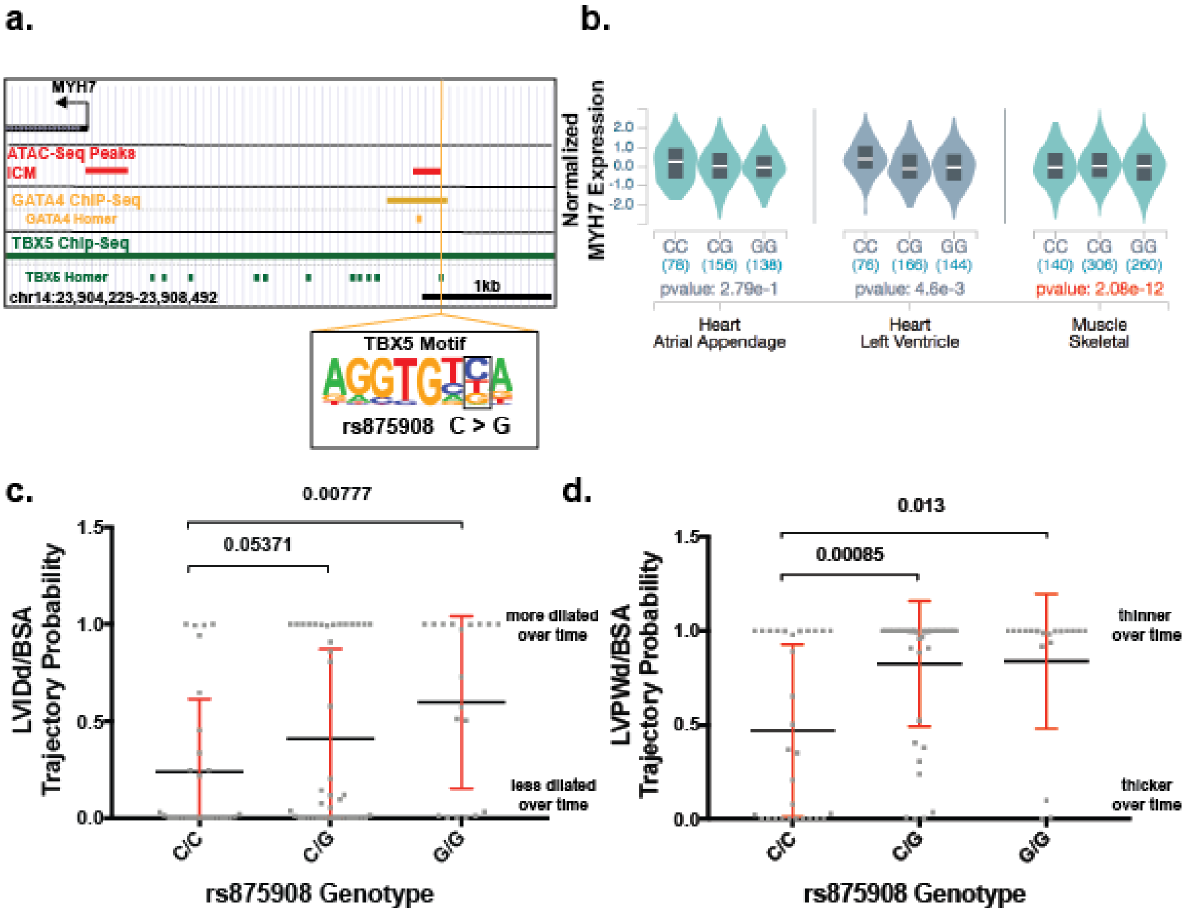


Figure 3.11. Correlation of *MYH6/7* rs875908 EMV with *MYH7* mRNA cardiac expression and longitudinal shift in left ventricular dimensions over time. A. UCSC genome browser screenshot showing the location of the *MYH6/7* regulatory variant and overlaps with epigenetic datasets. **B.** eQTL data from the GTEx project correlating regulatory variant genotype and *MYH7* expression in three muscle tissues. **C.** Association of variant status with LVIDd/BSA over time in cardiomyopathy cases from NU genomes cohort. **D.** Association of variant genotype with LVPWd/BSA over time in in cardiomyopathy cases from NU genomes cohort. Significance determined using a linear regression model corrected for race and sex. LVIDd/BSA, left ventricular internal diameter during diastole corrected for body surface area. LVPWd/BSA, left ventricular posterior wall thickness during diastole corrected for body surface area.

pipeline, is an EMV located ~2kb upstream of *MYH7* (**Figure 3.10A**). We deleted the region harboring this variant, *MYH7*-C6, in IPSCs (**Figure 3.6**). Heterozygous removal of this region in iCMs caused a reduction in *MYH7* expression but no change in *MYH6* expression in iCMs (**Figure 3.10B).** Homozygous deletion of this region showed an ~100 fold reduction in *MYH7* expression levels and a qualitative, but no significant increase in *MYH6* levels (**Figure 3.10B**). Homozygous deleted cells also showed a significant increase in the α/β-MHC protein ratio (**Figure 3.10C&D**). The rs875908 variant was identified because it is bound by *GATA4* and *TBX5* and also disrupts a *TBX5* motif (**Figure 3.11A**). GTEx eQTL data shows this variant correlates with lower *MYH7* expression in skeletal muscle with trending significance for expression in left ventricle (**Figure 3.11B**).

To ascertain whether rs875908 correlates cardiac outcomes, we evaluated trajectory probabilities of left ventricular dimensions over time using genomic and echocardiographic information derived from the Northwestern biobank. This approach assigns a probability of maintaining an echocardiographic change overtime (32). The rs875908 variant correlated with a more dilated left ventricle over time in participants selected with cardiomyopathy diagnosis codes (**Figure 3.11C**). This correlation was not observed when using clinical data from nonselected biobank participants (**Figure 3.12**). The rs875908 also variant correlated with a thinner left ventricle posterior wall thickness at end-diastole (LVPWd) over time in those with cardiomyopathy diagnostic codes (**Figure 3.11B**). Variant association with left ventricular wall thickness was also present

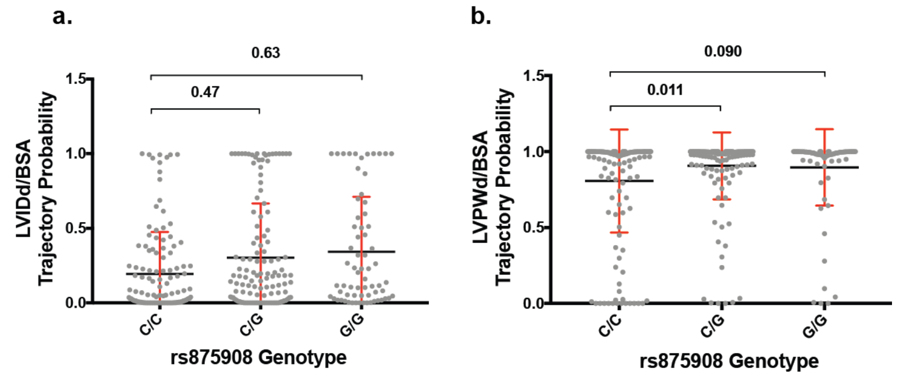


Figure 3.12. Phenotypic regressions using the NU genomes cohort. A. Association of variant status with LVIDd/BSA over time in the NU genomes cohort (n=387). **B.** Association of variant genotype with LVPWd/BSA overtime in the NU genomes cohort. Significance determined using a linear regression model corrected for race and sex. LVIDd/BSA, left ventricular internal diameter during diastole corrected for body surface area. LVPWd/BSA, left ventricular posterior wall thickness during diastole corrected for body surface area

with all subjects, but with a weaker signal (**Figure 3.12**). The majority of these diagnostic codes were for dilated cardiomyopathy, in which a thinner wall over time translates to a more diseased heart. These data support that the EMV rs875908 correlates with a more severe dilated cardiomyopathy phenotype.

Discussion

An *MYH7/6* Super Enhancer. Promoter capture Hi-C data from human cardiomyocytes (37) indicates that the *MYH7* and *MYH6* gene promoters contact each other within 3-dimensional space. Further, an enhancer cluster positioned ~7kb upstream of *MYH7* also interacts with the *MYH7* gene promoter. Since multiple individual parts of this enhancer cluster have activity in human cardiomyocytes, it is likely this cluster represents a super enhancer (159). Super enhancers are known to regulate critical cell identity genes (160). We now showed that deletion of the C3 enhancer region reduced *MYH7* expression in iCMs, and, correspondingly, deletion of the C3 enhancer increased *MYH6* expression resulting in an αMHC/βMHC ratio that increased heart function in EHTs. These data favor a model where the *MYH6* and *MYH7* promoter regions form a 3-dimensional complex with the super enhancer upstream of *MYH7* (**Figure 3.13**). In this model, the super enhancer, containing C3 and additional *MYH7*-specific enhancer regions induce *MYH7* expression, which is critical during heart development, and this same region may

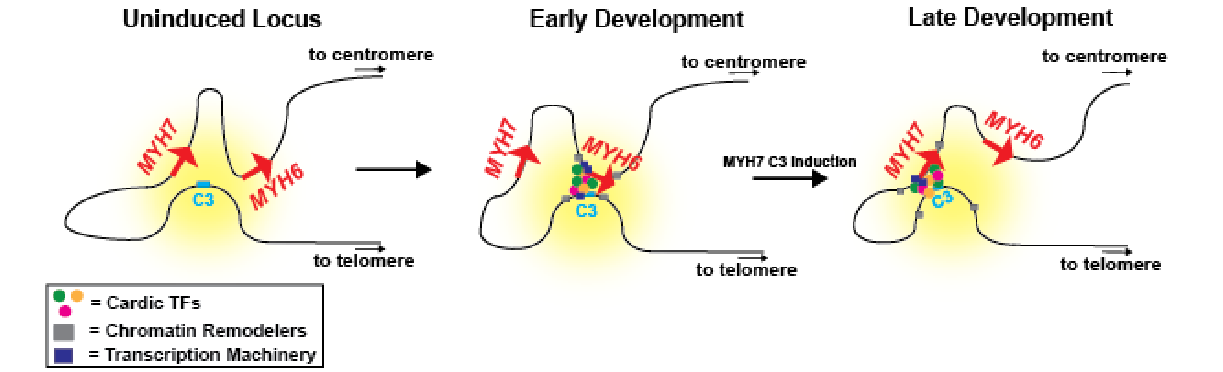


Figure 3.13. Schematic diagram of the *MYH6/7* locus during cardiac development. At the uninduced locus, the *MYH7* and *MYH6* promoter regions form a complex with a super enhancer containing the MYH7 C3 enhancer (the effect of the super enhancer is depicted as the yellow hue). In early development, this super enhancer is primarily in contact with the *MYH6* promoter, which recruits cardiac transcription factors, chromatin remodelers, and transcription machinery to drive *MYH6* expression. During development, activation of the MYH7 C3 region reorganizes the complex and the *MYH7* promoter preferentially contacts the regulatory regions. Through competition for transcriptional machinery or through an independent separate mechanism, the *MYH6* gene is downregulated. TFs, transcription factors

be employed in heart failure. The increase in *MYH6* expression that we observed may be due to an inhibitory function in C3 or an independent mechanism that compensates for reduced *MYH7* expression. These findings are reminiscent of the murine *Scn5A-Scn10A* locus, a region important for regulating electrical control of the heart (97).

Integrated genomics to identify EMVs. The rs875908 variant upstream of *MYH7* mapped to the C6 enhancer region. This variant correlated with lower *MYH7* expression levels and with having a more severe dilated cardiomyopathy phenotype over time, as marked by a more dilated, thinner walled ventricle. The *MYH6/7* ratio is known to shift during heart failure, with end stage hearts exhibiting an increase in *MYH7* and a decrease in *MYH6*. With prolonged shift of myosin expression, or a specific magnitude of shift, this change in myosin expression may actually contribute to heart failure (161). Supporting this, the *MYH6*/*7* ratio has previously been implicated in heart failure phenotypes (162). A distinct contributory mechanism could involve variants within *MYH6/7* enhancers, variants in linkage disequilibrium or even pathogenic coding mutations. Varied expression of pathogenic *MYH7* mutations has been shown to affect cardiomyopathy phenotypes (163, 164). A region related to the C6 enhancer, containing the EMV rs875908, was previously deleted in a mouse. Mice missing this C6 orthologous region had reduced *MYH7*/β-MHC but no change in *MYH6*/α-MHC (115), similar to what was shown here in human cells. This study measured *MYH7* expression in the mouse embryonic heart, which differs from the human developing and mature heart. Consistent with the human genetic findings, mouse hearts lacking this enhancer region demonstrated reduced fractional shortening and higher amounts of myofiber disarray, which additionally support the functionality of this region.

As deep sequencing data of intergenic regions becomes more available, the importance of noncoding annotation of disease genes will become vital and permit the integration of this information into clinical care. Targeted assessment of EMVs annotated by specific epigenetic marks can have clinical utility.

Name	Sequence(s)	Experiments/Notes
MYH7_C3_KO_G1	GCCTAGAAGTCCGGACACCG	Guide used to remove C3 Enhancer
MYH7_C3_KO_G2	GTGGTGTGGAACAAAGCGAA	Guide used to remove C3 Enhancer
MYH7_C3_KO_Homo_G1	CCTAGAAGTCCGGACAACCG	Guide used to remove C3 Enhancer in het. IPSCs
MYH7_C3_KO_Homo_G2	TGGTGTGGAACAAAGCCGAA	Guide used to remove C3 Enhancer in het. IPSCs
MYH7_C4_KO_G1	ATGGGATTGTGAACAGCGGA	Guide used to remove C4 Enhancer
MYH7_C4_KO_G2	CGTGATTTGGACTGGCGATC	Guide used to remove C4 Enhancer
MYH7_C6_KO_G1	CAGAGCCTCCCAAACCCGAA	Guide used to remove C6 Enhancer
MYH7_C6_KO_G2	TTTGTGGGGAGTGACCGGTC	Guide used to remove C6 Enhancer
MYH7_C6_KO_Homo_G1	CAGAGCCTCCCAAACCGAA	Guide used to remove C4 Enhancer in het. IPSCs
MYH7_C3_3PrimerGenotyping_Mix	1.AAGACAGTGGAGTGACGAGG 2.AAAGACCTCTAGTGCACCCC 3.AGAAGAGAACGAAGCGGGAA	Primers used for genotyping C3 enhancer KO IPSCs
MYH7_C4_3PrimerGenotyping_Mix	1.GAGAGGGTGGAGGAGGGT 2.TGCATTCCAGGCTGAGTGA 3.CCCCTTGGTACTGTCCTCAC	Primers used for genotyping C4 enhancer KO IPSCs
MYH7_C6_3PrimerGenotyping_Mix	AAAGGGTGCTTGGGACGTAG CCTCACTCTCCCCACAAGG GCCTGAGTAGCCCTGGAAA	Primers used for genotyping C6 enhancer KO IPSCs
hsMYH7_qPCR	F.GCAGCTAAAGGTCAAGGCC R.AGCTACTCCTCATTCAAGCC	Gene expression in IPSC-CMs Efficiency= 1.04
hsMYH6_qPCR	F.AAGTCCTCCCTCAAGCTCATGGC R.ATTTTCCCGGTGGAGAGC	Gene expression in IPSC-CMs Efficiency= 0.96
hsTNNT2_qPCR	F.AGGAGACCAGGGCAGAAGATG R.CTGGGCTTTGGTTTGGACTCC	Gene expression in IPSC-CMs Efficiency=0.98
hsMYBPC3_qPCR	F.CCCCATCTGAGTACGAGCG R.AGCCAGTTCCACGGTCAG	Gene expression in IPSC-CMs Efficiency= 0.95
hsSLC8A1_qPCR	F.AGTGCTGGGGAAGATGATGACGACG R.AGGATGGAGACAATGAAACACGCCC	Gene expression in IPSC-CMs Efficiency= 1.02
hsTNNI3_qPCR	F.CGTGTGGACAAGGTGGATGA R.CCGCTTAAACTTGCCTCGAA	Gene expression in IPSC-CMs Efficiency=1.06
hsMYOZ2_qPCR	F.AACACCCCAGATCCACGAAG R.GCCTCTAAAAGCTCCGGATC	Gene expression in IPSC-CMs Efficiency=1.02
hsGAPDH_qPCR	F.GTGGACCTGACCTGCCGTCT R.GGAGGAGTGGGTGTCGCTGT	Gene expression in IPSC-CMs Efficiency= 0.96

Table 3.2. Guides and primers used in this study.

