MUTATIONS IN GENOMIC DNA ASSOCIATED WITH SUDDEN CARDIAC DEATH

By

TIERNEY MULLANEY

Bachelor of Science in Biology

South Dakota State University

Brookings, South Dakota

2016

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
MASTER OF SCIENCE
July, 2018

MUTATIONS IN GENOMIC DNA ASSOCIATED

WITH SUDDEN CARDIAC DEATH

Thesis Approved:

Dr. Robert Allen

Thesis Adviser

Dr. Gerwald Koehler

Committee Member

Dr. Mark Payton

Committee Member

Ms. Jane Pritchard

Committee Member

ACKNOWLEDGEMENTS

I would first like to acknowledge and thank those who have significantly contributed to this study. The Office of the Chief Medical Examiner in Oklahoma, Tulsa office, helped me to gather samples from individuals who have died from cardiac episodes for my research. Dr. Mark Payton also helped to calculate and analyze my statistical results. Dr. Jun Fu was extremely supportive with the laboratory procedures and instruments in the biology student laboratories. I would also like to express gratitude to Dr. Di Shen for sharing her past research on sudden cardiac death and its correlating genes.

Secondly, I would like to thank all of my professors and classmates that have encouraged me to see my education and research to the end of this degree. I would like to thank Dr. Robert Allen, my advisor, for helping me with every step of my thesis and teaching me a lot about techniques and genetics. My committee members also deserve my expressions of thanks for helping guide me in my research and studies.

Lastly, I need to thank my supportive parents and family who have always pushed me to pursue my dreams with passion, especially when the going gets tough. I would also like to thank Tommy Culver who supported me through all of the anxiety and pressures of a graduate thesis degree. I would not have made it this far without the unending support and love they have provided.

There are many people I appreciate who I have not named personally that have

helped me through my graduate school. It is an honor to have been given the opportunity to study at OSU-CHS School of Forensic Sciences and I hope to protect and serve the public with the knowledge I have gained here.

Name: TIERNEY MULLANEY

Date of Degree: JULY, 2018

Title of Study: MUTATIONS IN GENOMIC DNA ASSOCIATED WITH SUDDEN CARDIAC DEATH

Major Field: FORENSIC SCIENCES

Abstract: Sudden cardiac death, or SCD, results from the heart failing to pump blood to the rest of the body.[1] Sudden cardiac death, the leading cause of natural death in the United States, and is caused from a variety of factors including diet, lifestyle, and genetics. This research focuses on the underlying genetic cause(s) of heart failure leading to SCD. There are hundreds of genes that can affect cardiac function and to study them all represents an enormous task. Therefore, this study concentrated on a handful of genes suspected to be associated with defects in cardiac function and possibly SCD. The genes studied included: EYA4, TNNI3, MYH6, and NEXN. These genes chosen were based upon the scientific literature and also on some preliminary studies performed in the lab using next generation sequencing on a sample of DNA obtained from a child who died from no known cause and was highly suspected of being an SCD victim. Within these genes, five single nucleotide polymorphisms (SNPs) were identified as candidates for research. SNaPshot™ is a method that allows of several steps that ultimately allow the nucleotide at a particular position within a DNA sequence to be identified. Thus, rather than producing a continuous sequence of nucleotides in a DNA fragment, SNaPshot™ technology allows the nucleotide at one specific position within a DNA fragment to be identified. Using this method, the five SNPs were evaluated to see if the nucleotide in the SNP site was mutated or wildtype. Two groups of individuals (one defined as “normal” or from individuals of older age with no known heart history and the other as “medical examiner samples” or samples from individuals who passed away from a cardiac incident) were tested to determine if there was any correlation between the SNP genotypes and death from pathology associated with the heart.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

CHAPTER I

INTRODUCTION

On July 27, 1993, at an off-season practice a 27-year-old Boston Celtics player collapsed on the court. After several attempts of resuscitation, Reggie Lewis was pronounced dead due to heart failure. The sudden death sparked public awareness of the devastating finality of sudden cardiac death. Sudden cardiac death, or SCD, is an umbrella term for the unforeseen loss of life caused by a sudden stop in heart function. Lewis's death opened the eyes of the public to the fact that SCD is not just an effect of not eating properly or exercising enough but could be caused by nucleotide sequence variations in someone's genome involving genes associated with cardiac function. Unfortunately, many different genetic defects can result in the heterogenetic condition of SCD that makes diagnostic screening for SCD difficult at best. Understanding the consequences of the large number of genetic variations that exist in genes associated with cardiac function is unclear and considerable research will have to be performed to identify the constellation of variants that, individually or in combination, lead to elevated risk for SCD. With enough time and effort, the genes that cause sudden cardiac death can be identified and put to diagnostic use in a clinical setting. Once a diagnostic protocol is developed, screening for the pathogenic nucleotides in genomic DNA can be performed

and early death from SCD can be reduced. In this study, a gene panel of 4 genes exhibiting 5 single nucleotide variations was sequenced to explore the correlation of these variations with cardiac function defects that could be related to death.

