AWARD NUMBER: W81XWH-18-1-0161

TITLE: Interrogating the Functional Impact of Regulatory Sequences in Congenital Heart Disease

PRINCIPAL INVESTIGATOR: Nikhil V. Munshi, MD, PhD

CONTRACTING ORGANIZATION: University of Texas Southwester Medical Center
Dallas, TX 75390

REPORT DATE: July 2019

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
July 2019	Annual	1 Jul 2018 - 30 Jun 2019

4. TITLE AND SUBTITLE
Interrogating the functional impact of regulatory sequences in

Congenital heart disease

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-18-1-0161

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Nikhil V. Munshi, MD, PhD; Gary C. Hon, PhD

E-Mail:nikhil.munshi@utsouthwestern.edu; gary.hon@utsouthwestern.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

UTSW Medical Center
5323 Harry Hines Blvd
Dallas, TX 75390

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Congenital heart disease (CHD) affects nearly 1% of all newborns and continues to carry a poor overall prognosis. This failure stems largely from an incomplete knowledge of the underlying pathogenetic mechanisms. Thus, there is a critical need to obtain a comprehensive understanding of the genetic factors that disrupt cardiac development and lead to human CHD. Despite considerable progress, most genetic contributors to CHD remain unknown. Here, we propose a novel strategy to annotate the remaining "dark matter" in the genome to potentially identify a key source of the "missing heritability" that limits the scope of current diagnostic testing. Previous GWAS on patients with CHD identified common variants in the loci of 17 different genes linked to congenital abnormalities. Most of these variants are non-coding and lie within a class of regulatory elements called transcriptional enhancers. Given the key role that enhancers play in development, we postulate that enhancer variants cause cardiac developmental defects that contribute significantly to CHD. But direct evidence to support this notion is lacking. Therefore, we will test our central hypothesis that specific enhancers are required for cardiac development. The objective of this proposal is to develop a robust enhancer annotation pipeline for human cardiac development that we can use to rigorously evaluate our central hypothesis. Our rationale is that attainment of our objective will prioritize enhancers for causative association with human CHD in future proposals.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
Unclassified	Unclassified	Unclassified	Unclassified		19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

Page

1. INTRODUCTION

Congenital heart disease (CHD) affects nearly 1% of all newborns and continues to carry a poor overall prognosis. This failure stems largely from an incomplete knowledge of the underlying pathogenetic mechanisms. Thus, there is a critical need to obtain a comprehensive understanding of the genetic factors that disrupt cardiac development and lead to human CHD. Traditional linkage analysis has uncovered many single-gene mutations that explain a subset of CHD, but their allele frequencies are exceedingly rare so that the majority of familial CHD currently eludes molecular diagnosis. To identify more common mutations, investigators have used whole-exome sequencing (WES) and genome-wide association studies (GWAS) to correlate specific genetic variants with CHD and other diseases. Despite considerable progress, most genetic contributors to CHD remain unknown. Here, we propose a novel strategy to annotate the remaining “dark matter” in the genome to potentially identify a key source of the “missing heritability” that limits the scope of current diagnostic testing. Our long term goal is to understand how enhancer variation contributes to CHD. Previous GWAS on patients with CHD identified common variants in the loci of 17 different genes linked to congenital abnormalities. Most of these variants are non-coding and lie within a class of regulatory elements called transcriptional enhancers. The prevailing model is that common genetic variation of enhancers causes mild alterations in target gene expression, which contributes modestly to overall disease risk. However, GWAS cannot resolve rare genetic variants with large phenotypic effects, and WES only interrogates the coding genome. Although whole-genome sequencing can be performed, it remains difficult to distinguish deleterious from “bystander” variants. Given the key role that enhancers play in development, we postulate that enhancer variants cause cardiac developmental defects that contribute significantly to CHD. But direct evidence to support this notion is lacking. Therefore, we will test our central hypothesis that specific enhancers are required for cardiac development. The objective of this proposal is to develop a robust enhancer annotation pipeline for human cardiac development that we can use to rigorously evaluate our central hypothesis. Our rationale is that attainment of our objective will prioritize enhancers for causative association with human CHD in future proposals.