Region Name	Primers	Size (bp)	Coordinates (hg19)
MYH7-C1-2	AGTTCAGCCCCATGAGGTAG GGTACCGAGGCGAGGGATATGGTGAAGG	673	chr14:23870150-23870823
MYH7-C1-3	GGGTCAGGTCTTTCACAAGC TTTTCCTCCTGTGCCCAAGAC	698	chr14:23870761-23871458
MYH7-C1-4	TCTTGGGCACAGGAGGAAAATTC TCCCTTCCTCCATTCACCC	697	chr14:23871436-23872136
MYH7-C2	CTGGCCTTGGCTTTTCTCCAG CAAACCAGGGTGGCCTCAAG	2072	chr14:23876121-23878188
MYH7-C2-1	AAACCTCCTCTTACCTGGGC TTGGGGAACAGAAGGAGACC	694	chr14:23877446-23878141
MYH7-C2-2	GCCCTACTCACCTTCCCATTC TGCCTCTCTGCTTCTAACCC	838	chr14:23876221-23877058
MYH7-C2-3	ACCTGGTTATCCCTTCACGG TGTCACCTCCAGAGCCAAAGG	844	chr14:23876782-23877626
MYH7-C3	GBLOCK	961	chr14:23912000-23912961
MYH7-C4-1	TGTTCACAATCCCATCCCCA AGTGGGTCTCTGAAAAGGCA	1400	chr14:23913940-23915344
MYH7-C4-2	TGGCTGGATTCCTGATGTG CGGACTTTGCCCTTCATAGCACC	2209	chr14:23915187-23917391
MYH7-C5	GCCAGAGGCTGAGCGTGAATTAG GCAATTTGAATATGATATGCCCAGG	2223	chr14:23922666-23924886
MYH7-C5-1	GBLOCK	790	chr14:23923381-23924171
LMNA-C1-1	CCTGTCCTGGAGTGGCTAAATC GGGCAGGGGTTAGAATTCCTG	1156	chr1:155937201-155938359
LMNA-C1-2	CATTCGGACTCTCTCTCCCC TTTAGCCACTCCAGGACAGG	1210	chr1:155936009-155937220
LMNA-C2	GTTAGGTGCCGGGTTTTCTG TGATATGTGCATGTACGGCG	928	chr14:23904382-23905597
LMNA-C3-1	CTCTCTCGTCCATCCTCCAC GCTCCTCTTCGGGTCTTGAAAG	1108	chr1:156074366-156075480
LMNA-C3-2	ACTCCTCTAACAGCTGTGGG CCCCTTGGTGAATGGATCCA	1199	chr1:156073216-156074415
LMNA-C4-1	GAAAGGGATTGGAGCGGAAAG CAGCAGCCCCTTAACTCTC	1211	chr1:156092084-156093294
LMNA-C4-2	TAACACTGCCACCTTCTGC TTGGCTAGTCTGTGGGTCTG	1392	chr1:156093103-156094494
LMNA-C5	TGAGATCACCTGGGCGAC AGAAGGGCTGGGCATCCTG	850	chr1:156095724-156096574
LMNA-C6	CCAGAAAAGGTGAGGGAGGTG GGGAGGGCCTAGGTAGAAGAG	1101	chr1:156099538-156100640

Table 3.3. Luciferase constructs tested in iCMs and HL-1 cells

Target	Clone	Genotype Call	Allele 1		Allele 2	
			Guide 1 Site	Guide 2 Site	Guide 1 Site	Guide 2 Site
MYH7 C3	1	Heterozygous	WT +1 (T)	WT +1 (C)	Deletion +0	Deletion +0
MYH7 C3	18	Homozygous	Deletion +2 (CT)	Deletion +2 (CT)	Deletion -1 (T)	Deletion -1 (T)
MYH7 C4	2	Heterozygous	WT +1 (G)	WT +0	Deletion -6	Deletion -3
MYH7 C4	4	Homozygous	Deletion +0	Deletion +1 (G)	Deletion +0	Deletion -10
MYH7 C6	9	Heterozygous	WT +1 (G)	WT +0	Deletion -9	Deletion -10
MYH7 C6	2	Homozygous	Deletion +0	Deletion +0	Deletion -9	Deletion -10

Table 3.4. Genotypes of enhancer deleted cells as determined by Sanger sequencing.

#	Guide	#Mismatches	Location (hg19)	Annotation (Gene)	Result
1	MYH7_C3_KO_G1	3	chr3:43948037-43948059-	Intergenic (RP4-672N11.1-RP4-555D20.3)	Negative
2	MYH7_C3_KO_G2	2	chr3:8242570-8242592:+	Intron (LMCD1-AS1)	Negative
3	MYH7_C3_KO_G2	3	chr2:196514628-196514650:-	Intron (SLC39A10)	Negative
4	MYH7_C3_KO_G2	3	chr4:25160613-25160635:-	Exon (SEPSECS)	Negative
5	MYH7_C3_KO_G2	4	chr2:179579180-179579202:-	Exon (TTN)	Negative
6	MYH7_C4_KO_G1	2	chr2:237534886-237534908:-	Intergenic (ACKR3-AC011286.1)	Negative
7	MYH7_C4_KO_G1	3	chr18:55971940-55971962-	Intron (NEDD4L)	Negative
8	MYH7_C4_KO_G1	3	chr1:237019608-237019630:+	Intron (MTR)	Negative
9	MYH7_C4_KO_G1	4	chrX:33011884-33011906:+	Intron (DMD)	Negative
10	MYH7_C4_KO_G2	3	chr5:37853503-37853525:+	Intron (GDNF-AS1)	Negative
11	MYH7_C4_KO_G2	3	chr2:43370103-43370125:-	Intergenic (AC093609.1-THADA)	Negative
12	MYH7_C6_KO_G1	3	chr2:19143341-19143363:+	Intergenic (AC106053.1-AC092594.1)	Negative
13	MYH7_C6_KO_G1	3	chr9:134519754-134519776:-	Intron (RAPGEF1)	Negative
14	MYH7_C6_KO_G2	3	chr22:18336723-18336745:+	Intron (MICAL3)	Negative
15	MYH7_C6_KO_G2	4	chr2:224012528-224012550+	Intron (KCNE4)	Negative
16	MYH7_C6_KO_G2	4	chr1:32712994-32713016:-	Exon (FAM167B)	Negative

Table 3.5. Off target analysis in CRISPr-treated IPSCs**.**

Chapter 4.

A transcriptional method for assaying IPSC-derived cardiomyocyte purity and maturity level

Abstract

The field of cardiac genetics is missing a model system that closely recapitulates human left ventricular biology. Induced pluripotent stem cell derived cardiomyocytes (iCMs) offer an alternative, but are limited by technical challenges including variable purity and maturation. When measuring the effect of noncoding variation on cardiac gene expression, transcriptional maturity is vital. In order to study the regulation of the *MYH6-MYH7* gene cluster, we set out to determine the best normalization factors for *MYH6/7* expression in iCMs. Using publicly available RNA-seq data of iCM differentiation we identified gene clusters that shared temporal expression patterns. We tested genes that correlated with *MYH6/7* expression values during differentiation as normalization factors. We differentiated an IPSC line 30 times and tested the ability of our normalization factors to reduced MYH6/7 expression variability. Our normalizers performed well for *MYH6/7* and can likely be applied to other genes that are expressed adequately in iCMs.

Introduction

The field of cardiac genetics is in need of a human model system that can adequately evaluate genetic perturbations. Mouse model systems are powerful, but physiological, genomic, and transcriptomic differences between human and mouse hearts can complicate translation (12, 165). Current cell line models fail to express many important cardiac genes (14, 166). Induced pluripotent stem cell (IPSCs) technology offers an alternative. IPSCs can be generated from a variety of human somatic cell sources. Using a Wnt modulation protocol, IPSCs can be differentiated into cardiomyocyte-like cells (iCMs) that express many cardiac genes, contract in vitro, and have measurable action potentials and calcium transients(18). While iCMs recapitulate some features of adult cardiomyocyte biology, their transcriptional phenotype more closely resembles immature fetal cardiomyocytes (20, 167). Additionally, the process of iCM differentiation is complex and prone to technical variations both between and within labs. The variability of iCMs is a major challenge and needs to be addressed before iCMs can reach their full potential as research tools and potential therapeutics.

There is currently no clear agreement in the cardiac field as to what defines an adequately mature iCM. In studying the effect of noncoding variation, the transcriptional maturity of iCMs is of paramount importance. When compared to human ventricular cardiomyocytes, iCMs show lower expression of many important tissue-specific genes, including *MYH7*, *TTN*, *TNNI3*, *SCN5A*, and *RYR2* (168). These genes are of particular focus because mutations in these genes are known to cause cardiomyopathy and arrhythmias (1, 169). Many biochemical and physical techniques have been employed improve iCM maturity (for a review see (168)), but the field has not settled on a standard toolset.

iCM purity is related to, but distinct from iCM maturity. Purity refers to the percentage of cells that have committed to the iCM lineage. Maturity is more concerned with how the iCMs change their functional properties overtime. However, in order to determine iCM purity, they must be mature enough to differentiate from non-cardiomyocyte cells. The variability of the differentiation process can affect iCM purity/maturity between groups and lead to spurious findings. Currently, the field has settled on assaying the percentage of cells expressing cTnT (cardiac troponin T) protein, which is cardiomyocyte specific, to assay iCM purity. Standard protocols can consistently generate iCM populations with 80-90% cTnT+ cells and biochemical methods have been reported to increase that percentage to > 99% (170). Transcriptionally, *TNNT2* is expressed early during differentiation and therefore may be a good marker of cardiomyocyte-committed cells. However, cTnT positive cells may still be quite immature and fail to demonstrate expected phenotypes.

The myosin heavy chain genes, *MYH6* and *MYH7*, encode for the major molecular motor protein of the sarcomere. Coding mutations in *MYH7* are a common cause of hypertrophic cardiomyopathy (HCM) and mutations in *MYH6* have been linked to various cardiomyopathy phenotypes (1, 65). During development of the human left ventricle, *MYH6* expression switches to *MYH7*. This process is recapitulated in iCMs. In order to study the regulatory regions responsible for *MYH7/6* expression, we needed to determine baseline expression levels. We also needed normalization factors for *MYH6/7* expression that can reduce technical variability. To meet these needs, we downloaded a publicly available RNA-seq dataset of iCM differentiation (171). This data demonstrated clusters of genes that change expression in similar patterns over time. We selected potential normalization genes from genes that were within the same clusters as *MYH6/7*. To test our predictions, we differentiated an IPSC line 30 times and measured the variability of *MYH6/7* expression when normalized by each gene and identified a robust normalization strategy. Our method was useful for *MYH6/7*, but can be applied other genes adequately expressed in iCMs.

Methods

RNA-Seq Download and Count Normalization

We used the SRA toolkit to download the raw fastq files of RNA-seq data representing iCM differentiation (GSE81585) (171, 172). Raw RNA-seq reads were trimmed with trimmomatic (v0.36) and aligned to the human genome (hg19) using STAR with default settings (118).Uniquely aligned reads were assigned to genes using htseq-count using the Ensembl GTF file version 87 (Downloaded May 2016) as annotations (124). Raw count matrices were inputted into EdgeR for normalization (125). Genes were required to have at least 1 count per million in at least 3 different samples. We used EdgeR to generate an MDS plot using the “common” option to compare the same genes across samples.

Gene Clustering and Gene Ontology Analysis

The 500 genes with the most variability across samples was determined by calculating the variance of each gene and keeping genes with the 500 largest values. R’s hclust function was used with default settings to cluster samples and genes. Heatmaps were generated using ggplot’s heatmap.2 function with row normalization. GeneID’s from each cluster were inputted into the PANTHER gene ontology online tool (123).

Linear Regressions and Pearson Correlation Matrix Calculations

For RNA-seq data, we averaged the three replicates to determine an expression value for each day of differentiation. We inputted the log transformed averages into PRISM and ran a linear regression analysis to determine an R^2 value. To calculate Pearson correlation coefficients, we used the corr.plot function in R on a matrix of average expression values across differentiation.

IPSC-CM Differentiation, qPCR Data Generation and qPCR Analysis

IPSCs were differentiated into cardiomyocytes (iCMs) using Wnt modulation as previously described(18). Differentiation was conducted in CDM3 (RPMI 1640 with L-glutamine, 213 μg/mL L-asorbic acid 2-phosphate, 500μg/mL recombinant human albumin) (18). Cells were grown to ~95% confluency and treated with 6μM CHIR99021 for 24 hours and allowed to recover for 24 hours. Cells were treated with 2μM Wnt-C59 for 48 hours and then media was changed with CDM3 every two days until beating cardiomyocytes were obtained (~day 6-10). In order to prevent cell detachment, beating cardiomyocytes re-plated on to new plates using TrypLE (Thermo Fisher). 1 million IPSC-derived cardiomyocytes were plated on a well of 12-well plate. At ~day 20, cells were washed with PBS and 400μl of TRIzol (Thermo Fisher) was added directly to the well. Cells were collected into an Eppendorf tube using a cell scraper. Trizol was kept at -80 ^{o}C until further processing. Six hundred μl of additional TRIzol was added to the cells and the entire sample was added to a tube containing 250μl of silica-zirconium beads. Tubes were placed in a bead beater homogenizer (BioSpec) for 1 minute and immediately cooled on ice. Samples were incubated at room temperature for 5min and then centrifuged at 12,000g for 5min to remove unhomogenized cell aggregates. Supernatant was transferred to a new tube and 200μl of chloroform was added. After vigorous shaking for 30 seconds followed by 10 min incubation with periodic shaking, samples were centrifuged at 12,000g for 15 min. The upper aqueous layer was added to an equal volume of fresh 70% ethanol and used an input to the Aurum Total RNA Mini Kit (Biorad). RNA was processed according to manufacturer's instructions including on-column DNase digestion. RNA was eluted twice with 30μl of warmed water and the concentration was measured using a nanodrop spectrophotometer.

The qScript cDNA SuperMix (Quantabio) was used to generate a 100ng cDNA library. A 1:10 dilution was used as a template in a 3-step SYBR-green qPCR region with a 57^{o}C annealing temperature. We used a panel of primers targeting cardiomyocyte genes (*MYH6*, *MYH7*,*TNNT2*, *MYBPC3*, *TNNI3*, *SLC8A1*, *MYOZ2* and *GAPDH*) that passed optimization studies confirming primer specificity and efficiency. The delta-delta Cq method was used with various second normalizations to determine the variation of *MYH6/7* expression compared to the average of all differentiations. Primers used are shown in **Table 3.2.**

Results

RNA-Seq of IPSC-CM Differentiation Identifies Clusters of Gene Expression Patterns

We downloaded the raw RNA-seq data from a published dataset of iCM differentiation (171). This study differentiated IPSCs into cardiomyocytes with the commonly used Wnt modulation protocol three separate times. They collected RNA every day of differentiation until day 9 and then at three additional time points at day 14, day 30, and day 90 (**Figure 4.1A**). We analyzed the raw data using standard methods and assayed the library size-normalized gene expression values. An MDS plot demonstrated a bell-shaped curve that separated samples according to differentiation day along PC2 (**Figure 4.1B**). In order to understand what genes were contributing to this clustering pattern, we focused on the 500 genes with the most variability across all days. Hierarchical clustering separated samples based on differentiation state, matching our MDS plot (**Figure 4.2A**). The largest transcriptional difference occurred at day 6. Same-day replicates generally clustered together, but not always, underscoring the variability of the differentiation process. We repeated this clustering process on the expression values of the 500 most variable genes to identify shared temporal expression patterns. Genes clustered into 6 different groups (**Figure 4.2B**). Gene ontology (GO) analysis identified both shared and unique gene functions (**Figure 4.2C**). Cluster 1 contained 156 genes and represented genes that increased gradually with differentiation and were mainly associated with muscle development GO terms. Cluster 2 genes also were induced with differentiation and showed highest expression in the long term cultures (day 30 and day 90). Cluster 2 GO terms included muscle development, but also more specific striated muscle development terms. Cluster 3 contained 108 genes that were expressed early in differentiation and decreased over time. Cluster 3 GO terms included gastrulation and germ layer differentiation. Cluster 4 genes were induced with differentiation, but were lost in long term cultures. These genes were enriched for extracellular structure organization and generalized tissue morphogenesis terms. Cluster 5 genes were similar to cluster 1, but showed a more variable induction time. Cluster 5 GO terms were more specific for heart development.