Sudden cardiac death defined, in the broadest sense, is an unexpected death that is caused by a sudden loss of cardiac function.[1] However, SCD is not a heart attack. Sudden cardiac death occurs when the electrical system of the heart is not working properly and causes irregular heartbeats. The irregularity in beating dangerously reduces the oxygen in the body, which can cause severe brain damage if not regulated. In contrast, heart attacks damage hearts when there is a defect in the coronary arteries. The defect prevents the heart from obtaining the oxygenated blood, and many people can survive the episodes. In both situations, the brain not having access to oxygen is a main concern. Emergency treatment can sometimes save the SCD victim's life, but the treatment is needed immediately to help chances of survival. Actions that can help save a victim of SCD are cardiopulmonary resuscitation, or CPR, and defibrillation. CPR helps by pushing oxygen in the lungs to the brain to maintain brain function until the heart is back to its normal rhythm, and the defibrillation helps to push the heart back into its normal contraction rhythm.[1]

In the United States of America, sudden cardiac death accounts for half of all heart disease deaths, killing around 325,000 adults each year.[1] The number of deaths attributed to SCD each year makes it the largest cause of natural death in the USA. Adults in their mid-30s are most likely to be affected by SCD, and sudden cardiac death appears twice as often in men compared to women. Sudden cardiac deaths are usually caused by arrhythmias, which are abnormal heart rhythms.[2] Arrhythmias can be caused

by any number of conditions such as coronary artery disease, electrolyte imbalances in blood (diet), heart muscle changes, and genetics. Underlying genetic causes of sudden cardiac death were the focus of this research.

Genetic causation of sudden cardiac death can stem from a single mutation that eliminates the critical functioning of a gene essential for cardiac function. However, the literature suggests that the less that optimal functioning of the products of a constellation of genes associated with cardiac function is the more likely cause of SCD. Single nucleotide polymorphisms, or SNPs, are nucleotide variations that change a DNA sequence and, when they occur in genes, SNPs can alter the amino acid sequence of a gene product (changing the amino acid) or even silence a gene altogether. SNPs are the most common type of nucleotide sequence variation in the human genome and can be detected through DNA sequencing, RFLP mapping, or other molecular methods.

The study of SCD has involved the use of massively parallel DNA sequencing (MPS) techniques[3], also known as next generation DNA sequencing because the method allows one to identify variations in a large cohort of genes that may be associated with a particular physiological function in a single molecular assay. While this approach may be time and cost effective to produce large amounts of sequence data, managing and interpreting the data within the context of a particular disease state is not straightforward, not to mention the costs associated with generating the data to begin with. Selective gene panels offer an alternative to whole genome or exome sequencing. Gene panels look at a limited number of genes instead of the whole genome or exome and can vary in number from a few to hundreds of genes depending on how correlated defects in the genes chosen are to a disease state.

When working with known SNPs in a limited number of genes associated with a particular condition, a number of molecular techniques can be used to characterize the nucleotide resident at the SNP site. This study used a method known as SNaPshot™ (Thermo Fisher Scientific, Waltham, MA) to specifically identify the nucleotides residing at 5 SNP sites located in 4 genes known to be involved in cardiac muscle structure and in the normal functioning of the heart. These SNPs therefore constituted the gene panel for the study. The SNaPshot™ method uses oligonucleotide primers that hybridize to genomic DNA template one nucleotide upstream from a SNP site. Once hybridized, the primer can be extended through the SNP site through the action of DNA polymerase adding a di-deoxynucleotide that is complementary to the template. If the ddNTP incorporated is coupled to one of 4 different fluorescent dyes, the color of the incorporated ddNDP will identify the nucleotide resident in the SNP.

Samples from 2 different population groups were obtained for this research: a collection of post mortem samples provided by the Office of the Chief Medical Examiner, Eastern Office, Tulsa, OK that were collected from cardiac incident victims and a group of samples from heart healthy, aged 40 or above, living individuals provided by the Oklahoma State University Human Identity Testing Laboratory. These 2 types of samples were tested with the chosen panel of 5 SNPs in the 4 different genes: EYA4, EYA4-2, MYH6, TNNI3, and NEXN. Nucleotides resident at these SNP sites have been identified in the literature as either "wild type" or "pathogenic" depending upon the nucleotide at the site and their associations with heart function.[3,4] Our goal was to determine SNP genotypes in a cohort of young decedents subjected to autopsy by the Office of the Chief Medical Examiner and found either to have died from cardiac

incidents or to be candidates for SCD. The incidence of pathogenic SNPs in this group was then compared with the incidence of pathogenic SNPs in the cohort of presumably normal controls.

Sudden cardiac death can be a deadly consequence of genetic mutations that occur in genes associated with heart function. Unfortunately, the rather large amount of natural genetic variation in the human genome often complicates research aimed at an understanding of multi-genic conditions such as SCD. The research performed here aims to determine whether there is any amount of correlation between the nucleotides residing at the five SNP sites in the four genes studied and cardiac related deaths. Understanding such associations will be the key to identify elevated risk through genetic testing and thereby develop prevention measures to alleviate early death from SCD.

CHAPTER II

REVIEW OF LITERATURE

Sudden cardiac death, or SCD, results from the heart failing to pump blood to the rest of the body. A SCD incident is different from a heart attack, however. A heart attack, which occurs while the individual is fully conscious, is more of a stutter in the heartbeat (which can cause a heart to stop) while a sudden cardiac attack makes the person lose consciousness and the heart stops beating altogether.[1] Sudden cardiac death is the leading cause of natural death in the United States and is caused by a variety of factors including diet, lifestyle, and genetics. The research presented here focuses on the underlying genetic cause(s) of heart failure.

Technological advances that allow researchers to examine either small or large percentages of the human genome economically and in a timely fashion have led to an increase in studies devoted to the association of gene defects with sudden cardiac death. Due to the sudden deaths in young athletes being reported more and more frequently, notoriety for SCD has also increased.

There are hundreds of genes that can affect cardiac function and to study them all would requires considerable expense; interpreting the extensive genetic data produced is complicated and often does not provide a compelling picture of the cause of an SCD

event, probably because of the coordinated failure of many genes that can contribute to the condition. Therefore, for both expediency and to keep with budget constraints this study was narrowed down to 4 genes suspected of being associated with defects in cardiac function. The 4 genes and their associated SNPs were chosen were based upon the literature and some preliminary studies performed here in the OSU Human Identify Testing laboratory using next generation sequencing on a sample of DNA obtained from a child who died from no known cause who was highly suspected of being an SCD victim. This chapter will summarize the published literature connected with background information on SCD, current methodology used to analyze the genetics associated with SCD, selection of genes for SCD research, and the specific SNPs that were chosen for this research. It is hoped our research will add to the body of knowledge surrounding the underlying genetic causes of SCD. While causative associations of particular gene defects that are uncovered may not help the victim of SCD, such information would be of value to his or her family to take preventative measures to reduce the chance of SCD in other family members.