2. KEY WORDS

congenital heart disease
cardiac development
linkage analysis
allele frequency
whole-exome sequencing (WES)
genome-wide association studies (GWAS)
enhancer
common variant
genomic loci
non-coding variant
regulatory element
gene expression
rare variant
CRISPR
dCas9-KRAB
single-cell RNA-seq (scRNA-seq)
embryonic stem cell (ESC)

3. ACCOMPLISHMENTS

What were the major goals of the project?

1) **Assess 10 putative enhancers of *NKX2-5* for their roles in human cardiac development.** As a pilot study, we will apply Mosaic-seq to identify essential cardiac enhancers of the key cardiac development gene *NKX2-5*. We will identify enhancers that alter *NKX2-5* expression, examine how their perturbation affects cell state during cardiac differentiation, and compare enhancer repression with complete silencing of the *NKX2-5* gene.
2) **Evaluate the importance of 500 enhancers in a model of human cardiac development.** We have defined a list of 500 putative cardiac developmental enhancers through extensive bioinformatic analysis of existing datasets and careful literature review. Applying Mosaic-Seq, we will assess the phenotypic consequences of endogenous enhancer repression to identify those required for normal cardiomyocyte differentiation.
3) **Validate candidate enhancers by targeted genetic deletion.** To independently verify the essential cardiac enhancers identified above, we will create individual CRISPR/Cas9-mediated knockout cell lines for the top 5 candidate enhancers. We will thoroughly characterize the cardiac differentiation capacity of each line.

What was accomplished under these goals?

We have accomplished the following milestones:

1. Generated multiple stable dCas9-KRAB expressing hESC lines. **Figure 1**.
2. Confirmed the functionality of dCas9-KRAB hESCs with sgRNAs targeting transcriptional start sites (TSSs). **Figures 2-4**.
3. Determined that sgRNAs are stably expressed in dCas9-KRAB hESCs during differentiation and across multiple passages. **Figure 5.**
4. Established that dCas9-KRAB hESCs differentiate appropriately. **Figure 6**.
5. Established a single-cell atlas for embryoid body (EB) differentiation. **Figure 7**.
6. Determined the lentiviral titer required for adequate sgRNA library coverage. **Figure 8.**
7. Generated hESC lines expressing both dCas9-KRAB and a library of sgRNAs targeting 50 transcription factors (TFs) implicated in congenital heart disease. **Figure 8.**
8. Established a library of sgRNAs targeting all TBX5 sense and antisense transcripts.

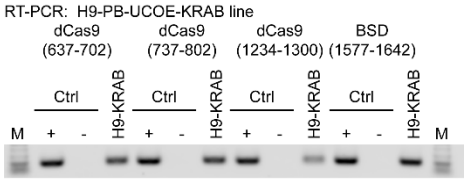

Figure 1. Confirmation of dCas9-KRAB and Blasticidin expression in stable dCas9-KRAB hESC line.