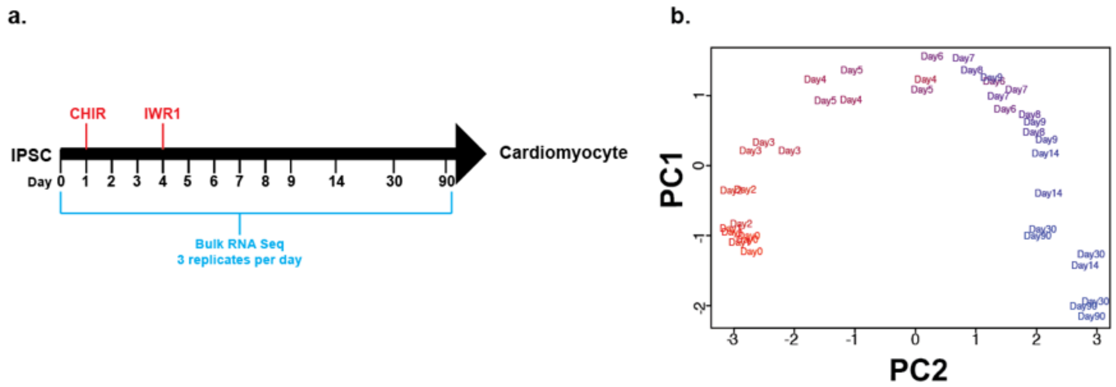
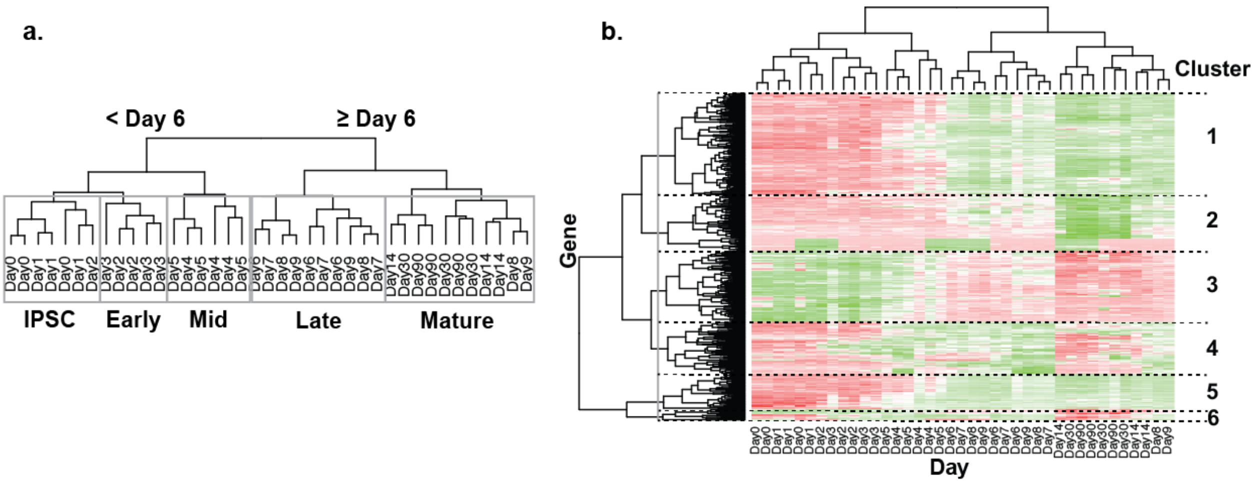


Figure 4.1. RNA-seq captures transcriptional changes during IPSC-CM differentiation. A. Schematic of IPSC differentiation protocol. Bulk RNA-seq data was obtained at each day from three separate differentiations. **B.** MDS plot of samples for 500 genes with the largest standard deviations across samples color coded by day.

Selected Cardiomyocyte Genes Normalize iCM *MYH6* and *MYH7* Expression Values

We set out to use the clustering analysis above to identify genes that can serve as normalization factors for *MYH6* and *MYH7* expression. *MYH6* and *MYH7* expression in iCMs follows a pattern similar to human development, where *MYH6* expression switches to *MYH7* expression over time (**Figure 4.3A&B**) (152). We selected *TNNT2*, *MYBPC3*, *SLC8A1*, and *MYOZ2* as they were members of the same expression pattern clusters as *MYH6* and *MYH7*. We also included *TNNI3* because this gene has recently been used for normalization (173). We also included *GAPDH* as a control as it is a commonly used qPCR normalization factor.

We used the iCM RNA-seq data to assay the relationship between *MYH6/7* expression and our selected cardiomyocyte reference genes expression. We hypothesized that genes with expression values that correlate across differentiation, will be good choices for normalization across differentiations. If well correlated, technical variation in the expression of in the normalization gene should correct for technical variation in *MYH6/7* expression levels. We assayed the linear relationships between *MYH6/7* expression changes and reference gene expression changes by calculating a “goodness of fit” R^2 value for each comparison. *MYH7* expression was most correlated with *MYBPC3* and *MYOZ2* expression, indicating that



c.

Cluster	# of Genes	Top 5 Significant GO Terms
1	156	muscle structure development muscle cell differentiation muscle system process muscle tissue development striated muscle cell differentiation
2	85	muscle tissue development striated muscle tissue development striated muscle cell development striated muscle cell differentiation muscle structure development
3	108	Gastrulation endodermal cell fate specification cell fate specification endoderm formation regionalization
4	80	extracellular structure organization tissue morphogenesis regulation of developmental process circulatory system development tube morphogenesis
5	53	heart development cardiac muscle tissue development circulatory system development striated muscle tissue development muscle tissue development
6	18	NA

Figure 4.2. Unsupervised clustering of RNA-seq data identifies modules of gene expression Changes. A. Dendrogram of hierarchical clustering results on the normalized counts values for the 500 most variable genes. **B.** Row-normalized heatmap of the expression values of the 500 most variable genes. Dendrogram of hierarchical clustering of genes was cut at the gray line to generate 6 clusters. **C.** Top 5 significantly enriched gene ontology terms for the clusters identified in B.

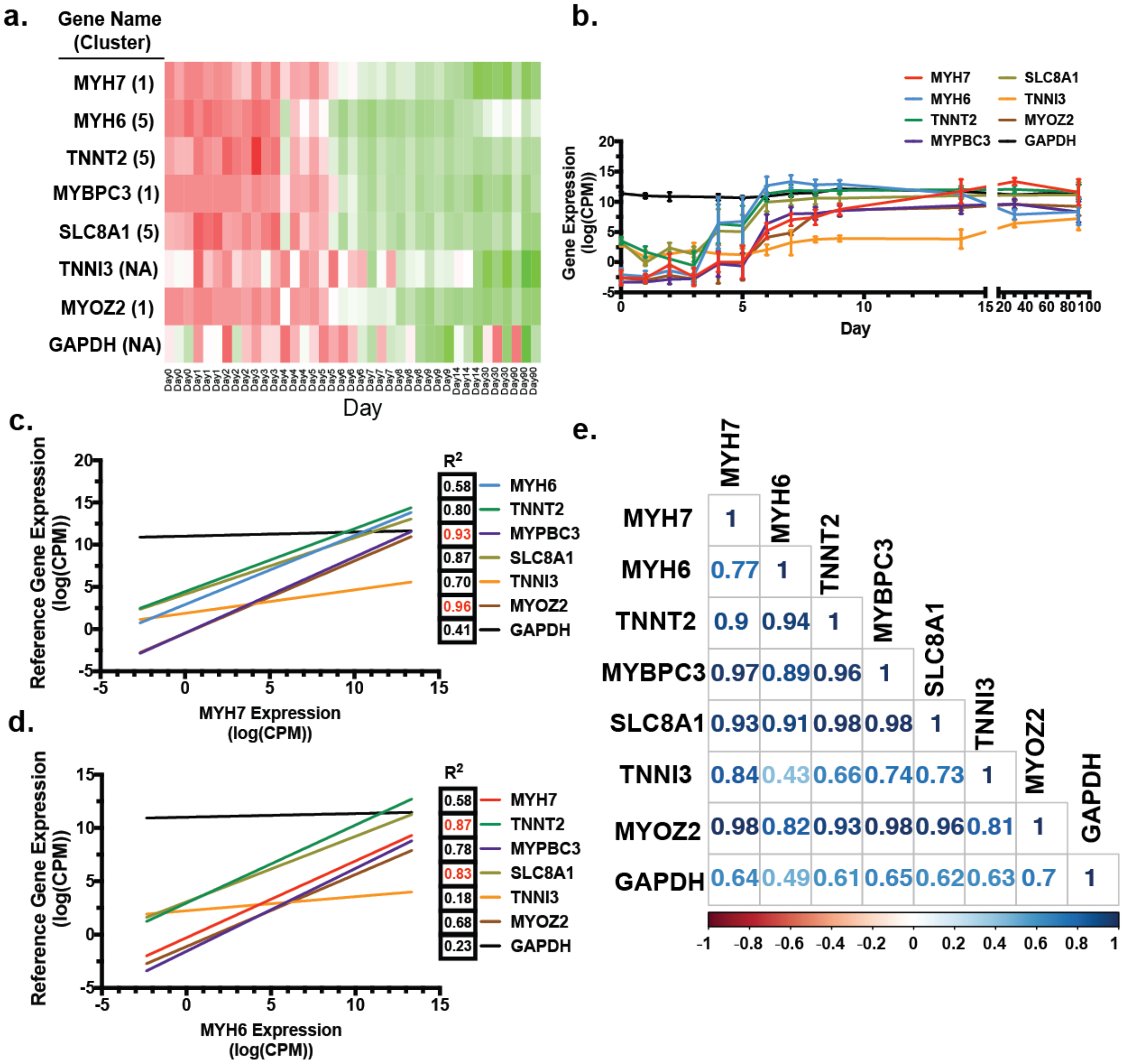
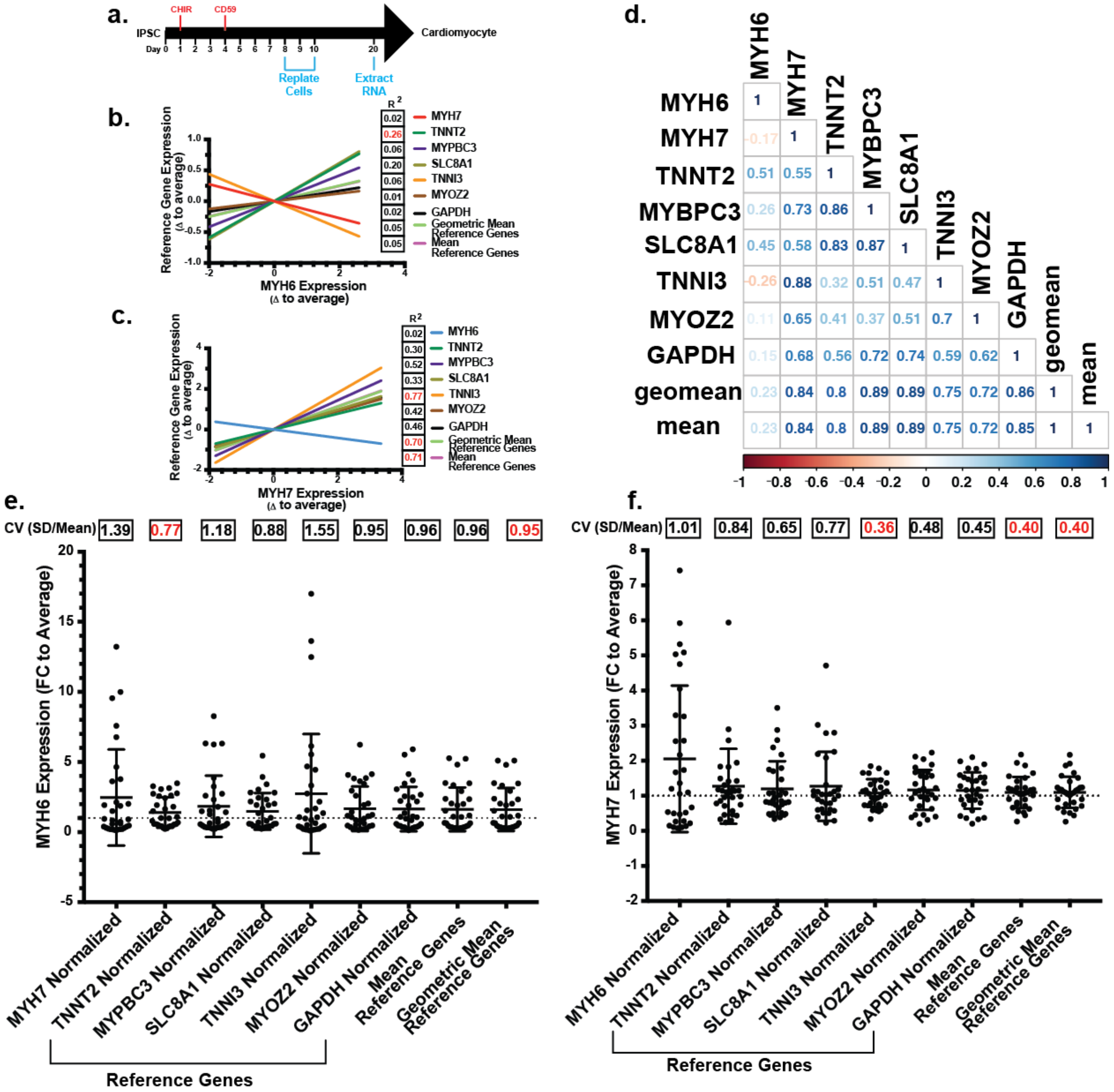


Figure 4.3. RNA-seq data demonstrates relationships between selected cardiomyocyte genes. A. Row-normalized heatmap of selected cardiomyocyte gene RNA-seq expression levels during differentiation. **B.** Time course of gene RNA-seq expression levels during differentiation. **C.** Linear relationships between *MYH7* expression and other cardiomyocyte genes throughout differentiation. **D.** Linear relationships between *MYH6* expression and other cardiomyocyte genes throughout differentiation. **E.** Matrix of Pearson correlation coefficients between selected cardiomyocyte genes.



	MYH6	MYH7	TNNT2	MYBPC3	SLC8A1	TNNI3	MYOZ2	GAPDH	geomean	mean
MYH6	1									
MYH7	-0.17	1								
TNNT2	0.51	0.55	1							
MYBPC3	0.26	0.73	0.86	1						
SLC8A1	0.45	0.58	0.83	0.87	1					
TNNI3	-0.26	0.88	0.32	0.51	0.47	1				
MYOZ2	0.11	0.65	0.41	0.37	0.51	0.7	1			
GAPDH	0.15	0.68	0.56	0.72	0.74	0.59	0.62	1		
geomean	0.23	0.84	0.8	0.89	0.89	0.75	0.72	0.86	1	
mean	0.23	0.84	0.8	0.89	0.89	0.75	0.72	0.85	1	1

Figure 4.4. Normalization genes reduce variation of *MYH6/7* qPCR expression data. A. Schematic of IPSC-CM differentiation protocol used in this study. All differentiations were cultured until day 20. **B.** Linear relationships between *MYH6* expression and cardiomyocyte genes at day 20 of differentiation. **C.** Linear relationships between *MYH7* expression and cardiomyocyte genes at day 20 of differentiation. **D.** Matrix of Pearson correlation coefficients of gene expression values at day 20 of differentiation. **E**. *MYH6* expression values normalized to expression value of each cardiomyocyte gene and the geometric mean and mean of reference genes. The coefficient of variation is listed at the top. **F.***MYH7* expression values normalized to expression value of each cardiomyocyte gene and the geometric mean and mean of reference genes. The coefficient of variation is listed at the top.

changes in *MYH7* are correlated with changes in *MYBPC3* and *MYOZ2* expression during differentiation (**Figure 4.3C**). *MYH6* expression was most correlated with *TNNT2* and *SLC8A1* expression **(Figure 4.3D**). In general, *MYH7* linear relationships were stronger than *MYH6* relationships. We also calculated Pearson correlation coefficients for pairwise comparisons between all genes. As expected, the Pearson correlation coefficients for *MYH6/7* matched the results we obtained from our linear regressions (**Figure 4.3E**). Overall, these findings indicate that *MYBPC3* and *MYOZ2* can normalize *MYH7* expression levels and *TNNT2* and *SLC8A1* can normalize *MYH6* expression levels.