Background Information on SCD

Sudden Cardiac Death Definition and Statistics

Sudden cardiac death is an umbrella term for the death of an individual as a result of an unexpected loss of cardiac function.[1] SCD is caused when the electrical signaling in the cardiac muscles fails to work properly and causes irregularity in the beating of the heart. Heartbeats can become either dangerously fast or dangerously slow, causing inappropriate amounts of blood to be delivered to the rest of the body. Likewise, sudden

cardiac death can result from defects in heart muscle traceable to defects in genes encoding those proteins involved in contraction of the heart. In both instances, a principal concern is that the brain will not receive sufficient oxygen due to the loss of blood flow. If immediate action is not taken the loss of oxygen can cause the victim to lose consciousness and to suffer severe brain damage. The disease can also be defined by the sudden loss of a pulse due to an abnormal heart beat rhythm with no known underlying cardiac cause.[5] The definition of SCD is important in regards to research on SCD because either a loose or a strict definition will define the source of biological samples to be used for study on the underlying cause of SCD. For example, if SCD is broadly defined, then patients used as a source of samples for research will be equally broad and studies could lead to erroneous conclusions regarding the pathology underlying SCD. In contrast, if the definition is narrow, patient samples will be fewer and some underlying causes may be missed.

A factor that sets SCD apart from other cardiac conditions is that victims suddenly lose consciousness due to the stoppage of the heart. The sudden unconsciousness experienced by victims of SCD complicate delivering medical attention because unconsciousness may result from many other more benign causes, and the loss of cardiac function may not be recognized and treatment may not be begun quickly enough for survival.

Most victims of sudden cardiac death are young; sometimes adults in their mid-30s. Men are twice as likely to be affected by SCD as women. Yearly, around 325 000 adults die from this disease in the United States.[1] This number can vary, however, depending upon how broad or narrow the definition of SCD. The number of deaths in the

United States, using the broadest definition of SCD, is as high as 400,000 individuals.[6] Even with the strong advances in cardiac resuscitation, many people still do not survive a sudden cardiac incident.[5] Although statistically uncommon in young individuals, sudden cardiac death causes from 1 to 10 deaths per 100,000 children (infant to 18-years-old) in America each year.[7] A young person's sudden cardiac death is often greatly publicized and receives considerable national attention.

The most common cause for sudden cardiac death in people under the age of 35 stems from ion channelopathies.[8] Ion channels represent "gates" inside the heart tissue that regulate the flow of ions and other molecules into heart muscle cells and help regulate heartbeat. Channelopathies occur when ion gates are not functioning properly and the heart rhythm is at risk of breaking down. The suffix -opathy denotes a disorder.[9] Appearing often in young athletes, channelopathies and hard exercise can trigger the fatal cardiac episode due to the high demand for blood flow athletic exercise causes.[8] Cardiomyopathy, a disorder of the heart muscle structure, was reported in 35% of young athletes who have died in the United States.[8] SCD can therefore result from both cardiomyopathies and ion channelopathies.

Genetic Look at SCD

The traditional method of studying a disorder involving a genetic cause is to compare suspected defective genes in the patient and in family members as well. In the case of SCD, affected individuals often do not know they carry a defect until it is too late. Scientific interest in SCD as a disease resulting from gene defects has grown over the past 12 years.[10] Moreover, advances in our ability to identify defective genes involved

with cardiac function has advanced as well, enabling gene defects that could result in SCD to be detected before the death of the individual in question. Using genetic testing to identify family members that may be at risk for a disease is known as family cascade testing.[10]

One of the problems associated with uncovering genetic defects that elevate the risk of SCD is the large number of genes involved with proper cardiac function and the fact that individual genetic defects may not precipitate SCD. Rather, defects in a number of genes, each individually not capable of causing SCD, can coordinately compromise cardiac function to the point that a high demand for blood flow during exercise for example results in SCD.[8] Many people die from this disease without even knowing they are at risk because the effects of the gene variants do not reveal themselves during normal activity. The more research that is done on the genetics of SCD, the more people can be warned and saved from an early death.

The Genes Associated with SCD Related Cardiomyopathies

Cardiomyopathy is defined as a disease that affects the muscles of the heart.[11] While not always lethal, cardiomyopathies can have serious consequences for cardiac function and discomfort in victims with this condition. There are 4 basic types of cardiomyopathy: dilated cardiomyopathy (DCM), hypertrophic cardiomyopathy (HCM), arrhythmogenic right ventricular cardiomyopathy (ARVC), and restrictive cardiomyopathy (RCM). These 4 types of cardiomyopathies differ in many ways but lead to similar pathologic conditions; some cardiomyopathies can cause sudden cardiac death.