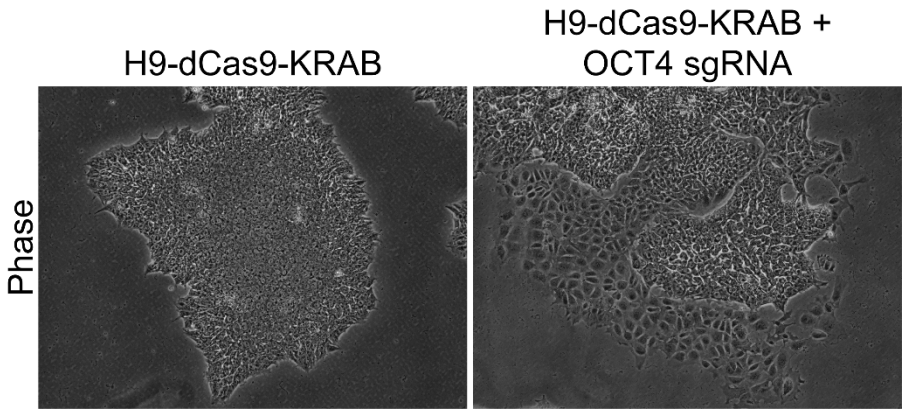

Figure 2. Transduction of a TSS-targeted OCT4 sgRNA alters colony morphology in H9-dCas9-KRAB ES cells.

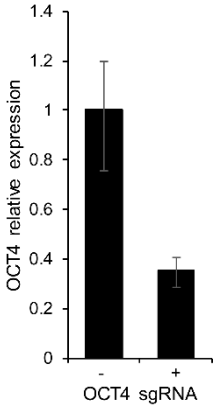

Figure 3. Transduction of a TSS-targeted sgRNA inhibits OCT4 mRNA expression.

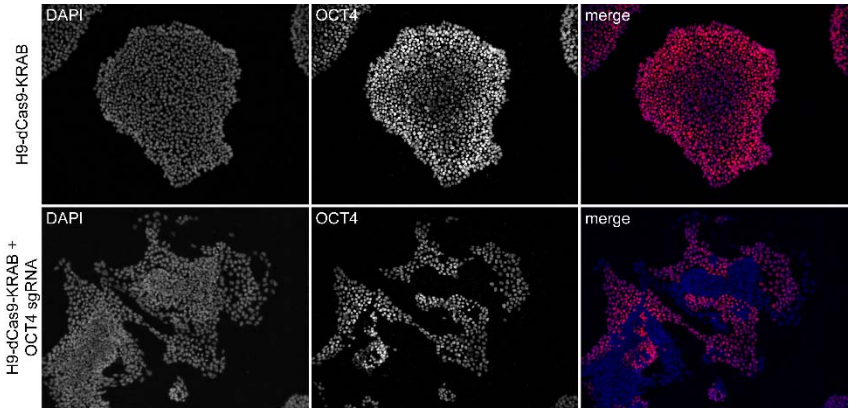

Figure 4. Transduction of a TSS-targeted sgRNA inhibits OCT4 protein expression.

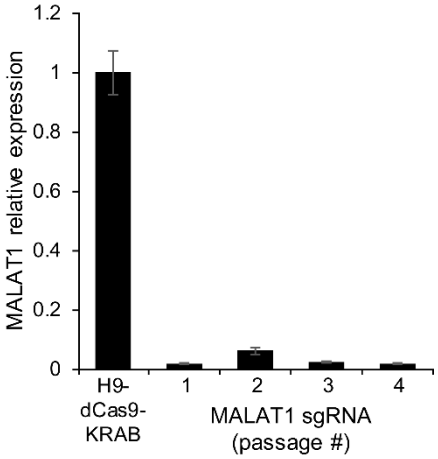

Figure 5. Stability of MALAT1 sgRNA expression and function across multiple passages.

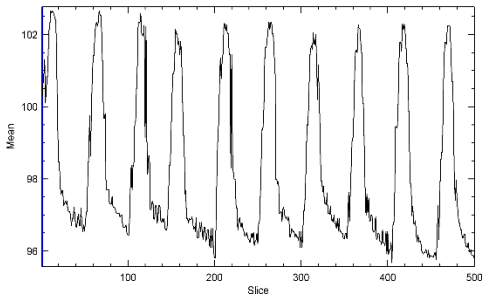

Figure 6. Normal formation of beating cardiomyocytes from dCas9-KRAB hESCs.