We set out to test these predictions by generating qPCR data from iCMs. We differentiated a single IPSC line (Coriell, GM03348) 30 times using a standard protocol (**Figure 4.4A**). All differentiations were matured to day 20 and processed similarity to generate qPCR data for *MYH6*, *MYH7*, *TNNT2*, *MYBPC3*, *SLC8A1*, *TNNI3*, *MYOZ2*, and *GAPDH*. For each differentiation, we calculated each gene's difference to the mean value over all differentiations. Using a linear model approach, we assayed the relationship between *MYH6/7*'s variation from the mean and reference gene variation from the mean. For example, if one differentiation had lower *MYH6/7* expression compared to the mean, we tested if the reference genes were also reduced. *MYH6* expression was most related to *TNNT2* and *SLC8A1* expression, matching the RNA-seq data predictions (**Figure 4.4B**). *MYH7* expression was most related to *TNNI3* expression and was also related to *MYBPC3* expression (**Figure 4.4C)**. Interestingly, when we averaged (both mean and geometric mean) the difference values across all the reference genes (*TNNT2*, *MYBPC3*, *SLC8A1*, *TNNI3*, *MYOZ2*, *GAPDH*), we saw a strong relationship with *MYH7* expression, but not *MYH6*. Overall qPCR R^2 values were lower than RNA-seq R^2 values likely due to the technical variability of qPCR and measurements across many more iCM differentiations. We also calculated Pearson correlation coefficients for the mean difference values and our results matched those from linear regression findings.

Finally, we calculated *MYH6* and *MYH7* expression normalized for each reference gene. As all differentiations were identical, we expect the expression of *MYH6/7* to be consistent across samples. For each normalization, we calculated the coefficient of variation (CV) by dividing the standard deviation by the mean value. For *MYH6*, *TNNT2* and geometric mean showed the lowest CV (**Figure 4.4E**). For

MYH7, *TNNI3* and the mean values showed the lowest CV (**Figure 4.4F**). As hypothesized, the normalization genes with the best linear fits and Person coefficients resulted in the lowest CV values.

Discussion

Gene Expression Changes During iCM Differentiation Cluster into Common Patterns

RNA-seq analysis of iCM differentiation identified multiple clusters of genes that share a similar expression pattern. Some clusters represent the induction of cardiomyocyte/muscle genes, while others represent downregulation of pluripotency genes. Induction clusters are enriched for cardiomyocyte specific genes and cardiomyopathy-associated genes. Additional genes within this cluster may represent new cardiomyopathy-associated genes and analysis of variation within these genes will likely yield interesting findings.

Another cluster of genes was identified that turns on during differentiation but is lost in longer term cultures. The most enriched term for these genes was extracellular structure organization. These changes could represent the genes that help cardiomyocytes remodel the extracellular matrix, which has been implicated in cardiomyopathy and heart failure (174). These genes could also explain why long-term cultures of iCMs tend to delaminate from surfaces. Further analyses of these genes may identify modifiers of cardiomyopathy and lead to technical advances in iCM maturation.

Gene-Specific iCM Transcriptional Maturity Measure

We determined normalization factors for *MYH6/7* expression by using RNA-seq data to find genes that are co-expressed. The RNA-seq data was able to predict normalizers for *MYH6/7*, but the variation in the *MYH6* qPCR data was still high. Additional RNA-seq datasets from a variety of IPCS over many more differentiations would be revealing and likely identify even better normalizers. qPCR is also more prone to technical variation than RNA-seq and with the falling costs of sequencing, targeted RNA-sequencing may be more commonly used.

The results with *MYH6/7* are supportive of our approach. iCMs express high levels *MYH6/7* and their relative expression changes significantly over time. Other genes that are expressed at lower levels or in different patterns may not have strong normalizers. However, many cardiomyopathy genes are cardiac-specific and expressed at high levels in iCMs. We propose that our system can readily be applied to these genes to determine gene-specific normalization factors.

Chapter 5.

Summary and Conclusions

Summary and Conclusions

The phenotypic variability associated with cardiomyopathy mutations has intrigued scientists for many years. The search for genetic modifiers began in the coding genome with some success, but a large proportion of variance was left unexplained. This thesis expanded the search of genetic modifiers into the noncoding genome. We applied both genome wide techniques and targeted analyses around the cardiomyopathy genes *MYH7* and *LMNA*. Our analysis identified promoters and enhancers that change activity in the heart failure state, which are likely to contain modifying sequence variants. The targeted analysis presented here identified two enhancer regions responsible for *MYH7* expression in iCMs. We identified sequence variants within these enhancers that altered enhancer function. Importantly, a common enhancer variant upstream of *MYH7* correlated with cardiomyopathy phenotypes over time. These findings indicate that genome wide and targeted analysis can identify functional noncoding modifiers of cardiomyopathy phenotypes.

Healthy and Failed Human LV Promoter/Enhancer Map

In **Chapter 2**, I evaluated CAGE-seq data from healthy and failing human left ventricles (LV). CAGE-seq allowed the detection and comparison of promoter and enhancer RNA expression in healthy and failed human LVs. We identified heart failure-specific promoters and generated an evidence-based promoter map of the human left ventricle. This map will be a useful source of information for LV biology and aid in future studies of LV gene expression. This map can improve variant annotation tools by prioritizing transcripts expressed in the human LV. Future chromatin conformation experiments can also utilize this map to ensure the assessment of LV-relevant promoter interactions.

We also found that many eRNA producing enhancer regions are located with the first intron of genes and also identified failure-induced enhancers. The first intron of genes has been implicated in enhancer function before and should be considered when assaying enhancer regions around genes (130). Enhancers that change expression during heart failure are of particular interest. The robust gene expression changes seen in the failure state are presumably driven by changes in enhancer function. Sequence variants that modify the function of these enhancers are well positioned to modify the gene expression changes in heart failure and therefore affect organ function and clinical presentations. Future analyses of sequence variants within these regions, with a particular focus on common eQTL variants, are likely to yield novel noncoding modifiers.

Regulation of the MYH6-MYH7 Gene Cluster

The regulation of the *MYH6*-*MYH7* gene cluster has been the target of study for decades. *MYH6* and *MYH7* encode a vital components of the sarcomere in the heart and, the two encoded proteins have differences in ATP-utilization and force generation, which are specified by differences in their protein coding regions. During development in humans, the left ventricle downregulates *MYH6* and upregulates *MYH7*. Interestingly, a similar downregulation of *MYH6* occurs in the failing LV. In **Chapter 3**, I identified two genomic regions important for *MYH7* expression in iCMs. One region ~7kb upstream of *MYH7* contains a super enhancer that is required to switch from *MYH6* to *MYH7* during iCM differentiation, since deletion of this region affected the expression of both genes in coordinated manner. Another region, ~2kb upstream of *MYH7*, emerged as more specific for *MYH7* expression as homozygous deletion caused ~100-fold reduction in transcript levels, with comparatively less effect on *MYH6* expression. As part of these studies, I developed a transcriptional assay of multiple cardiomyocyte genes that when used together helps to reduce the variable purity/maturity of iCM populations (**Chapter 4**). I propose a specific model where the *MYH6* and *MYH7* promoter regions are part of a complex with the upstream super enhancer. In this model, during development changes in transcription factor expression increase *MYH7* promoter-enhancer interactions and concomitantly decrease *MYH6* promoter-enhancer interactions. It is likely that the separate *MYH7*-specific enhancer is also part of this complex, but more important for *MYH7* induction and less important for *MYH6* reduction. I hypothesize a common super enhancer model may also be present at other gene clusters like *TBX3*-*TBX5*, *SCN5A*-*SCN10A*, and *NPPA*-*NPPB*, genes which likely arose from gene duplication events and where each gene evolved to adopt spatiotemporal-specific regulation.

Application to Clinical Care of Cardiomyopathy

The overarching goal of this body of work is to improve the mechanistic understanding of the clinical variability in cardiomyopathy. Current clinical genetic testing focuses on assaying coding regions of known cardiomyopathy-causing genes to find a causative mutation. If the regulatory regions of known cardiomyopathy genes are evaluated, then targeted noncoding sequencing could be integrated to inform clinical decision making. With large datasets of noncoding variants and phenotypic data, machine learning algorithms may be able to provide estimations of expected phenotypic courses. There are likely gene-specific and general noncoding modifiers. For example, the *MYH7* noncoding modifier rs875908 variant is likely a general noncoding modifier as there were phenotypic correlations across individuals without pathogenic coding *MYH7* mutations. Knowledge of an individual's precise genetic risk can allow a clinician to tailor a clinical care strategy that maximizes effectiveness.

Future Directions

There is already a massive amount of epigenomic data readily available relevant to cardiomyopathy phenotypes (outlined in **Chapter 1.**) Future studies around other cardiomyopathy genes like *TTN*, *DES* and *ACTN1* can utilize this data to predict potential regulatory regions. Many cardiac model systems lack the transcriptional maturity to accurately validate many of these predictions. However, the field is rapidly advancing and new biochemical and physical techniques are constantly being developed to overcome this challenge. The regulatory map of *MYH7* enhancer regions created in this work is the first step to developing a map of enhancer regions associated with all cardiomyopathy genes and delivering on the potential of clinical genetic testing.

References

1.McNally EM, Mestroni L. Dilated Cardiomyopathy: Genetic Determinants and Mechanisms. Circ Res. 2017;121(7):731-48. Epub 2017/09/16. doi: 10.1161/CIRCRESAHA.116.309396. PubMed PMID: 28912180; PMCID: PMC5626020.

2.Makalowski W. The human genome structure and organization. Acta Biochim Pol. 2001;48(3):587-98. Epub 2002/02/09. PubMed PMID: 11833767.

3.Ohno S. So much "junk" DNA in our genome. Brookhaven Symp Biol. 1972;23:366-70. Epub 1972/01/01. PubMed PMID: 5065367.

4.Venters BJ, Pugh BF. How eukaryotic genes are transcribed. Crit Rev Biochem Mol Biol. 2009;44(2-3):117-41. Epub 2009/06/12. doi: 10.1080/10409230902858785. PubMed PMID: 19514890; PMCID: PMC2718758.

5.Orphanides G, Reinberg D. A unified theory of gene expression. Cell. 2002;108(4):439-51. Epub 2002/03/23. doi: 10.1016/s0092-8674(02)00655-4. PubMed PMID: 11909516.

6.Gacita AM, Dellefave-Castillo L, Page PGT, Barefield DY, Waserstrom JA, Puckelwartz MJ, Nobrega MA, McNally EM. Enhancer and promoter usage in the normal and failed human heart. bioRxiv. 2020:2020.03.17.988790. doi: 10.1101/2020.03.17.988790.

7.Rigau M, Juan D, Valencia A, Rico D. Intronic CNVs and gene expression variation in human populations. PLoS Genet. 2019;15(1):e1007902. Epub 2019/01/25. doi: 10.1371/journal.pgen.1007902. PubMed PMID: 30677042; PMCID: PMC6345438.

8.Chen Y, Yao B, Zhu Z, Yi Y, Lin X, Zhang Z, Shen G. A constitutive super-enhancer: homologous region 3 of Bombyx mori nucleopolyhedrovirus. Biochem Biophys Res Commun. 2004;318(4):1039-44. Epub 2004/05/19. doi: 10.1016/j.bbrc.2004.04.136. PubMed PMID: 15147978.

9.Szabo Q, Bantignies F, Cavalli G. Principles of genome folding into topologically associating domains. Sci Adv. 2019;5(4):eaaw1668. Epub 2019/04/17. doi: 10.1126/sciadv.aaw1668. PubMed PMID: 30989119; PMCID: PMC6457944.

10.Banerji J, Rusconi S, Schaffner W. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. Cell. 1981;27(2 Pt 1):299-308. Epub 1981/12/01. doi: 10.1016/0092-8674(81)90413-x. PubMed PMID: 6277502.

11.Banerji J, Olson L, Schaffner W. A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. Cell. 1983;33(3):729-40. Epub 1983/07/01. doi: 10.1016/0092-8674(83)90015-6. PubMed PMID: 6409418.

12.Hamlin RL, Altschuld RA. Extrapolation from mouse to man. Circ Cardiovasc Imaging. 2011;4(1):2-4. Epub 2011/01/20. doi: 10.1161/CIRCIMAGING.110.961979. PubMed PMID: 21245362.

13.England J, Loughna S. Heavy and light roles: myosin in the morphogenesis of the heart. Cell Mol Life Sci. 2013;70(7):1221-39. Epub 2012/09/08. doi: 10.1007/s00018-012-1131-1. PubMed PMID: 22955375; PMCID: PMC3602621.

14.Claycomb WC, Lanson NA, Jr., Stallworth BS, Egeland DB, Delcarpio JB, Bahinski A, Izzo NJ, Jr. HL-1 cells: a cardiac muscle cell line that contracts and retains phenotypic characteristics of the adult cardiomyocyte. Proc Natl Acad Sci U S A. 1998;95(6):2979-84. Epub 1998/04/18. doi: 10.1073/pnas.95.6.2979. PubMed PMID: 9501201; PMCID: PMC19680.

15.Spurrell CH, Barozzi I, Mannion BJ, Blow MJ, Fukuda-Yuzawa Y, Afzal SY, Akiyama JA, Afzal V, Tran S, Plajzer-Frick I, Novak CS, Kato M, Lee E, Garvin TH, Pham QT, Harrington AN, Lisgo S, Bristow J, Cappola TP, Morley MP, Margulies KB, Pennacchio LA, Dickel DE, Visel A. Genome-Wide Fetalization of Enhancer Architecture in Heart Disease. bioRxiv. 2019:591362. doi: 10.1101/591362.

16.Kim Y, Rim YA, Yi H, Park N, Park SH, Ju JH. The Generation of Human Induced Pluripotent Stem Cells from Blood Cells: An Efficient Protocol Using Serial Plating of Reprogrammed Cells by Centrifugation. Stem Cells Int. 2016;2016:1329459. Epub 2016/09/01. doi: 10.1155/2016/1329459. PubMed PMID: 27579041; PMCID: PMC4989082.

17. Lowry WE, Richter L, Yachechko R, Pyle AD, Tchieu J, Sridharan R, Clark AT, Plath K. Generation of human induced pluripotent stem cells from dermal fibroblasts. Proc Natl Acad Sci U S A. 2008;105(8):2883-8. Epub 2008/02/22. doi: 10.1073/pnas.0711983105. PubMed PMID: 18287077; PMCID: PMC2268554.

18. Burridge PW, Matsa E, Shukla P, Lin ZC, Churko JM, Ebert AD, Lan F, Diecke S, Huber B, Mordwinkin NM, Plews JR, Abilez OJ, Cui B, Gold JD, Wu JC. Chemically defined generation of human cardiomyocytes. Nat Methods. 2014;11(8):855-60. Epub 2014/06/16. doi: 10.1038/nmeth.2999. PubMed PMID: 24930130; PMCID: PMC4169698.

19. Schaaf S, Eder A, Vollert I, Stohr A, Hansen A, Eschenhagen T. Generation of strip-format fibrin-based engineered heart tissue (EHT). Methods Mol Biol. 2014;1181:121-9. Epub 2014/07/30. doi: 10.1007/978-1-4939-1047-2_11. PubMed PMID: 25070332.