Dilated cardiomyopathy, or DCM, causes heart chamber dilation and therefore poor contraction in the left (or sometimes both) ventricles of the heart. DCM can be caused by a number of factors including myocarditis, alcohol abuse, metabolic disorders, autoimmune disease, familial factors, and genetic factors to name a few.[11] The genes EYA4, MYH6, NEXN, and TNNI3 have all been associated with this type of cardiomyopathy as well (Table 1).[3]

Hypertrophic cardiomyopathy, or HCM, results when the left and/or right ventricles exhibit hypertrophy, or thickening of the heart muscle.[11] HCM can cause arrhythmias and sudden death in individuals. HCM is the most common cause of SCD in young adults under the age of 30. HCM also is the cause for a high proportion of sudden deaths in athletes. The main cause of hypertrophic cardiomyopathy is familial inheritance with an autosomal dominant inheritance pattern. Autosomal dominant inheritance will be expressed in one generation to the next and will not skip generations. Although the disease does not skip generations, the expression in members may vary from severely affected to not so severely affected.[11] In other words the penetrance of the gene defect can vary suggesting other, as yet unknown gene defects may also be involved in the disease pathology. The genes MYH6, NEXN, and TNNI3 are all correlated with this myopathy (Table 1).[3]

Arrhythmogenic right ventricular cardiomyopathy, or ARVC, causes atrophy in the right ventricular muscle and gradually replaces the heart muscle with fatty tissue.[11] The replacement of heart muscle by fat can create bulges and weakness in the cardiac wall and eventually lead to heart failure. ARVC can be found mostly in young adults but is also seen in older age groups. ARVC can cause sudden cardiac death in young people

who would otherwise have a very sound heart. ARVC, like HCM, is inherited in an autosomal dominant pattern.[11]

Restrictive cardiomyopathy, or RCM is the least common cardiomyopathy and causes impaired ventricular filling.[11] RCM can result from increased stiffness in the myocardium (heart muscle) that creates pressure in the ventricle(s). Sometimes RCM can be familial and inherited; in 30% of cases, there is a family history of cardiomyopathy.[11]

Table 1. Sudden cardiac death related genes for current research.

Gene Name	Chromosome	DCM	HCM
EYA4	6	X	
MYH6	14	X	X
TNNI3	19	X	X
NEXN	1	X	X

Methods for Genetic Analysis: Next Generation DNA Sequencing (NGS)

Identifying the mutations or pathogenic variations that lead to sudden cardiac death has become more feasible due to the massive parallel DNA sequencing (MPS) technology (also known as NGS) available today. Using MPS, it is now possible to determine the nucleotide sequence of hundreds of genes in a single sequence analysis and to identify variations in the nucleotide sequence of DNA from SCD patients when compared against the same genes in a normal healthy individual. These gene deviations may represent single nucleotide substitutions (SNPs), small insertions or deletions of nucleotides, or other variations not seen in a "normal" genome. Several NGS platforms

are available commercially at present and each has advantages and disadvantages over competitors.

Ion Torrent

One NGS platform that is used widely for sequencing to identify candidate mutations/variations responsible for cardiomyopathy is known as the Ion Torrent.[4] The Ion Torrent uses ion semiconductor sequencing to determine the order of nucleotides in DNA molecules in a very cost effective way. The technology behind the Ion Torrent platform detects hydrogen ions that are liberated as nucleotides are added stepwise to a replicating DNA template. Every time a nucleotide is added to a growing chain being synthesized from a template DNA molecule being sequenced, a hydrogen ion is released as a byproduct of the reaction and this hydrogen ion is detected and recorded. The Ion Torrent technology couples target DNA sequences (perhaps representing a collection of 50-100 genes known to be associated with cardiac function) to micro beads that are situated in microwells of a metal-oxide semiconductor chip much like computer chips used in electronic devices. Such "DNA chips" can have millions of these microwells, each harboring a particular fragment of DNA to be sequenced. These wells are situated atop an ion-sensitive layer of the DNA chip that is able to detect the release of a hydrogen ion, which lowers the pH of the solution. As individual dNTP solutions are washed over the chip loaded with the sequencing beads harboring DNA templates, a nucleotide in the dNTP solution may be incorporated into the complementary DNA chain as it is synthesized from the sequencing template. When this occurs, a H+ ion will be released and detected. Over the course of a sequencing run, 200-400 flows of dNTPs will

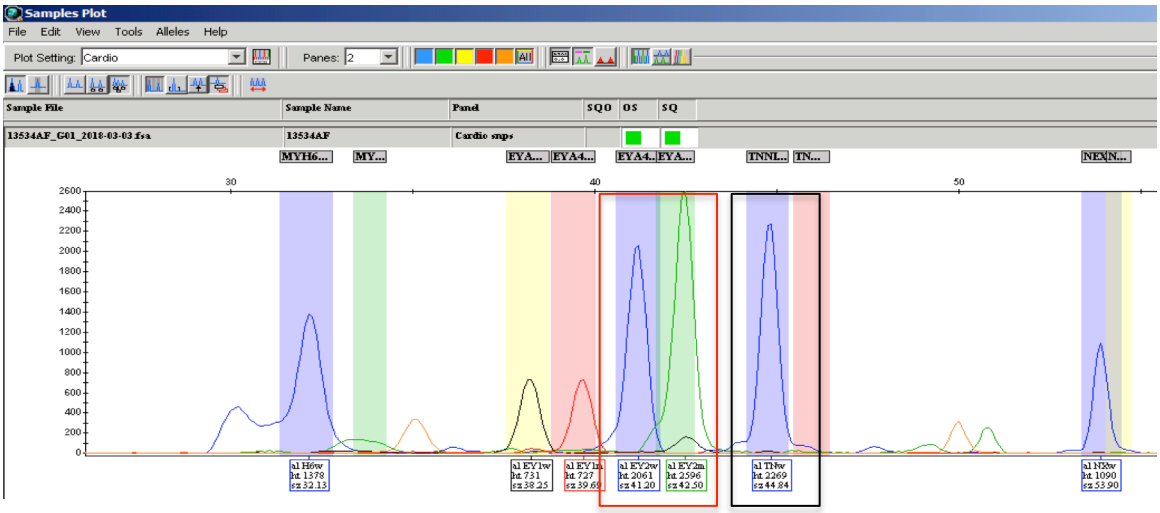
flood the chip sequentially and therefore the length of sequence produced may be 200-400 nucleotides.[12]