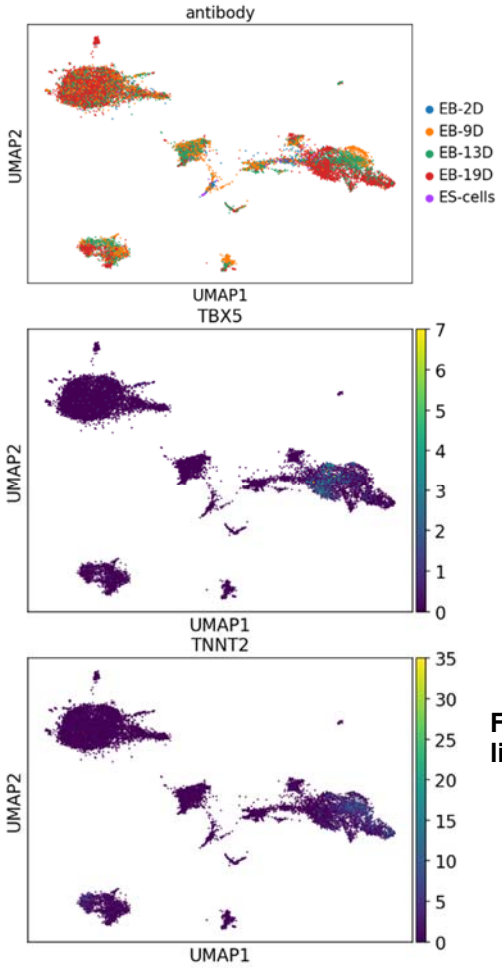

Figure 7. scRNA-seq atlas of embryoid bodies demonstrate cardiomyocyte differentiation by 19 days.

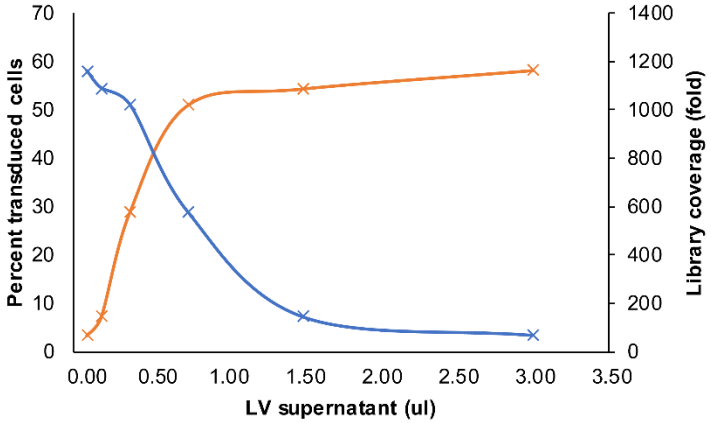

Figure 8. Transduction efficiency and library sgRNA library coverage as a function of lentivirus added.

What opportunities for training and professional development has the project provided?

This project has provided project personnel with important training and professional development opportunities. Specifically, through bimonthly joint lab meetings, trainees present their work, respond to question, and receive valuable constructive feedback. In addition, Dr. Hon and Dr. Munshi provide individualized one-on-one mentorship to Dr. Duan.

How were the results disseminated to communities of interest?

Nothing to Report.

What do you plan to do during the next reporting period to accomplish the goals?

During the next reporting period, we expect to accomplish the following milestones:

1. Perform a comprehensive analysis of the TBX5 locus to identify which of the multiple sense and antisense transcripts are required for human cardiomyocyte development.

2. Analyze loss-of-function phenotypes for 50 TFs implicated in congenital heart disease during human cardiac development.
3. Annotate TBX5 enhancers for cardiac developmental phenotypes.

4. IMPACT

What was the impact on the development of the principal discipline(s) of the project?

This first year of funding has focused on establishing the systems to accomplish the proposed Aims. As a result, the findings and results are relatively premature and incomplete. However, our experiences in establishing this system indicate that applying single-cell screens to differentiating hESCs will yield important insights on mammalian development and the molecules driving congenital heart disease.

What was the impact on other disciplines? Nothing to Report.

What was the impact on technology transfer? Nothing to Report.