20.van den Berg CW, Okawa S, Chuva de Sousa Lopes SM, van Iperen L, Passier R, Braam SR, Tertoolen LG, del Sol A, Davis RP, Mummery CL. Transcriptome of human foetal heart compared with cardiomyocytes from pluripotent stem cells. Development. 2015;142(18):3231-8. Epub 2015/07/26. doi: 10.1242/dev.123810. PubMed PMID: 26209647.

21. Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, Amin V, Whitaker JW, Schultz MD, Ward LD, Sarkar A, Quon G, Sandstrom RS, Eaton ML, Wu YC, Pfenning AR, Wang X, Claussnitzer M, Liu Y, Coarfa C, et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015;518(7539):317-30. Epub 2015/02/20. doi: 10.1038/nature14248. PubMed PMID: 25693563; PMCID: PMC4530010.

22. Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, Hilton JA, Jain K, Baymuradov UK, Narayanan AK, Onate KC, Graham K, Miyasato SR, Dreszer TR, Strattan JS, Jolanki O, Tanaka FY, Cherry JM. The Encyclopedia of DNA elements (ENCODE): data portal update. Nucleic Acids Res. 2018;46(D1):D794-D801. Epub 2017/11/11. doi: 10.1093/nar/gkx1081. PubMed PMID: 29126249; PMCID: PMC5753278.

23. Shahid Z, Simpson B, Singh G. Genetics, Histone Code. StatPearls. Treasure Island (FL)2020.

24. Rothbart SB, Strahl BD. Interpreting the language of histone and DNA modifications. Biochimica et biophysica acta. 2014;1839(8):627-43. Epub 2014/03/19. doi: 10.1016/j.bbagrm.2014.03.001. PubMed PMID: 24631868; PMCID: PMC4099259.

25. May D, Blow MJ, Kaplan T, McCulley DJ, Jensen BC, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Afzal V, Simpson PC, Rubin EM, Black BL, Bristow J, Pennacchio LA, Visel A. Large-scale discovery of enhancers from human heart tissue. Nat Genet. 2011;44(1):89-93. Epub 2011/12/06. doi: 10.1038/ng.1006. PubMed PMID: 22138689; PMCID: PMC3246570.

26. Ang YS, Rivas RN, Ribeiro AJS, Srivas R, Rivera J, Stone NR, Pratt K, Mohamed TMA, Fu JD, Spencer CI, Tippens ND, Li M, Narasimha A, Radzinsky E, Moon-Grady AJ, Yu H, Pruitt BL, Snyder MP, Srivastava D. Disease Model of GATA4 Mutation Reveals Transcription Factor Cooperativity in Human Cardiogenesis. Cell. 2016;167(7):1734-49 e22. Epub 2016/12/17. doi: 10.1016/j.cell.2016.11.033. PubMed PMID: 27984724; PMCID: PMC5180611.

27. Benaglio P, D'Antonio-Chronowska A, Ma W, Yang F, Young Greenwald WW, Donovan MKR, DeBoever C, Li H, Drees F, Singhal S, Matsui H, van Setten J, Sotoodehnia N, Gaulton KJ, Smith EN, D'Antonio M, Rosenfeld MG, Frazer KA. Allele-specific NKX2-5 binding underlies multiple genetic associations with human electrocardiographic traits. Nat Genet. 2019;51(10):1506-17. Epub 2019/10/02. doi: 10.1038/s41588-019-0499-3. PubMed PMID: 31570892; PMCID: PMC6858543.

28. He A, Kong SW, Ma Q, Pu WT. Co-occupancy by multiple cardiac transcription factors identifies transcriptional enhancers active in heart. Proc Natl Acad Sci U S A. 2011;108(14):5632-7. Epub 2011/03/19. doi: 10.1073/pnas.1016959108. PubMed PMID: 21415370; PMCID: PMC3078411.

29. Akerberg BN, Gu F, VanDusen NJ, Zhang X, Dong R, Li K, Zhang B, Zhou B, Sethi I, Ma Q, Wasson L, Wen T, Liu J, Dong K, Conlon FL, Zhou J, Yuan GC, Zhou P, Pu WT. A reference map of murine cardiac transcription factor chromatin occupancy identifies dynamic and conserved enhancers. Nat Commun. 2019;10(1):4907. Epub 2019/10/30. doi: 10.1038/s41467-019-12812-3. PubMed PMID: 31659164; PMCID: PMC6817842.

30.van den Boogaard M, Wong LY, Tessadori F, Bakker ML, Dreizehnter LK, Wakker V, Bezzina CR, t Hoen PA, Bakkers J, Barnett P, Christoffels VM. Genetic variation in T-box binding element functionally affects SCN5A/SCN10A enhancer. J Clin Invest. 2012;122(7):2519-30. Epub 2012/06/19. doi: 10.1172/JCI62613. PubMed PMID: 22706305; PMCID: PMC3386824.

31. Liu Q, Jiang C, Xu J, Zhao MT, Van Bortle K, Cheng X, Wang G, Chang HY, Wu JC, Snyder MP. Genome-Wide Temporal Profiling of Transcriptome and Open Chromatin of Early Cardiomyocyte Differentiation Derived From hiPSCs and hESCs. Circ Res. 2017;121(4):376-91. Epub 2017/07/01. doi: 10.1161/CIRCRESAHA.116.310456. PubMed PMID: 28663367; PMCID: PMC5576565.

32. Pottinger TD, Pesce LL, Gacita A, Montefiori L, Hodge N, Kearns S, Salamone IM, Pacheco JA, Rasmussen-Torvik LJ, Smith ME, Chisholm R, Nobrega MA, McNally EM, Puckelwartz MJ. Trajectory analysis of cardiovascular phenotypes from biobank data uncovers novel genetic associations. bioRxiv. 2020:2020.05.10.087130. doi: 10.1101/2020.05.10.087130.

33. Arvanitis M, Tampakakis E, Zhang Y, Wang W, Auton A, andMe Research T, Dutta D, Glavaris S, Keramati A, Chatterjee N, Chi NC, Ren B, Post WS, Battle A. Genome-wide association and multi-omic analyses reveal ACTN2 as a gene linked to heart failure. Nat Commun. 2020;11(1):1122. Epub 2020/03/01. doi: 10.1038/s41467-020-14843-7. PubMed PMID: 32111823; PMCID: PMC7048760.

34. Noguchi S, Arakawa T, Fukuda S, Furuno M, Hasegawa A, Hori F, Ishikawa-Kato S, Kaida K, Kaiho A, Kanamori-Katayama M, Kawashima T, Kojima M, Kubosaki A, Manabe RI, Murata M, Nagao-Sato S, Nakazato K, Ninomiya N, Nishiyori-Sueki H, Noma S, Saijyo E, Saka A, Sakai M, Simon C, Suzuki N, et al. FANTOM5 CAGE profiles of human and mouse samples. Sci Data. 2017;4:170112. Epub 2017/08/30. doi: 10.1038/sdata.2017.112. PubMed PMID: 28850106; PMCID: PMC5574368.

35. Wei C, Qiu J, Zhou Y, Xue Y, Hu J, Ouyang K, Banerjee I, Zhang C, Chen B, Li H, Chen J, Song LS, Fu XD. Repression of the Central Splicing Regulator RBFox2 Is Functionally Linked to Pressure Overload-Induced Heart Failure. Cell Rep. 2015;10(9):1521-33. Epub 2015/03/11. doi: 10.1016/j.celrep.2015.02.013. PubMed PMID: 25753418; PMCID: PMC4559494.

36. Leung D, Jung I, Rajagopal N, Schmitt A, Selvaraj S, Lee AY, Yen CA, Lin S, Lin Y, Qiu Y, Xie W, Yue F, Hariharan M, Ray P, Kuan S, Edsall L, Yang H, Chi NC, Zhang MQ, Ecker JR, Ren B. Integrative analysis of haplotype-resolved epigenomes across human tissues. Nature. 2015;518(7539):350-4. Epub 2015/02/20. doi: 10.1038/nature14217. PubMed PMID: 25693566; PMCID: PMC4449149.

37. Montefiori LE, Sobreira DR, Sakabe NJ, Aneas I, Joslin AC, Hansen GT, Bozek G, Moskowitz IP, McNally EM, Nobrega MA. A promoter interaction map for cardiovascular disease genetics. Elife. 2018;7. Epub 2018/07/11. doi: 10.7554/eLife.35788. PubMed PMID: 29988018; PMCID: PMC6053306.

38. Jung I, Schmitt A, Diao Y, Lee AJ, Liu T, Yang D, Tan C, Eom J, Chan M, Chee S, Chiang Z, Kim C, Masliah E, Barr CL, Li B, Kuan S, Kim D, Ren B. A compendium of promoter-centered long-range chromatin interactions in the human genome. Nat Genet. 2019;51(10):1442-9. Epub 2019/09/11. doi: 10.1038/s41588-019-0494-8. PubMed PMID: 31501517; PMCID: PMC6778519.

39. Choy MK, Javierre BM, Williams SG, Baross SL, Liu Y, Wingett SW, Akbarov A, Wallace C, Freire-Pritchett P, Rugg-Gunn PJ, Spivakov M, Fraser P, Keavney BD. Promoter interactome of human embryonic stem cell-derived cardiomyocytes connects GWAS regions to cardiac gene networks. Nat Commun. 2018;9(1):2526. Epub 2018/06/30. doi: 10.1038/s41467-018-04931-0. PubMed PMID: 29955040; PMCID: PMC6023870.

40. Kim S, Yu NK, Kaang BK. CTCF as a multifunctional protein in genome regulation and gene expression. Exp Mol Med. 2015;47:e166. Epub 2015/06/06. doi: 10.1038/emm.2015.33. PubMed PMID: 26045254; PMCID: PMC4491725.

41. Grunert M, Dorn C, Rickert-Sperling S. Cardiac transcription factors and regulatory networks. Congenital Heart Diseases: The Broken Heart: Springer; 2016. p. 139-52.

42. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010;38(4):576-89. Epub 2010/06/02. doi: 10.1016/j.molcel.2010.05.004. PubMed PMID: 20513432; PMCID: PMC2898526.

43. Song L, Crawford GE. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. Cold Spring Harb Protoc. 2010;2010(2):pdb prot5384. Epub 2010/02/13. doi: 10.1101/pdb.prot5384. PubMed PMID: 20150147; PMCID: PMC3627383.

44. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr Protoc Mol Biol. 2015;109:21 9 1- 9 9. Epub 2015/01/07. doi: 10.1002/0471142727.mb2129s109. PubMed PMID: 25559105; PMCID: PMC4374986.

45. Kaikkonen MU, Halonen P, Liu OH, Turunen TA, Pajula J, Moreau P, Selvarajan I, Tuomainen T, Aavik E, Tavi P, Yla-Herttuala S. Genome-Wide Dynamics of Nascent Noncoding RNA Transcription in Porcine Heart After Myocardial Infarction. Circ Cardiovasc Genet. 2017;10(3). Epub 2017/06/15. doi: 10.1161/CIRCGENETICS.117.001702. PubMed PMID: 28611032.

46. Ding M, Liu Y, Liao X, Zhan H, Liu Y, Huang W. Enhancer RNAs (eRNAs): New Insights into Gene Transcription and Disease Treatment. J Cancer. 2018;9(13):2334-40. Epub 2018/07/22. doi: 10.7150/jca.25829. PubMed PMID: 30026829; PMCID: PMC6036709.

47. De Santa F, Barozzi I, Mietton F, Ghisletti S, Polletti S, Tusi BK, Muller H, Ragoussis J, Wei CL, Natoli G. A large fraction of extragenic RNA pol II transcription sites overlap enhancers. PLoS Biol. 2010;8(5):e1000384. Epub 2010/05/21. doi: 10.1371/journal.pbio.1000384. PubMed PMID: 20485488; PMCID: PMC2867938.

48. Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M, Barbara-Haley K, Kuersten S, Markenscoff-Papadimitriou E, Kuhl D, Bito H, Worley PF, Kreiman G, Greenberg ME. Widespread transcription at neuronal activity-regulated enhancers. Nature. 2010;465(7295):182-7. Epub 2010/04/16. doi: 10.1038/nature09033. PubMed PMID: 20393465; PMCID: PMC3020079.

49. Melo CA, Drost J, Wijchers PJ, van de Werken H, de Wit E, Oude Vrielink JA, Elkon R, Melo SA, Leveille N, Kalluri R, de Laat W, Agami R. eRNAs are required for p53-dependent enhancer activity and gene transcription. Mol Cell. 2013;49(3):524-35. Epub 2013/01/01. doi: 10.1016/j.molcel.2012.11.021. PubMed PMID: 23273978.

50. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Chen Y, Zhao X, Schmidl C, Suzuki T, Ntini E, Arner E, Valen E, Li K, Schwarzfischer L, Glatz D, Raithel J, Lilje B, Rapin N, Bagger FO, Jorgensen M, Andersen PR, Bertin N, Rackham O, Burroughs AM, et al. An atlas of active enhancers across human cell types and tissues. Nature. 2014;507(7493):455-61. Epub 2014/03/29. doi: 10.1038/nature12787. PubMed PMID: 24670763; PMCID: PMC5215096.

51. Murata M, Nishiyori-Sueki H, Kojima-Ishiyama M, Carninci P, Hayashizaki Y, Itoh M. Detecting expressed genes using CAGE. Methods Mol Biol. 2014;1164:67-85. Epub 2014/06/15. doi: 10.1007/978-1-4939-0805-9_7. PubMed PMID: 24927836.

52. Bianchi V, Geeven G, Tucker N, Hilvering CRE, Hall AW, Roselli C, Hill MC, Martin JF, Margulies KB, Ellinor PT, de Laat W. Detailed Regulatory Interaction Map of the Human Heart Facilitates Gene Discovery for Cardiovascular Disease. bioRxiv. 2019:705715. doi: 10.1101/705715.

53. Schoenfelder S, Javierre BM, Furlan-Magaril M, Wingett SW, Fraser P. Promoter Capture Hi-C: High-resolution, Genome-wide Profiling of Promoter Interactions. J Vis Exp. 2018(136). Epub 2018/07/17. doi: 10.3791/57320. PubMed PMID: 30010637; PMCID: PMC6102006.

54. Koopmann TT, Adriaens ME, Moerland PD, Marsman RF, Westerveld ML, Lal S, Zhang T, Simmons CQ, Baczko I, dos Remedios C, Bishopric NH, Varro A, George AL, Jr., Lodder EM, Bezzina CR. Genome-wide identification of expression quantitative trait loci (eQTLs) in human heart. PLoS One. 2014;9(5):e97380. Epub 2014/05/23. doi: 10.1371/journal.pone.0097380. PubMed PMID: 24846176; PMCID: PMC4028258.

55. Consortium GT. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013;45(6):580-5. Epub 2013/05/30. doi: 10.1038/ng.2653. PubMed PMID: 23715323; PMCID: PMC4010069.

56. Smale ST. Luciferase assay. Cold Spring Harb Protoc. 2010;2010(5):pdb prot5421. Epub 2010/05/05. doi: 10.1101/pdb.prot5421. PubMed PMID: 20439408.

57. Visel A, Minovitsky S, Dubchak I, Pennacchio LA. VISTA Enhancer Browser--a database of tissue-specific human enhancers. Nucleic Acids Res. 2007;35(Database issue):D88-92. Epub 2006/11/30. doi: 10.1093/nar/gkl822. PubMed PMID: 17130149; PMCID: PMC1716724.

58. Melnikov A, Zhang X, Rogov P, Wang L, Mikkelsen TS. Massively parallel reporter assays in cultured mammalian cells. J Vis Exp. 2014(90). Epub 2014/09/02. doi: 10.3791/51719. PubMed PMID: 25177895; PMCID: PMC4364389.

59. Muerdter F, Boryn LM, Arnold CD. STARR-seq - principles and applications. Genomics. 2015;106(3):145-50. Epub 2015/06/15. doi: 10.1016/j.ygeno.2015.06.001. PubMed PMID: 26072434.