SNaPshot™ Method

The SNaPshot™ minisequencing kit from Thermo Fisher Scientific is an efficient tool for determining the nucleotide present at a particular position within a piece of DNA.[13] Related to Sanger sequencing, SNaPshot™ utilizes the chain-terminating characteristic of ddNTPs that are fluorescently labeled. However, rather than a mixture of dNTPs and ddNTPs, SNaPshot™ methods only use ddNTPs. In the method, an oligonucleotide primer is synthesized that terminates one nucleotide 5' to the site of a SNP site (perhaps the SNP responsible for a disease). The DNA to be analyzed is heat denatured to single strands, hybridized to the oligonucleotide primer, and DNA polymerase extends the primer one nucleotide using the ddNTP and the nucleotide incorporated will be complementary to that existing at the SNP site. Each ddNTP is covalently coupled to a different fluorescent dye: ddATP is green, ddGTP is blue, ddCTP is yellow (seen as black for enhanced visibility), and ddTTP is red. The resulting short fluorescent DNA fragment is then separated by size using the 3130XL Genetic Analyzer (Applied Biosystems Inc., Foster City, CA) and the color of the fragment represents the ddNTP incorporated and therefore the nucleotide at that particular SNP site in the genomic DNA. The SNaPshot™ method is robust and extremely efficient for SNP typing in genes for which SNP type polymorphisms are known. SNaPshot™ technology is available in kit form from Applied Biosystems, Inc. and can be used to diagnose SNPs that are associated with genetic disorders. In addition, by careful design of oligonucleotide primers, multiple SNPs, located within a single gene or among several

genes, can be interrogated simultaneously.[13] An example of an electropherogram produced with SNP analysis is shown in Figure 1.

Figure 1. An example of an electropherogram with labeled peaks. Notice how TNNI3 is homozygous (black box) for G and EYA4-1 is heterozygous for C and T (red box).



Post-Mortem Molecular Autopsies

The detection of gene mutations or pathogenic variations in genes associated with cardiac function and possibly leading to SCD often occurs after the death of an affected individual because so often death is the only clue for a possible gene defect existing in the victim. A lack of overt external signs leading to a diagnosis makes it difficult for a medical examiner to conclude SCD as the cause of death. Therefore, molecular genetic studies in the form of a "molecular autopsy" have gained popularity around the world to help confirm a cause of death. A molecular autopsy exploits the molecular technology available today in order to search for an underlying genetic defect that would explain the death. Without this information, the medical examiner is sometimes unable to

conclusively identify the cause of death and also is unable to provide any useful information for the family of the victim who may also be at risk for inheriting the defective gene(s).[10] This is especially true in the case of a victim of SCD where there is no physical or histological data available with which to identify the cause of death. With a convincing diagnosis available to the ME resulting from the molecular autopsy, surviving family members can take appropriate lifestyle and medication changes to minimize their chance of suffering the same SCD fate. Following the diagnosis genetically through the decedent's family is referred to as family cascade testing to identify gene defects and their potential severity in the health of other potential victims in the family.[7]

The search for biomarkers of SCD has occurred for some time. A study by Carvajal-Zarrabal et al. used blood proteins as biomarkers that might be informative to identify the cause of death as SCD.[14] This study used an immunoassay technique post-mortem on 20 victims thought to have SCD. Post-mortem blood samples were collected for the testing and a panel of proteins in the blood was investigated immunologically. One conclusion from the study was that "blood troponin levels may be useful to support a diagnosis of SCD".[14] Molecular autopsies can also help pathologists determine manner of death for many other causes when the toxicological analysis and gross autopsy both come back negative.[15]

Selecting Genes

Literature exists describing possible associations of particular SNPs that might be associated with SCD.[3,16] Genes that might be of diagnostic significance would have to

meet certain criteria: the proposed SNP effect would have to be non-synonymous (i.e. alter the amino acid sequence of the protein, or silence the gene as opposed to being one that does not change the amino acid sequence of the protein) and in using the SNaPshot™ assay, the gene defect would involve a nucleotide sequence substitution rather than a deletion or insertion in the DNA. Tables and lists of SNPs in genes suspected to play a role in sudden cardiac death have been published and are summarized in a database made available to the public database hosted by NCBI (https://www.ncbi.nlm.nih.gov).[16–18] Because of the large number of suspected SNP associations with genes known to be involved with cardiac function, it can be a challenging task to single out a handful for a population study using the SNaPshot™ methods. Population studies are nonetheless needed to convincingly associate nucleotide sequence variations at SNP sites in relevant genes to SCD and thereby advance our ability to identify this condition as the cause of death in the medical examiner setting.[19]

Selecting Primers

To reliably genotype SNP sites in cardiac function genes using SNaPshot™ technology, one must design oligonucleotide primers that will hybridize to genomic DNA and terminate one nucleotide 5' to the SNP site. It is imperative to design a primer that will only find the specific SNP that a researcher is looking for.[19] Attributes of a primer include length and nucleotide sequence, both of which will affect the annealing temperature in the PCR reactions as well as the specificity of the primer to target just one genomic site. Software applications exist on the web that use the human genome database as a source of gene sequence to design primers that meet the required specifications. Ideally, a collection of primers can be designed for the multiplex

SNaPshot™ reaction to interrogate simultaneously or multiplex SNP sites either within the same gene or in several.[19]

National Database

Annotated genes and their associated polymorphisms are catalogued for numerous species, including humans from several resources. The National Center for Biotechnology Information, or NCBI (also called GenBank) is one such resource with DNA sequence information as well as some bioinformatics tools with which to study that sequence.[19] The online database allows researchers to search freely and easily for biological information concerning genes and their variations. Information from this database was used to identify candidate SNPs of potential importance to SCD and to identify specific nucleotide sequences to target with the primers used in the SNaPshot™ assay.