What was the impact on society beyond science and technology? Nothing to Report.

5. CHANGES / PROBLEMS

Changes in approach and reasons for change:

In Aim 1, we originally proposed to target the NKX2-5 gene, which has previously been implicated in mammalian cardiac development. However, a recent report demonstrated an unclear role for NKX2-5 in human cells undergoing in vitro cardiac differentiation. Furthermore, patients that bear NXK2-5 mutations are relatively rare and difficult to phenotype clinically. Therefore, we changed our approach to target TBX5 instead, which has a clearer role in cardiac differentiation of hESCs, and patients with Holt-Oram Syndrome are more easily phenotyped by clinicians.

In Aim 2, we originally proposed to target 500 candidate enhancers with potential roles in cardiac development. However, it is uncertain if genes with published roles in cardiac development will exhibit a readout during in vitro hESC differentiation. Therefore, to focus our study, we have added a Major Task of identifying genes with potential roles in cardiac development. By targeting putative enhancers of these genes, we will enrich for functional signals.

Actual or anticipated problems or delays and actions or plans to resolve them:

As a result of incorporating the Major Task described above, we anticipate that Aim 2 will be accomplished towards the end of the funding period, which will delay the completion of Aim 3. If needed, we plan to perform Aim 3 after the funding period ends.

Changes that had a significant impact on expenditures: Nothing to Report.

Significant changes in use or care of human subjects: Nothing to Report.

Significant changes in use or care of vertebrate animals: Nothing to Report.

Significant changes in use of biohazards and/or select agents: Nothing to Report.

6. PRODUCTS

Publications, conference papers, and presentations.

Nothing to Report.

Websites or other Internet sites.

Nothing to Report.

Technologies or techniques.

Nothing to Report.

Inventions, patent applications, and/or licenses.

Nothing to Report.

Other products.

Nothing to Report.

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

What individuals have worked on the project?

Name	Nikhil Munshi
Project Role	PD/PI
Research identifier	0000-0002-8397-100X
Nearest person month worked	1
Contribution to project	Dr. Munshi has co-directed the experimental directions of this project.
Funding support	University of Texas Southwestern Medical Center, National Institutes of Health, American Heart Association

Name	Sean Goetsch
Project Role	Research scientist
Research identifier	
Nearest person month worked	2.4
Contribution to project	Mr. Goetsch has performed work in the area of human embryonic stem cell culture and manipulation.
Funding support	University of Texas Southwestern Medical Center, National Institutes of Health

Name	Gary Hon
Project Role	PD/PI
Research identifier	0000-0002-1615-0391
Nearest person month worked	1
Contribution to project	Dr. Hon has co-directed the computational directions of this project.
Funding support	University of Texas Southwestern Medical Center, National Institutes of Health, Cancer Prevention and Research Institute of Texas, Burroughs Wellcome Fund

Name	Jialei Duan
Project Role	Research scientist
Research identifier	
Nearest person month worked	6
Contribution to project	Dr. Duan has performed work in the area of genomics and bioinformatics.
Funding support	National Institutes of Health, Cancer Prevention and Research Institute of Texas

Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?

RP190451 (PI: Hon) 3/1/2019 - 2/28/2022 1.8 calendar
Cancer Prevention and Research Institute of Texas (CPRIT)
Comprehensive Evaluation of Functional Enhancers in Breast Cancer Risk Susceptibility Loci
Goal: Systematically examine the top breast cancer risk loci to identify the causal enhancers, genes, and pathways contributing to breast cancer risk.
Role: PI

1019804 (PI: Mahendroo / Hon) 6/1/2019 - 4/31/2023 0.6 calendar
Burroughs Wellcome Fund
Defining the spatio-temporal drivers of cervical remodeling in pregnancy and parturition
Goal: Systematically examine the transcriptome and epigenome of mouse cervical remodeling during pregnancy and parturition.
Role: PI

What other organizations were involved as partners?

Nothing to Report.