60. Merkle FT, Neuhausser WM, Santos D, Valen E, Gagnon JA, Maas K, Sandoe J, Schier AF, Eggan K. Efficient CRISPR-Cas9-mediated generation of knockin human pluripotent stem cells lacking undesired mutations at the targeted locus. Cell Rep. 2015;11(6):875-83. Epub 2015/05/06. doi: 10.1016/j.celrep.2015.04.007. PubMed PMID: 25937281; PMCID: PMC5533178.

61. Beaudoin M, Gupta RM, Won HH, Lo KS, Do R, Henderson CA, Lavoie-St-Amour C, Langlois S, Rivas D, Lehoux S, Kathiresan S, Tardif JC, Musunuru K, Lettre G. Myocardial Infarction-Associated SNP at 6p24 Interferes With MEF2 Binding and Associates With PHACTR1 Expression Levels in Human Coronary Arteries. Arterioscler Thromb Vasc Biol. 2015;35(6):1472-9. Epub 2015/04/04. doi: 10.1161/ATVBAHA.115.305534. PubMed PMID: 25838425; PMCID: PMC4441556.

62. Maeder ML, Linder SJ, Cascio VM, Fu Y, Ho QH, Joung JK. CRISPR RNA-guided activation of endogenous human genes. Nat Methods. 2013;10(10):977-9. Epub 2013/07/31. doi: 10.1038/nmeth.2598. PubMed PMID: 23892898; PMCID: PMC3794058.

63. Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, Stern-Ginossar N, Brandman O, Whitehead EH, Doudna JA, Lim WA, Weissman JS, Qi LS. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell. 2013;154(2):442-51. Epub 2013/07/16. doi: 10.1016/j.cell.2013.06.044. PubMed PMID: 23849981; PMCID: PMC3770145.

64. McNally EM, Golbus JR, Puckelwartz MJ. Genetic mutations and mechanisms in dilated cardiomyopathy. J Clin Invest. 2013;123(1):19-26. Epub 2013/01/03. doi: 10.1172/JCI62862. PubMed PMID: 23281406; PMCID: PMC3533274.

65. Konno T, Chang S, Seidman JG, Seidman CE. Genetics of hypertrophic cardiomyopathy. Curr Opin Cardiol. 2010;25(3):205-9. Epub 2010/02/04. doi: 10.1097/HCO.0b013e3283375698. PubMed PMID: 20124998; PMCID: PMC2932754.

66. Maron BJ, Rowin EJ, Maron MS. Global Burden of Hypertrophic Cardiomyopathy. JACC Heart Fail. 2018;6(5):376-8. Epub 2018/05/05. doi: 10.1016/j.jchf.2018.03.004. PubMed PMID: 29724362.

67. Lesyuk W, Kriza C, Kolominsky-Rabas P. Cost-of-illness studies in heart failure: a systematic review 2004-2016. BMC Cardiovasc Disord. 2018;18(1):74. Epub 2018/05/03. doi: 10.1186/s12872-018-0815-3. PubMed PMID: 29716540; PMCID: PMC5930493.

68. Mestroni L, Maisch B, McKenna WJ, Schwartz K, Charron P, Rocco C, Tesson F, Richter A, Wilke A, Komajda M. Guidelines for the study of familial dilated cardiomyopathies. Collaborative Research Group of the European Human and Capital Mobility Project on Familial Dilated Cardiomyopathy. Eur Heart J. 1999;20(2):93-102. Epub 1999/04/01. doi: 10.1053/euhj.1998.1145. PubMed PMID: 10099905.

69. Franaszczyk M, Chmielewski P, Truszkowska G, Stawinski P, Michalak E, Rydzanicz M, Sobieszczanska-Malek M, Pollak A, Szczygiel J, Kosinska J, Parulski A, Stoklosa T, Tarnowska A, Machnicki MM, Foss-Nieradko B, Szperl M, Sioma A, Kusmierczyk M, Grzybowski J, Zielinski T, Ploski R, Bilinska ZT. Titin Truncating Variants in Dilated Cardiomyopathy - Prevalence and Genotype-Phenotype Correlations. PLoS One. 2017;12(1):e0169007. Epub 2017/01/04. doi: 10.1371/journal.pone.0169007. PubMed PMID: 28045975; PMCID: PMC5207678.

70. Rankin J, Ellard S. The laminopathies: a clinical review. Clin Genet. 2006;70(4):261-74. Epub 2006/09/13. doi: 10.1111/j.1399-0004.2006.00677.x. PubMed PMID: 16965317.

71.de Leeuw R, Gruenbaum Y, Medalia O. Nuclear Lamins: Thin Filaments with Major Functions. Trends Cell Biol. 2018;28(1):34-45. Epub 2017/09/13. doi: 10.1016/j.tcb.2017.08.004. PubMed PMID: 28893461.

72. Ho CY, Day SM, Ashley EA, Michels M, Pereira AC, Jacoby D, Cirino AL, Fox JC, Lakdawala NK, Ware JS, Caleshu CA, Helms AS, Colan SD, Girolami F, Cecchi F, Seidman CE, Sajeev G, Signorovitch J, Green EM, Olivotto I. Genotype and Lifetime Burden of Disease in Hypertrophic Cardiomyopathy: Insights from the Sarcomeric Human Cardiomyopathy Registry (SHaRe). Circulation. 2018;138(14):1387-98. Epub 2018/10/10. doi: 10.1161/CIRCULATIONAHA.117.033200. PubMed PMID: 30297972; PMCID: PMC6170149.

73. Sabater-Molina M, Perez-Sanchez I, Hernandez Del Rincon JP, Gimeno JR. Genetics of hypertrophic cardiomyopathy: A review of current state. Clin Genet. 2018;93(1):3-14. Epub 2017/04/04. doi: 10.1111/cge.13027. PubMed PMID: 28369730.

74. Toepfer CN, Wakimoto H, Garfinkel AC, McDonough B, Liao D, Jiang J, Tai AC, Gorham JM, Lunde IG, Lun M, Lynch TLt, McNamara JW, Sadayappan S, Redwood CS, Watkins HC, Seidman JG, Seidman CE. Hypertrophic cardiomyopathy mutations in MYBPC3 dysregulate myosin. Sci Transl Med. 2019;11(476). Epub 2019/01/25. doi: 10.1126/scitranslmed.aat1199. PubMed PMID: 30674652; PMCID: PMC7184965.

75. Adhikari AS, Trivedi DV, Sarkar SS, Song D, Kooiker KB, Bernstein D, Spudich JA, Ruppel KM. beta-Cardiac myosin hypertrophic cardiomyopathy mutations release sequestered heads and increase enzymatic activity. Nat Commun. 2019;10(1):2685. Epub 2019/06/20. doi: 10.1038/s41467-019-10555-9. PubMed PMID: 31213605; PMCID: PMC6582153.

76. Hershberger RE, Hedges DJ, Morales A. Dilated cardiomyopathy: the complexity of a diverse genetic architecture. Nat Rev Cardiol. 2013;10(9):531-47. Epub 2013/08/01. doi: 10.1038/nrcardio.2013.105. PubMed PMID: 23900355.

77. Fananapazir L, Epstein ND. Genotype-phenotype correlations in hypertrophic cardiomyopathy. Insights provided by comparisons of kindreds with distinct and identical beta-myosin heavy chain gene mutations. Circulation. 1994;89(1):22-32. Epub 1994/01/01. doi: 10.1161/01.cir.89.1.22. PubMed PMID: 8281650.

78. McNally EM, Barefield DY, Puckelwartz MJ. The genetic landscape of cardiomyopathy and its role in heart failure. Cell Metab. 2015;21(2):174-82. Epub 2015/02/05. doi: 10.1016/j.cmet.2015.01.013. PubMed PMID: 25651172; PMCID: PMC4331062.

79. Jacoby D, McKenna WJ. Genetics of inherited cardiomyopathy. Eur Heart J. 2012;33(3):296-304. Epub 2011/08/04. doi: 10.1093/eurheartj/ehr260. PubMed PMID: 21810862; PMCID: PMC3270042.

80. Mialet Perez J, Rathz DA, Petrashevskaya NN, Hahn HS, Wagoner LE, Schwartz A, Dorn GW, Liggett SB. Beta 1-adrenergic receptor polymorphisms confer differential function and predisposition to heart failure. Nat Med. 2003;9(10):1300-5. Epub 2003/09/23. doi: 10.1038/nm930. PubMed PMID: 14502278.

81. Cappola TP, Li M, He J, Ky B, Gilmore J, Qu L, Keating B, Reilly M, Kim CE, Glessner J, Frackelton E, Hakonarson H, Syed F, Hindes A, Matkovich SJ, Cresci S, Dorn GW, 2nd. Common variants in HSPB7 and FRMD4B associated with advanced heart failure. Circ Cardiovasc Genet. 2010;3(2):147-54. Epub 2010/02/04. doi: 10.1161/CIRCGENETICS.109.898395. PubMed PMID: 20124441; PMCID: PMC2957840.

82. Cappola TP, Matkovich SJ, Wang W, van Booven D, Li M, Wang X, Qu L, Sweitzer NK, Fang JC, Reilly MP, Hakonarson H, Nerbonne JM, Dorn GW, 2nd. Loss-of-function DNA sequence variant in the CLCNKA chloride channel implicates the cardio-renal axis in interindividual heart failure risk variation. Proc Natl Acad Sci U S A. 2011;108(6):2456-61. Epub 2011/01/21. doi: 10.1073/pnas.1017494108. PubMed PMID: 21248228; PMCID: PMC3038744.

83. Heydemann A, Ceco E, Lim JE, Hadhazy M, Ryder P, Moran JL, Beier DR, Palmer AA, McNally EM. Latent TGF-beta-binding protein 4 modifies muscular dystrophy in mice. J Clin Invest. 2009;119(12):3703-12. Epub 2009/11/04. doi: 10.1172/JCI39845. PubMed PMID: 19884661; PMCID: PMC2786802.

84. Tesson F, Dufour C, Moolman JC, Carrier L, al-Mahdawi S, Chojnowska L, Dubourg O, Soubrier E, Brink P, Komajda M, Guicheney P, Schwartz K, Feingold J. The influence of the angiotensin I converting enzyme genotype in familial hypertrophic cardiomyopathy varies with the disease gene mutation. J Mol Cell Cardiol. 1997;29(2):831-8. Epub 1997/02/01. doi: 10.1006/jmcc.1996.0332. PubMed PMID: 9140839.

85. Marian AJ. Modifier genes for hypertrophic cardiomyopathy. Curr Opin Cardiol. 2002;17(3):242-52. Epub 2002/05/17. PubMed PMID: 12015473; PMCID: PMC2775140.

86. Edwards SL, Beesley J, French JD, Dunning AM. Beyond GWASs: illuminating the dark road from association to function. Am J Hum Genet. 2013;93(5):779-97. Epub 2013/11/12. doi: 10.1016/j.ajhg.2013.10.012. PubMed PMID: 24210251; PMCID: PMC3824120.

87. Wang X, Tucker NR, Rizki G, Mills R, Krijger PH, de Wit E, Subramanian V, Bartell E, Nguyen XX, Ye J, Leyton-Mange J, Dolmatova EV, van der Harst P, de Laat W, Ellinor PT, Newton-Cheh C, Milan DJ, Kellis M, Boyer LA. Discovery and validation of sub-threshold genome-wide association study loci using epigenomic signatures. Elife. 2016;5. Epub 2016/05/11. doi: 10.7554/eLife.10557. PubMed PMID: 27162171; PMCID: PMC4862755.

88. Smemo S, Campos LC, Moskowitz IP, Krieger JE, Pereira AC, Nobrega MA. Regulatory variation in a TBX5 enhancer leads to isolated congenital heart disease. Hum Mol Genet. 2012;21(14):3255-63. Epub 2012/05/01. doi: 10.1093/hmg/dds165. PubMed PMID: 22543974; PMCID: PMC3384386.

89.van den Boogaard M, van Weerd JH, Bawazeer AC, Hooijkaas IB, van de Werken HJG, Tessadori F, de Laat W, Barnett P, Bakkers J, Christoffels VM. Identification and Characterization of a Transcribed Distal Enhancer Involved in Cardiac Kcnh2 Regulation. Cell Rep. 2019;28(10):2704-14 e5. Epub 2019/09/05. doi: 10.1016/j.celrep.2019.08.007. PubMed PMID: 31484079.

90. Sotoodehnia N, Isaacs A, de Bakker PI, Dorr M, Newton-Cheh C, Nolte IM, van der Harst P, Muller M, Eijgelsheim M, Alonso A, Hicks AA, Padmanabhan S, Hayward C, Smith AV, Polasek O, Giovannone S, Fu J, Magnani JW, Marciante KD, Pfeufer A, Gharib SA, Teumer A, Li M, Bis JC, Rivadeneira F, et al. Common variants in 22 loci are associated with QRS duration and cardiac ventricular conduction. Nat Genet. 2010;42(12):1068-76. Epub 2010/11/16. doi: 10.1038/ng.716. PubMed PMID: 21076409; PMCID: PMC3338195.

91. Vincentz JW, Firulli BA, Toolan KP, Arking DE, Sotoodehnia N, Wan J, Chen PS, de Gier-de Vries C, Christoffels VM, Rubart-von der Lohe M, Firulli AB. Variation in a Left Ventricle-Specific Hand1 Enhancer Impairs GATA Transcription Factor Binding and Disrupts Conduction System Development and Function. Circ Res. 2019;125(6):575-89. Epub 2019/08/02. doi: 10.1161/CIRCRESAHA.119.315313. PubMed PMID: 31366290; PMCID: PMC6715539.

92.van Ouwerkerk AF, Bosada FM, van Duijvenboden K, Hill MC, Montefiori LE, Scholman KT, Liu J, de Vries AAF, Boukens BJ, Ellinor PT, Goumans M, Efimov IR, Nobrega MA, Barnett P, Martin JF, Christoffels VM. Identification of atrial fibrillation associated genes and functional non-coding variants. Nat Commun. 2019;10(1):4755. Epub 2019/10/20. doi: 10.1038/s41467-019-12721-5. PubMed PMID: 31628324; PMCID: PMC6802215.

93. Christophersen IE, Rienstra M, Roselli C, Yin X, Geelhoed B, Barnard J, Lin H, Arking DE, Smith AV, Albert CM, Chaffin M, Tucker NR, Li M, Klarin D, Bihlmeyer NA, Low SK, Weeke PE, Muller-Nurasyid M, Smith JG, Brody JA, Niemeijer MN, Dorr M, Trompet S, Huffman J, Gustafsson S, et al. Large-scale analyses of common and rare variants identify 12 new loci associated with atrial fibrillation. Nat Genet. 2017;49(6):946-52. Epub 2017/04/19. doi: 10.1038/ng.3843. PubMed PMID: 28416818; PMCID: PMC5585859.

94.van Setten J, Brody JA, Jamshidi Y, Swenson BR, Butler AM, Campbell H, Del Greco FM, Evans DS, Gibson Q, Gudbjartsson DF, Kerr KF, Krijthe BP, Lyytikainen LP, Muller C, Muller-Nurasyid M, Nolte IM, Padmanabhan S, Ritchie MD, Robino A, Smith AV, Steri M, Tanaka T, Teumer A, Trompet S, Ulivi S, et al. PR interval genome-wide association meta-analysis identifies 50 loci associated with atrial and atrioventricular electrical activity. Nat Commun. 2018;9(1):2904. Epub 2018/07/27. doi: 10.1038/s41467-018-04766-9. PubMed PMID: 30046033; PMCID: PMC6060178.

95. Wilde AAM, Amin AS. Clinical Spectrum of SCN5A Mutations: Long QT Syndrome, Brugada Syndrome, and Cardiomyopathy. JACC Clin Electrophysiol. 2018;4(5):569-79. Epub 2018/05/26. doi: 10.1016/j.jacep.2018.03.006. PubMed PMID: 29798782.