U.S. National Library of Medicine (NLM)

Another resource for information pertinent to the study reported here is the National Library of Medicine. The United States National Library of Medicine, or NLM, maintains a sizeable database of print and electronic health information that is available to any individual.[20] The library is the largest biomedical library in the world and is also one of the largest providers of digital content from the federal government.[20] The free digital collection contains the identities and functions of numerous human genes and, where possible, gene defects responsible for health related phenotypes associated with those genes. In addition to the brief description of a gene, references to published works are provided in the database for additional information. The NLM database was of great

value in deciding which genes to include in a limited study of SNP polymorphisms in genes known to be involved in cardiac function.[20]

Genes of interest to SCD

EYA4

The gene EYA4 encodes a protein that is a putative oncogene linked to cancerous outcomes in previous research.[20] EYA4 controls DNA repair and apoptosis, or controlled death of a cell, and these functions occur naturally in a living system. Researchers have found that nucleotide sequence variations in this gene correlate with a cardiomyopathy causing ventricular dilation and impaired systolic function. Dilated cardiomyopathy normally begins in the left ventricle and causes the chamber to stretch until it becomes too thin to effectively pump blood.[17] Therefore, the gene EYA4's mutations/variations are excellent candidates for our research with SCD.[20]

MYH6

MYH6 is an important cardiac gene directing the production of cardiac alpha (α)-myosin heavy chain protein.[20] The protein is one subunit of a larger protein in heart muscle cells called type II myosin. These muscle proteins help convert energy into a mechanical force that is necessary for the heart to function and pump blood to the rest of the body. Type II myosin helps to create the sarcomeres, the main cellular structures involved in the contraction of cardiac muscle. Genetic mutations and pathogenic variations in this gene have been associated with dilated cardiomyopathy and premature deaths due to cardiac function failure. MYH6's correlation to cardiomyopathy made it a good candidate to include in our study.[20]

TNNI3

TNNI3 is a gene that encodes the protein cardiac troponin I.[20] Troponin I is a protein found only in the heart and is part of a troponin protein complex in the muscle cells connected with sarcomeres. The troponin protein helps to regulate and coordinate the contractions of the heart. Sequence variations in the TNNI3 gene have been found to cause a variety of cardiomyopathies: hypertrophic cardiomyopathy (thickening of the heart), restrictive cardiomyopathy (stiffening of cardiac muscle), and dilated cardiomyopathy. These diseases can impair the heart function and hurt blood flow, which can lead to a premature cardiac death.[20,21]

NEXN

The human gene NEXN encodes for a protein reportedly known to help with cell adhesion and migration.[20] SNP-type variations in NEXN can cause pathologic conditions such as hypertrophic cardiomyopathy and dilated cardiomyopathy. Both of these cardiomyopathies are hereditary. Both of these disorders can also be worsened and provoked with intense exercise. Hypertrophic cardiomyopathy can cause heart collapse, palpitations, and severe chest pain.[20]

Summary

Sudden cardiac death is an unexpected, tragic event that typically results in loss of life and shock for surviving family members. SCD is likely a heterogenetic condition that may result from mutations or nucleotide sequence variations in a single gene, or from mutations/variations in a collection of genes that together compromise cardiac function and put the host at risk for cardiac failure under conditions of physical exertion. To

better understand SCD, candidate SNPs that may affect cardiac function can be studied in a population of presumably normal individuals as well as among victims of unexplained death possibly linked to defects in cardiac function. In this study, a small collection of SNP sites residing in several genes associated with cardiac function were studied using a single nucleotide sequencing technology known as SNaPshot™. While our study involved only a small sampling of genes and SNPs, we nonetheless anticipate adding to an understanding of SCD.

CHAPTER III

METHODOLOGY

Single nucleotide polymorphisms, or SNPs, that may be correlated with sudden cardiac death (SCD) were identified using a method called SNaPshot™. Through the use of this method, it was possible to genotype 5 SNPs residing in 4 genes known to be associated with cardiac function and to attempt to correlate those genotypes with the cause of death in a cohort of DNA samples obtained from the Medical Examiner for the State of Oklahoma. Many people fall victim to SCD in the United States each year and this research could help identify the mutations in cardiac function genes that underlie these untimely deaths.

Choosing the Genes and SNPs

The genes and associated SNPs were chosen based upon the scientific literature available in this area that suggests these SNP/gene combinations are correlated with an elevated risk for SCD.[3,5,10,16,18,20] The genome databases available to researchers contain the locations of SNPs within most human genes and also summarize studies that implicate those SNPs with disease. In the cases of SNPs within genes associated with SCD, the pathogenicity of those SNPs is also suggested in some of the databases

available. Those resources were used to choose the genes and SNPs for this study. Once the candidate SNP was chosen, the National Center of Biotechnology Information (NCBI) website was used to obtain the nucleotide sequence of the DNA flanking the SNP and thus to identify primer binding sites used to amplify a DNA fragment harboring the SNP. The individual webpages of the SNPs have sections that contain the nucleic acid sequences of the genes along with the highlighted SNPs contained within the genes. Right clicking the SNP and clicking "Reveal in Sequence View" shows a sequence that can be copied and pasted into software that will allow for the creation of primers for the SNaPshot™ assay to interrogate the SNP site. This process was followed for each of the five SNPs chosen for this research: *EYA4, EYA4-2, MYH6, TNNI3*, and *NEXN*. It is noted that the SNPs in EYA4 and EYA4-2 reside in different areas of the same gene and both have been reported to be "pathogenic".[19]