96.van den Boogaard M, Smemo S, Burnicka-Turek O, Arnolds DE, van de Werken HJ, Klous P, McKean D, Muehlschlegel JD, Moosmann J, Toka O, Yang XH, Koopmann TT, Adriaens ME, Bezzina CR, de Laat W, Seidman C, Seidman JG, Christoffels VM, Nobrega MA, Barnett P, Moskowitz IP. A common genetic variant within SCN10A modulates cardiac SCN5A expression. J Clin Invest. 2014;124(4):1844-52. Epub 2014/03/20. doi: 10.1172/JCI73140. PubMed PMID: 24642470; PMCID: PMC3973109.

97. Man JCK, Mohan RA, Boogaard MVD, Hilvering CRE, Jenkins C, Wakker V, Bianchi V, Laat W, Barnett P, Boukens BJ, Christoffels VM. An enhancer cluster controls gene activity and topology of the SCN5A-SCN10A locus in vivo. Nat Commun. 2019;10(1):4943. Epub 2019/11/02. doi: 10.1038/s41467-019-12856-5. PubMed PMID: 31666509; PMCID: PMC6821807.

98. Sergeeva IA, Hooijkaas IB, Ruijter JM, van der Made I, de Groot NE, van de Werken HJ, Creemers EE, Christoffels VM. Identification of a regulatory domain controlling the Nppa-Nppb gene cluster during heart development and stress. Development. 2016;143(12):2135-46. Epub 2016/04/07. doi: 10.1242/dev.132019. PubMed PMID: 27048739.

99.van Weerd JH, Badi I, van den Boogaard M, Stefanovic S, van de Werken HJ, Gomez-Velazquez M, Badia-Careaga C, Manzanares M, de Laat W, Barnett P, Christoffels VM. A large permissive regulatory domain exclusively controls Tbx3 expression in the cardiac conduction system. Circ Res. 2014;115(4):432-41. Epub 2014/06/26. doi: 10.1161/CIRCRESAHA.115.303591. PubMed PMID: 24963028.

100. Rau CD, Lusis AJ, Wang Y. Genetics of common forms of heart failure: challenges and potential solutions. Curr Opin Cardiol. 2015;30(3):222-7. Epub 2015/03/15. doi: 10.1097/HCO.0000000000000160. PubMed PMID: 25768955; PMCID: PMC4406340.

101. Bagnall RD, Molloy LK, Kalman JM, Semsarian C. Exome sequencing identifies a mutation in the ACTN2 gene in a family with idiopathic ventricular fibrillation, left ventricular noncompaction, and sudden death. BMC Med Genet. 2014;15:99. Epub 2014/09/17. doi: 10.1186/s12881-014-0099-0. PubMed PMID: 25224718; PMCID: PMC4355500.

102. Prondzynski M, Lemoine MD, Zech AT, Horvath A, Di Mauro V, Koivumaki JT, Kresin N, Busch J, Krause T, Kramer E, Schlossarek S, Spohn M, Friedrich FW, Munch J, Laufer SD, Redwood C, Volk AE, Hansen A, Mearini G, Catalucci D, Meyer C, Christ T, Patten M, Eschenhagen T, Carrier L. Disease modeling of a mutation in alpha-actinin 2 guides clinical therapy in hypertrophic cardiomyopathy. EMBO Mol Med. 2019;11(12):e11115. Epub 2019/11/05. doi: 10.15252/emmm.201911115. PubMed PMID: 31680489; PMCID: PMC6895603.

103. Heinig M, Adriaens ME, Schafer S, van Deutekom HWM, Lodder EM, Ware JS, Schneider V, Felkin LE, Creemers EE, Meder B, Katus HA, Ruhle F, Stoll M, Cambien F, Villard E, Charron P, Varro A, Bishopric NH, George AL, Jr., Dos Remedios C, Moreno-Moral A, Pesce F, Bauerfeind A, Ruschendorf F, Rintisch C, et al. Natural genetic variation of the cardiac transcriptome in non-diseased donors and patients with dilated cardiomyopathy. Genome Biol. 2017;18(1):170. Epub 2017/09/15. doi: 10.1186/s13059-017-1286-z. PubMed PMID: 28903782; PMCID: PMC5598015.

104. Razeghi P, Young ME, Alcorn JL, Moravec CS, Frazier OH, Taegtmeyer H. Metabolic gene expression in fetal and failing human heart. Circulation. 2001;104(24):2923-31. Epub 2001/12/12. doi: 10.1161/hc4901.100526. PubMed PMID: 11739307.

105. Miyata S, Minobe W, Bristow MR, Leinwand LA. Myosin heavy chain isoform expression in the failing and nonfailing human heart. Circ Res. 2000;86(4):386-90. Epub 2000/03/04. doi: 10.1161/01.res.86.4.386. PubMed PMID: 10700442.

106. Yin Z, Ren J, Guo W. Sarcomeric protein isoform transitions in cardiac muscle: a journey to heart failure. Biochimica et biophysica acta. 2015;1852(1):47-52. Epub 2014/12/03. doi: 10.1016/j.bbadis.2014.11.003. PubMed PMID: 25446994; PMCID: PMC4268308.

107. Anderson PA, Malouf NN, Oakeley AE, Pagani ED, Allen PD. Troponin T isoform expression in humans. A comparison among normal and failing adult heart, fetal heart, and adult and fetal skeletal muscle. Circ Res. 1991;69(5):1226-33. Epub 1991/11/01. doi: 10.1161/01.res.69.5.1226. PubMed PMID: 1934353.

108. Makarenko I, Opitz CA, Leake MC, Neagoe C, Kulke M, Gwathmey JK, del Monte F, Hajjar RJ, Linke WA. Passive stiffness changes caused by upregulation of compliant titin isoforms in human dilated cardiomyopathy hearts. Circ Res. 2004;95(7):708-16. Epub 2004/09/04. doi: 10.1161/01.RES.0000143901.37063.2f. PubMed PMID: 15345656.

109. Beqqali A. Alternative splicing in cardiomyopathy. Biophys Rev. 2018;10(4):1061-71. Epub 2018/07/28. doi: 10.1007/s12551-018-0439-y. PubMed PMID: 30051286; PMCID: PMC6082314.

110. Mestroni L, Brun F, Spezzacatene A, Sinagra G, Taylor MR. Genetic Causes of Dilated Cardiomyopathy. Prog Pediatr Cardiol. 2014;37(1-2):13-8. Epub 2015/01/15. doi: 10.1016/j.ppedcard.2014.10.003. PubMed PMID: 25584016; PMCID: PMC4288017.

111. He A, Gu F, Hu Y, Ma Q, Ye LY, Akiyama JA, Visel A, Pennacchio LA, Pu WT. Dynamic GATA4 enhancers shape the chromatin landscape central to heart development and disease. Nat Commun. 2014;5:4907. Epub 2014/09/25. doi: 10.1038/ncomms5907. PubMed PMID: 25249388; PMCID: PMC4236193.

112. Wamstad JA, Alexander JM, Truty RM, Shrikumar A, Li F, Eilertson KE, Ding H, Wylie JN, Pico AR, Capra JA, Erwin G, Kattman SJ, Keller GM, Srivastava D, Levine SS, Pollard KS, Holloway AK, Boyer LA, Bruneau BG. Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. Cell. 2012;151(1):206-20. Epub 2012/09/18. doi: 10.1016/j.cell.2012.07.035. PubMed PMID: 22981692; PMCID: PMC3462286.

113. Paige SL, Thomas S, Stoick-Cooper CL, Wang H, Maves L, Sandstrom R, Pabon L, Reinecke H, Pratt G, Keller G, Moon RT, Stamatoyannopoulos J, Murry CE. A temporal chromatin signature in human embryonic stem cells identifies regulators of cardiac development. Cell. 2012;151(1):221-32. Epub 2012/09/18. doi: 10.1016/j.cell.2012.08.027. PubMed PMID: 22981225; PMCID: PMC3462257.

114. Rosa-Garrido M, Chapski DJ, Schmitt AD, Kimball TH, Karbassi E, Monte E, Balderas E, Pellegrini M, Shih TT, Soehalim E, Liem D, Ping P, Galjart NJ, Ren S, Wang Y, Ren B, Vondriska TM. High-Resolution Mapping of Chromatin Conformation in Cardiac Myocytes Reveals Structural Remodeling of the Epigenome in Heart Failure. Circulation. 2017;136(17):1613-25. Epub 2017/08/13. doi: 10.1161/circulationaha.117.029430. PubMed PMID: 28802249; PMCID: PMC5648689.

115. Dickel DE, Barozzi I, Zhu Y, Fukuda-Yuzawa Y, Osterwalder M, Mannion BJ, May D, Spurrell CH, Plajzer-Frick I, Pickle CS, Lee E, Garvin TH, Kato M, Akiyama JA, Afzal V, Lee AY, Gorkin DU, Ren B, Rubin EM, Visel A, Pennacchio LA. Genome-wide compendium and functional assessment of in vivo heart enhancers. Nat Commun. 2016;7:12923. Epub 2016/10/06. doi: 10.1038/ncomms12923. PubMed PMID: 27703156; PMCID: PMC5059478.

116. Kimura K, Wakamatsu A, Suzuki Y, Ota T, Nishikawa T, Yamashita R, Yamamoto J, Sekine M, Tsuritani K, Wakaguri H, Ishii S, Sugiyama T, Saito K, Isono Y, Irie R, Kushida N, Yoneyama T, Otsuka R, Kanda K, Yokoi T, Kondo H, Wagatsuma M, Murakawa K, Ishida S, Ishibashi T, et al. Diversification of transcriptional modulation: large-scale identification and characterization of putative alternative promoters of human genes. Genome Res. 2006;16(1):55-65. Epub 2005/12/14. doi: 10.1101/gr.4039406. PubMed PMID: 16344560; PMCID: PMC1356129.

117. Zou J, Tran D, Baalbaki M, Tang LF, Poon A, Pelonero A, Titus EW, Yuan C, Shi C, Patchava S, Halper E, Garg J, Movsesyan I, Yin C, Wu R, Wilsbacher LD, Liu J, Hager RL, Coughlin SR, Jinek M, Pullinger CR, Kane JP, Hart DO, Kwok PY, Deo RC. An internal promoter underlies the difference in disease severity between N- and C-terminal truncation mutations of Titin in zebrafish. Elife. 2015;4:e09406. Epub 2015/10/17. doi: 10.7554/eLife.09406. PubMed PMID: 26473617; PMCID: PMC4720518.

118. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15-21. Epub 2012/10/30. doi: 10.1093/bioinformatics/bts635. PubMed PMID: 23104886; PMCID: PMC3530905.

119. Haberle V, Forrest AR, Hayashizaki Y, Carninci P, Lenhard B. CAGEr: precise TSS data retrieval and high-resolution promoterome mining for integrative analyses. Nucleic Acids Res. 2015;43(8):e51. Epub 2015/02/06. doi: 10.1093/nar/gkv054. PubMed PMID: 25653163; PMCID: PMC4417143.

120. Thodberg MT, A..Vitting-Seerup, K.. Andersson, R.. Sandelin, A. CAGEfightR: Cap Analysis of Gene Expression (CAGE) in R/Bioconductor. bioRxiv. 2018. doi: https://doi.org/10.1101/310623

121. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841-2. Epub 2010/01/30. doi: 10.1093/bioinformatics/btq033. PubMed PMID: 20110278; PMCID: PMC2832824.

122. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. Genome Res. 2004;14(6):1188-90. Epub 2004/06/03. doi: 10.1101/gr.849004. PubMed PMID: 15173120; PMCID: PMC419797.

123. Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, Diemer K, Muruganujan A, Narechania A. PANTHER: a library of protein families and subfamilies indexed by function. Genome Res. 2003;13(9):2129-41. Epub 2003/09/04. doi: 10.1101/gr.772403. PubMed PMID: 12952881; PMCID: PMC403709.

124. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31(2):166-9. Epub 2014/09/28. doi: 10.1093/bioinformatics/btu638. PubMed PMID: 25260700; PMCID: PMC4287950.

125. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139-40. Epub 2009/11/17. doi: 10.1093/bioinformatics/btp616. PubMed PMID: 19910308; PMCID: PMC2796818.

126. Benner C, Konovalov S, Mackintosh C, Hutt KR, Stunnenberg R, Garcia-Bassets I. Decoding a signature-based model of transcription cofactor recruitment dictated by cardinal cis-regulatory elements in proximal promoter regions. PLoS Genet. 2013;9(11):e1003906. Epub 2013/11/19. doi: 10.1371/journal.pgen.1003906. PubMed PMID: 24244184; PMCID: PMC3820735.

127. Schlesinger J, Schueler M, Grunert M, Fischer JJ, Zhang Q, Krueger T, Lange M, Tonjes M, Dunkel I, Sperling SR. The cardiac transcription network modulated by Gata4, Mef2a, Nkx2.5, Srf, histone modifications, and microRNAs. PLoS Genet. 2011;7(2):e1001313. Epub 2011/03/08. doi: 10.1371/journal.pgen.1001313. PubMed PMID: 21379568; PMCID: PMC3040678.

128. Fantom, Riken, Forrest AR, Kawaji H, Rehli M, Baillie JK, de Hoon MJ, Haberle V, Lassmann T, Kulakovskiy IV, Lizio M, Itoh M, Andersson R, Mungall CJ, Meehan TF, Schmeier S, Bertin N, Jorgensen M, Dimont E, Arner E, Schmidl C, Schaefer U, Medvedeva YA, Plessy C, Vitezic M, et al. A promoter-level mammalian expression atlas. Nature. 2014;507(7493):462-70. Epub 2014/03/29. doi: 10.1038/nature13182. PubMed PMID: 24670764; PMCID: PMC4529748.

129. Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, Semple CA, Taylor MS, Engstrom PG, Frith MC, Forrest AR, Alkema WB, Tan SL, Plessy C, Kodzius R, Ravasi T, Kasukawa T, Fukuda S, Kanamori-Katayama M, Kitazume Y, Kawaji H, Kai C, Nakamura M, Konno H, Nakano K, et al. Genome-wide analysis of mammalian promoter architecture and evolution. Nat Genet. 2006;38(6):626-35. Epub 2006/04/29. doi: 10.1038/ng1789. PubMed PMID: 16645617.

130. Park SG, Hannenhalli S, Choi SS. Conservation in first introns is positively associated with the number of exons within genes and the presence of regulatory epigenetic signals. BMC Genomics. 2014;15:526. Epub 2014/06/27. doi: 10.1186/1471-2164-15-526. PubMed PMID: 24964727; PMCID: PMC4085337.

131. Porto AG, Brun F, Severini GM, Losurdo P, Fabris E, Taylor MRG, Mestroni L, Sinagra G. Clinical Spectrum of PRKAG2 Syndrome. Circ Arrhythm Electrophysiol. 2016;9(1):e003121. Epub 2016/01/06. doi: 10.1161/CIRCEP.115.003121 e003121
10.1161/CIRCEP.115.003121. PubMed PMID: 26729852; PMCID: PMC4704128.

132. Ortega A, Rosello-Lleti E, Tarazon E, Gil-Cayuela C, Lago F, Gonzalez-Juanatey JR, Martinez-Dolz L, Portoles M, Rivera M. TRPM7 is down-regulated in both left atria and left ventricle of ischaemic cardiomyopathy patients and highly related to changes in ventricular function. ESC Heart Fail. 2016;3(3):220-4. Epub 2016/11/08. doi: 10.1002/ehf2.12085. PubMed PMID: 27818786; PMCID: PMC5071679.

133. Sah R, Mesirca P, Van den Boogert M, Rosen J, Mably J, Mangoni ME, Clapham DE. Ion channel-kinase TRPM7 is required for maintaining cardiac automaticity. Proc Natl Acad Sci U S A. 2013;110(32):E3037-46. Epub 2013/07/24. doi: 10.1073/pnas.1311865110. PubMed PMID: 23878236; PMCID: PMC3740880.

134. Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, Richmond TA, Wu Y, Green RD, Ren B. A high-resolution map of active promoters in the human genome. Nature. 2005;436(7052):876-80. Epub 2005/07/01. doi: 10.1038/nature03877. PubMed PMID: 15988478; PMCID: PMC1895599.

135. Cooper SJ, Trinklein ND, Anton ED, Nguyen L, Myers RM. Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome. Genome Res. 2006;16(1):1-10. Epub 2005/12/14. doi: 10.1101/gr.4222606. PubMed PMID: 16344566; PMCID: PMC1356123.

136. Davuluri RV, Suzuki Y, Sugano S, Plass C, Huang TH. The functional consequences of alternative promoter use in mammalian genomes. Trends Genet. 2008;24(4):167-77. Epub 2008/03/11. doi: 10.1016/j.tig.2008.01.008. PubMed PMID: 18329129.

137. Oliveira SM, Zhang YH, Solis RS, Isackson H, Bellahcene M, Yavari A, Pinter K, Davies JK, Ge Y, Ashrafian H, Walker JW, Carling D, Watkins H, Casadei B, Redwood C. AMP-activated protein kinase phosphorylates cardiac troponin I and alters contractility of murine ventricular myocytes. Circ Res. 2012;110(9):1192-201. Epub 2012/03/30. doi: 10.1161/CIRCRESAHA.111.259952. PubMed PMID: 22456184.

138. Khalil H, Kanisicak O, Prasad V, Correll RN, Fu X, Schips T, Vagnozzi RJ, Liu R, Huynh T, Lee SJ, Karch J, Molkentin JD. Fibroblast-specific TGF-beta-Smad2/3 signaling underlies cardiac fibrosis. J Clin Invest. 2017;127(10):3770-83. Epub 2017/09/12. doi: 10.1172/JCI94753. PubMed PMID: 28891814; PMCID: PMC5617658.

139. Chen H, Moreno-Moral A, Pesce F, Devapragash N, Mancini M, Heng EL, Rotival M, Srivastava PK, Harmston N, Shkura K, Rackham OJL, Yu WP, Sun XM, Tee NGZ, Tan ELS, Barton PJR, Felkin LE, Lara-Pezzi E, Angelini G, Beltrami C, Pravenec M, Schafer S, Bottolo L, Hubner N, Emanueli C, et al. WWP2 regulates pathological cardiac fibrosis by modulating SMAD2 signaling. Nat Commun. 2019;10(1):3616. Epub 2019/08/11. doi: 10.1038/s41467-019-11551-9. PubMed PMID: 31399586; PMCID: PMC6689010.

140. Haggerty CM, Damrauer SM, Levin MG, Birtwell D, Carey DJ, Golden AM, Hartzel DN, Hu Y, Judy R, Kelly MA, Kember RL, Lester Kirchner H, Leader JB, Liang L, McDermott-Roe C, Babu A, Morley M, Nealy Z, Person TN, Pulenthiran A, Small A, Smelser DT, Stahl RC, Sturm AC, Williams H, et al. Genomics-First Evaluation of Heart Disease Associated With Titin-Truncating Variants. Circulation. 2019;140(1):42-54. Epub 2019/06/21. doi: 10.1161/circulationaha.119.039573. PubMed PMID: 31216868; PMCID: PMC6602806.

141. Roberts AM, Ware JS, Herman DS, Schafer S, Baksi J, Bick AG, Buchan RJ, Walsh R, John S, Wilkinson S, Mazzarotto F, Felkin LE, Gong S, MacArthur JA, Cunningham F, Flannick J, Gabriel SB, Altshuler DM, Macdonald PS, Heinig M, Keogh AM, Hayward CS, Banner NR, Pennell DJ, O'Regan DP, et al. Integrated allelic, transcriptional, and phenomic dissection of the cardiac effects of titin truncations in health and disease. Sci Transl Med. 2015;7(270):270ra6. Epub 2015/01/16. doi: 10.1126/scitranslmed.3010134. PubMed PMID: 25589632; PMCID: PMC4560092.

142. Barp A, Bello L, Politano L, Melacini P, Calore C, Polo A, Vianello S, Soraru G, Semplicini C, Pantic B, Taglia A, Picillo E, Magri F, Gorni K, Messina S, Vita GL, Vita G, Comi GP, Ermani M, Calvo V, Angelini C, Hoffman EP, Pegoraro E. Genetic Modifiers of Duchenne Muscular Dystrophy and Dilated Cardiomyopathy. PLoS One. 2015;10(10):e0141240. Epub 2015/10/30. doi: 10.1371/journal.pone.0141240. PubMed PMID: 26513582; PMCID: PMC4626372.

143. Verdonschot JAJ, Robinson EL, James KN, Mohamed MW, Claes GRF, Casas K, Vanhoutte EK, Hazebroek MR, Kringlen G, Pasierb MM, van den Wijngaard A, Glatz JFC, Heymans SRB, Krapels IPC, Nahas S, Brunner HG, Szklarczyk R. Mutations in PDLIM5 are rare in dilated cardiomyopathy but are emerging as potential disease modifiers. Mol Genet Genomic Med. 2020;8(2):e1049. Epub 2019/12/28. doi: 10.1002/mgg3.1049. PubMed PMID: 31880413; PMCID: PMC7005607.

144. Andersson R, Sandelin A. Determinants of enhancer and promoter activities of regulatory elements. Nat Rev Genet. 2020;21(2):71-87. Epub 2019/10/13. doi: 10.1038/s41576-019-0173-8. PubMed PMID: 31605096.

145. Tanjore R, Rangaraju A, Vadapalli S, Remersu S, Narsimhan C, Nallari P. Genetic variations of beta-MYH7 in hypertrophic cardiomyopathy and dilated cardiomyopathy. Indian J Hum Genet. 2010;16(2):67-71. Epub 2010/10/30. doi: 10.4103/0971-6866.69348. PubMed PMID: 21031054; PMCID: PMC2955954.

146. Kim EY, Barefield DY, Vo AH, Gacita AM, Schuster EJ, Wyatt EJ, Davis JL, Dong B, Sun C, Page P, Dellefave-Castillo L, Demonbreun A, Zhang HF, McNally EM. Distinct pathological signatures in human cellular models of myotonic dystrophy subtypes. JCI Insight. 2019;4(6). Epub 2019/02/08. doi: 10.1172/jci.insight.122686. PubMed PMID: 30730308; PMCID: PMC6482996.

147. Haeussler M, Schonig K, Eckert H, Eschstruth A, Mianne J, Renaud JB, Schneider-Maunoury S, Shkumatava A, Teboul L, Kent J, Joly JS, Concordet JP. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. Genome Biol. 2016;17(1):148. Epub 2016/07/07. doi: 10.1186/s13059-016-1012-2. PubMed PMID: 27380939; PMCID: PMC4934014.

148. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez JY, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A. Fiji: an open-source platform for biological-image analysis. Nat Methods. 2012;9(7):676-82. Epub 2012/06/30. doi: 10.1038/nmeth.2019. PubMed PMID: 22743772; PMCID: PMC3855844.

149. Sala L, van Meer BJ, Tertoolen LGJ, Bakkers J, Bellin M, Davis RP, Denning C, Dieben MAE, Eschenhagen T, Giacomelli E, Grandela C, Hansen A, Holman ER, Jongbloed MRM, Kamel SM, Koopman CD, Lachaud Q, Mannhardt I, Mol MPH, Mosqueira D, Orlova VV, Passier R, Ribeiro MC, Saleem U, Smith GL, et al. MUSCLEMOTION: A Versatile Open Software Tool to Quantify Cardiomyocyte and Cardiac Muscle Contraction In Vitro and In Vivo. Circ Res. 2018;122(3):e5-e16. Epub 2017/12/29. doi: 10.1161/CIRCRESAHA.117.312067. PubMed PMID: 29282212; PMCID: PMC5805275.

150. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297-303. Epub 2010/07/21. doi: 10.1101/gr.107524.110. PubMed PMID: 20644199; PMCID: PMC2928508.

151. Jones BL, Nagin DS. Advances in Group-Based Trajectory Modeling and an SAS Procedure for Estimating Them. Sociological Methods & Research. 2007;35(4):542-71. doi: 10.1177/0049124106292364.

152. Morkin E. Regulation of myosin heavy chain genes in the heart. Circulation. 1993;87(5):1451-60. Epub 1993/05/01. doi: 10.1161/01.cir.87.5.1451. PubMed PMID: 8490999.

153. VanBuren P, Harris DE, Alpert NR, Warshaw DM. Cardiac V1 and V3 myosins differ in their hydrolytic and mechanical activities in vitro. Circ Res. 1995;77(2):439-44. Epub 1995/08/01. doi: 10.1161/01.res.77.2.439. PubMed PMID: 7614728.

154. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, Gauthier LD, Brand H, Solomonson M, Watts NA, Rhodes D, Singer-Berk M, England EM, Seaby EG, Kosmicki JA, Walters RK, Tashman K, Farjoun Y, Banks E, Poterba T, Wang A, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. bioRxiv. 2020:531210. doi: 10.1101/531210.

155. Lundquist MR, Storaska AJ, Liu TC, Larsen SD, Evans T, Neubig RR, Jaffrey SR. Redox modification of nuclear actin by MICAL-2 regulates SRF signaling. Cell. 2014;156(3):563-76. Epub 2014/01/21. doi: 10.1016/j.cell.2013.12.035. PubMed PMID: 24440334; PMCID: PMC4384661.

156. Houweling AC, van Borren MM, Moorman AF, Christoffels VM. Expression and regulation of the atrial natriuretic factor encoding gene Nppa during development and disease. Cardiovasc Res. 2005;67(4):583-93. Epub 2005/07/09. doi: 10.1016/j.cardiores.2005.06.013. PubMed PMID: 16002056.

157. Hershberger RE, Pinto JR, Parks SB, Kushner JD, Li D, Ludwigsen S, Cowan J, Morales A, Parvatiyar MS, Potter JD. Clinical and functional characterization of TNNT2 mutations identified in patients with dilated cardiomyopathy. Circ Cardiovasc Genet. 2009;2(4):306-13. Epub 2009/12/25. doi: 10.1161/CIRCGENETICS.108.846733. PubMed PMID: 20031601; PMCID: PMC2900844.

158. Yilbas AE, Hamilton A, Wang Y, Mach H, Lacroix N, Davis DR, Chen J, Li Q. Activation of GATA4 gene expression at the early stage of cardiac specification. Front Chem. 2014;2:12. Epub 2014/05/03. doi: 10.3389/fchem.2014.00012. PubMed PMID: 24790981; PMCID: PMC3982529.

159. Pott S, Lieb JD. What are super-enhancers? Nat Genet. 2015;47(1):8-12. Epub 2014/12/31. doi: 10.1038/ng.3167. PubMed PMID: 25547603.

160. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA, Hoke HA, Young RA. Super-enhancers in the control of cell identity and disease. Cell. 2013;155(4):934-47. Epub 2013/10/15. doi: 10.1016/j.cell.2013.09.053. PubMed PMID: 24119843; PMCID: PMC3841062.

161. Nakao K, Minobe W, Roden R, Bristow MR, Leinwand LA. Myosin heavy chain gene expression in human heart failure. J Clin Invest. 1997;100(9):2362-70. Epub 1997/12/31. doi: 10.1172/JCI119776. PubMed PMID: 9410916; PMCID: PMC508434.

162. Abraham WT, Gilbert EM, Lowes BD, Minobe WA, Larrabee P, Roden RL, Dutcher D, Sederberg J, Lindenfeld JA, Wolfel EE, Shakar SF, Ferguson D, Volkman K, Linseman JV, Quaife RA, Robertson AD, Bristow MR. Coordinate changes in Myosin heavy chain isoform gene expression are selectively associated with alterations in dilated cardiomyopathy phenotype. Mol Med. 2002;8(11):750-60. Epub 2003/01/10. PubMed PMID: 12520092; PMCID: PMC2039952.

163. Tripathi S, Schultz I, Becker E, Montag J, Borchert B, Francino A, Navarro-Lopez F, Perrot A, Ozcelik C, Osterziel KJ, McKenna WJ, Brenner B, Kraft T. Unequal allelic expression of wild-type and mutated beta-myosin in familial hypertrophic cardiomyopathy. Basic Res Cardiol. 2011;106(6):1041-55. Epub 2011/07/20. doi: 10.1007/s00395-011-0205-9. PubMed PMID: 21769673; PMCID: PMC3228959.

164. Jiang J, Wakimoto H, Seidman JG, Seidman CE. Allele-specific silencing of mutant Myh6 transcripts in mice suppresses hypertrophic cardiomyopathy. Science. 2013;342(6154):111-4. Epub 2013/10/05. doi: 10.1126/science.1236921. PubMed PMID: 24092743; PMCID: PMC4100553.

165. Anzai T, Yamagata T, Uosaki H. Comparative Transcriptome Landscape of Mouse and Human Hearts. Front Cell Dev Biol. 2020;8:268. Epub 2020/05/12. doi: 10.3389/fcell.2020.00268. PubMed PMID: 32391358; PMCID: PMC7188931.

166. Davidson MM, Nesti C, Palenzuela L, Walker WF, Hernandez E, Protas L, Hirano M, Isaac ND. Novel cell lines derived from adult human ventricular cardiomyocytes. J Mol Cell Cardiol. 2005;39(1):133-47. Epub 2005/05/26. doi: 10.1016/j.yjmcc.2005.03.003. PubMed PMID: 15913645.

167. Karakikes I, Ameen M, Termglinchan V, Wu JC. Human induced pluripotent stem cell-derived cardiomyocytes: insights into molecular, cellular, and functional phenotypes. Circ Res. 2015;117(1):80-8. Epub 2015/06/20. doi: 10.1161/CIRCRESAHA.117.305365. PubMed PMID: 26089365; PMCID: PMC4546707.

168. Machiraju P, Greenway SC. Current methods for the maturation of induced pluripotent stem cell-derived cardiomyocytes. World J Stem Cells. 2019;11(1):33-43. Epub 2019/02/02. doi: 10.4252/wjsc.v11.i1.33. PubMed PMID: 30705713; PMCID: PMC6354100.

169. Spears DA, Gollob MH. Genetics of inherited primary arrhythmia disorders. Appl Clin Genet. 2015;8:215-33. Epub 2015/10/02. doi: 10.2147/TACG.S55762. PubMed PMID: 26425105; PMCID: PMC4583121.

170. Tohyama S, Hattori F, Sano M, Hishiki T, Nagahata Y, Matsuura T, Hashimoto H, Suzuki T, Yamashita H, Satoh Y, Egashira T, Seki T, Muraoka N, Yamakawa H, Ohgino Y, Tanaka T, Yoichi M, Yuasa S, Murata M, Suematsu M, Fukuda K. Distinct metabolic flow enables large-scale purification of mouse and human pluripotent stem cell-derived cardiomyocytes. Cell Stem Cell. 2013;12(1):127-37. Epub 2012/11/22. doi: 10.1016/j.stem.2012.09.013. PubMed PMID: 23168164.

171. Churko JM, Garg P, Treutlein B, Venkatasubramanian M, Wu H, Lee J, Wessells QN, Chen SY, Chen WY, Chetal K, Mantalas G, Neff N, Jabart E, Sharma A, Nolan GP, Salomonis N, Wu JC. Defining human cardiac transcription factor hierarchies using integrated single-cell heterogeneity analysis. Nat Commun. 2018;9(1):4906. Epub 2018/11/23. doi: 10.1038/s41467-018-07333-4. PubMed PMID: 30464173; PMCID: PMC6249224.

172. Leinonen R, Sugawara H, Shumway M, International Nucleotide Sequence Database C. The sequence read archive. Nucleic Acids Res. 2011;39(Database issue):D19-21. Epub 2010/11/11. doi: 10.1093/nar/gkq1019. PubMed PMID: 21062823; PMCID: PMC3013647.

173. Bedada FB, Chan SS, Metzger SK, Zhang L, Zhang J, Garry DJ, Kamp TJ, Kyba M, Metzger JM. Acquisition of a quantitative, stoichiometrically conserved ratiometric marker of maturation status in stem cell-derived cardiac myocytes. Stem Cell Reports. 2014;3(4):594-605. Epub 2014/11/02. doi: 10.1016/j.stemcr.2014.07.012. PubMed PMID: 25358788; PMCID: PMC4223713.

174. Frangogiannis NG. The Extracellular Matrix in Ischemic and Nonischemic Heart Failure. Circ Res. 2019;125(1):117-46. Epub 2019/06/21. doi: 10.1161/CIRCRESAHA.119.311148. PubMed PMID: 31219741; PMCID: PMC6588179.