Creation of Primers

Primers are needed to specify areas on the DNA strand that harbors a SNP that may be associated with SCD. The primers for the four genes and five SNPs were designed using the *Primer 3: WWW Primer Tool* available online from the Whitehead Institute for Biomedical Research (http://biotools.umassmed.edu/bioapps/primer3_www.cgi). This primer design tool allows the user to design the 5' and 3' primers that will direct the amplification of a DNA fragment, design a hybridization probe for that amplicon, and determine the melting temperature range (Tm) of the primers. The basepair size and Tm of the primers were the most important considerations in primer design for this study. As will be discussed below, each SNP primer for the SNaPshot™ assay was designed with a different basepair

length in order to allow for differentiation of the various SNPs when using capillary electrophoresis. Additionally, in order to maintain a consistent Tm for the primers to allow multiplexing of the SNaPshot™ reaction, adenosine residues were sometimes added to a primer in differing numbers. Adding adenosines residues to a primer did not affect the Tm (because the A residues did not hybridize to the genomic DNA template) but did allow for engineering a primer of the desired size. Thus, adenosine tails allow a PCR amplicon to become longer without affecting the Tm or the specificity of primer hybridization. A total of three primers were made for each SNP or 15 primers altogether: a forward and reverse primer for first round of PCR in which a small genomic DNA segment harboring the SNP is amplified, and a third primer terminating one nucleotide upstream from the SNP site that is used to detect the nucleotide at that site using the SNaPshot™ reaction. Table 2 lists the primers created for this study. The table gives the gene name, amplicon size, direction of amplification, and the 5'-3' sequence. After the primers were chosen, they were produced commercially (Invitrogen, Carlsbad, CA).

Table 2. Shown below are the primers being used in this research. The gene name, length, direction, and sequence are all included.

Gene	Primer size (bp)	Direction	Sequence 5'-3'
EYA4	20	Forward	TCTCTCTCCCATCCCTCCTT
EYA4	20	Reverse	TGTAGAGCCAGAGGCATTGA
EYA4	40	SNP	AAAAAAAAAAAACTGTGTTCTTTAGCCGGAGATC
EYA4-2	22	Forward	TCAGAAACAAATGGGGCTGAGT
EYA4-2	20	Reverse	ACGCTCCATACGTTGATGCT
EYA4-2	34	SNP	AAAAAAAATTTCAGGATTATCCATCCTATACAGCCTTT
MYH6	20	Forward	GCTGGAGTGAACAGGGACAT
MYH6	20	Reverse	ACTGATAGGCGTTGTCGGAG
MYH6	23	SNP	AAAAAAAAAACATGCTGACCTTCCTGCAC
TNNI3	21	Forward	GGTCTTTATCCTGAAGCCCCG
TNNI3	20	Reverse	AGAAACCTCGCATCCTTGGG
TNNI3	46	SNP	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAGAGGCGGAGGAGCG
NEXN	24	Forward	AGCAGTCAACAATAAAGGATCTGC
NEXN	21	Reverse	CCGCCACAGAAAAGATGGATG
NEXN	53	SNP	ATCAGCTAAAAAAGAGAAGAAGAAAAGTGATTTTAAGGAAAAAAAAATTAATA

Sample Collection

DNA samples were obtained from two different population sources: OSU-CHS Human Identity Testing lab and the Tulsa Medical Examiner's Office. A total of 56 extracted DNA samples that were deemed "normal" were obtained from the Human

Identity Testing laboratory at OSU-CHS. "Normal" in this study was defined as an individual surviving to age 40 or above. Sudden unexplained deaths normally occur in individuals under the age of 35 or 40.[22] The samples were labeled with a coding system that uses numbers and letters provided by the RT lab. Since these samples were provided as anonymous no Institutional Review Board, or IRB, approval was needed for the study. The Tulsa Medical Examiner's Office provided bloodstains from victims suspected of dying of cardiac related conditions. The ME placed blood samples on blood collection cards prior to sending them to the lab; the cards were stored at room temperature at the OSU-CHS RT lab. No IRB approval was needed for these samples since the cards were not labeled with names or other identifying information and the human subjects were deceased.

DNA Extraction

An organic extraction method routinely used by the Human Identity Testing laboratory was used to extract and recover DNA from the ME samples. The first step in DNA extraction was to prepare the bloodstain cuttings from the bloodstain cards supplied by the medical examiner. A cutting of about 1/3 of the sample was placed into a 0.6 mL centrifuge tube that was labeled with the sample number and date. After all the sample cuttings were made, fresh extraction buffer was prepared using 20 µL of 20-mg/ml proteinase K, 25 µL of 20% SDS, and 955 µL of TNE buffer (TNE is 10mM Tris-Cl, pH 8.0 containing 0.2 M NaCl and 0.1 mM EDTA). 300 µL of this buffer was added to each tube of cuttings. The tubes were mixed using a vortex mixer and placed in a 65°C heat block for 1 hour. After 1 hour, the tubes were removed from heat and placed into a 1.5 mL tubes that were labeled in the same way as the original tube. A hole was punched in

the bottom of the small tube before placement into the larger tube and, upon centrifugation; the liquid extract containing the DNA was collected into the larger tube. The smaller tube containing the stain card remnants was discarded while the larger tube was kept for the next steps. To each sample the following organic solvent mixture was added: 270 μL phenol and 30 μL chloroform/isoamyl mix (9:1). Each tube was vortexed briefly and then centrifuged again at 10 000 rpm for 1 minute. The emulsion created by vortex mixing is noticeably separated into an organic layer with phenol in the bottom phase and an aqueous phase containing the DNA in the top layer. The top, aqueous phase could then be pipetted into a new tube, each labeled with the sample name. There should be approximately 170 μL transferred. Another 300 μL of the chloroform/isoamyl mix (above) was added to each tube that was then briefly mixed using a vortex mixer and then centrifuged at 10,000 rpm for 1 minute. Again the top aqueous layer was transferred into a new 1.8mL tube. DNA was then purified from the aqueous phase and concentrated by binding to a silica spin column (Zymo Research, Orange, CA). This kit contains columns, wash buffer, and binding buffer to recover purified DNA from the initial extract. The Zymo column was placed in a 1.8mL tube and the sample and binding buffer (about 400 μL) were added. The tubes were then centrifuged at 10,000 rpm for 1 minute. Flow-through from the column in the bottom of the tube was discarded. Next, 200 μL of the wash buffer was added to the column and then it was centrifuged again for 10 000 rpm for 1 minute. The flow-through was not discarded after this step. Another 200 μL of wash buffer was added to the tube and centrifugation was repeated. All flow-through from this step and the step above was now discarded. In a new tube, the sample was eluted by adding 15 μL of TE^{-4} (TE^{-4} is a 10mM Tris-Cl, pH 8.0 containing 0.1mM

EDTA) on top of the spin column, incubated 1 minute at room temperature, and then centrifuged at 10,000 rpm for 30 seconds. The flow-through was retained and another 15 μL TE^{-4} was added to the silica column and centrifuged again at 10,000 rpm for 30 seconds. The retained flow-through contains the eluted DNA in a total volume of about 30 μL. The samples were stored in a -20°C freezer.

Preparing Primers

The primers received from Invitrogen are dry and must be reconstituted before use. Primers were rehydrated in TE^{-4} to a concentration of 100 μM. In order to dissolve the primers, TE^{-4} buffer was first added to each individual tube of the primers. The amount of TE buffer that must be added to each primer was determined using the following formula:

$$[\text{concentration}_1][\text{volume}_1]=[\text{concentration}_2][\text{volume}_2]$$

or

$$V_1=(C_2*V_2)/C_1$$

In this formula C stands for concentration, V for volume, 1 for initial, and 2 for final or desired result. There is also an easier way to determine this number. When the primers arrive, each is accompanied with its own sheet titled "Invitrogen Custom Primers-*Certificate of Analysis*." This sheet states the nano-moles (nmoles) of the primer; the amount varies for each primer depending upon each synthesis reaction. That amount was then multiplied by 10, which is the amount of TE^{-4} buffer (in μL) that will be needed to produce primer stocks of 100 μM to the proper concentration. Once all the primers were

resuspended to a concentration of 100 μM, 5 μL of the forward and reverse primers for each SNP was added to 40 μL of TE^{-4} in a 1.5 mL centrifuge tube to prepare a total of 50 μL of 10X primer stock containing both the forward and reverse primers directing the amplification of the genomic DNA fragment harboring the SNP interrogated using SNaPshot™. It was also possible to prepare a 10X primer concentrate containing the forward and revers primers for all 5 SNPs such that the SNaPshot™ templates could be produced by multiplex PCR. The final 10 μM stock was then used as the "working concentrate" for PCR or SNaPshot™ reactions. The 100 μM stocks of each primer were stored at minus 20°C.

Amplification of Genomic DNA Fragment Harboring a SNP

PCR amplifies a small region of the genomic DNA template to produce an amplicon with the SNP in the middle or near the middle of the amplicon. Genomic DNA samples were taken from the freezer and thawed along with the forward/reverse primers. Genomic DNA samples were amplified with PCR in 12.5 μL reactions composed of 1.25 μL primer mix, 7.5 μL GoTaq DNA polymerase (Promega Corp., Madison, WI), 3.25 μL nuclease free water, and 1 μL of genomic DNA. A master mix minus the DNA was created and aliquoted into the number of PCR reactions tubes desired and then 1 μL DNA was added to each sample. Once all the samples and a negative control have been prepared, they were amplified in an ABI9700 thermocycler (Applied Biosystems Inc., Carlsbad, CA). The cycling program was: 98.0°C for 2 minutes for initial denaturation of the genomic DNA template followed by 30 cycles of 98°C for 10 seconds, 60°C for 1 minute, 72°C for 30 seconds. The program also had a single cycle of 60°C for 20 minutes for completion of the elongation and terminal adenosine addition.

Following amplification of the DNA fragments to be used for the SNaPshot™ reaction, remaining PCR primers and dNTPs were destroyed by exposing amplification products to digestion with ExoSap (Affymetrix USB, Waltham, MA) which enzymatically degrades oligonucleotides and dephosphorylates dNTPs so that they cannot participate in further PCR amplification.

Agarose gel electrophoresis

The amount of amplified DNA produced in the first PCR reaction was roughly quantitated using agarose gel electrophoresis to ensure there was adequate DNA present in amplicons to obtain results from the SNaPshot™ reaction. A 1% (v/v) agarose gel was prepared using Molecular Biology Grade Agarose LE powder (Phenix, Candler, NC), 1x TAE buffer, and GelRed dye (Phenix, Candler, NC) diluted 10,000x in the final liquid agarose gel mix when it was cool enough to touch and just before pouring the molten agarose into the gel mold. Five microliters of each amplified sample was mixed with 1 μL of 5x loading dye (mix of bromphenol blue, xylene cyanol, and TAE buffer) in a centrifuge tube. The prepared mix (6 μL total) was then pipetted into a sample well in the gel. Once all the samples were loaded into the gel, the samples were electrophoresed at 100 Volts for a period of about 30 minutes or until the bromophenol blue tracking dye moved down the gel a distance of about 8-10 cm. Once electrophoresis was complete the gel was removed from the electrophoresis apparatus and placed in the InGenious LHR Gel Imaging System by Syngene for analysis using a program called GeneSnap (SynGene, Bangalore, India). The gel was illuminated on a UV light box and electrophoresis results were captured with a digital camera. Sample results are shown in Figure 2.

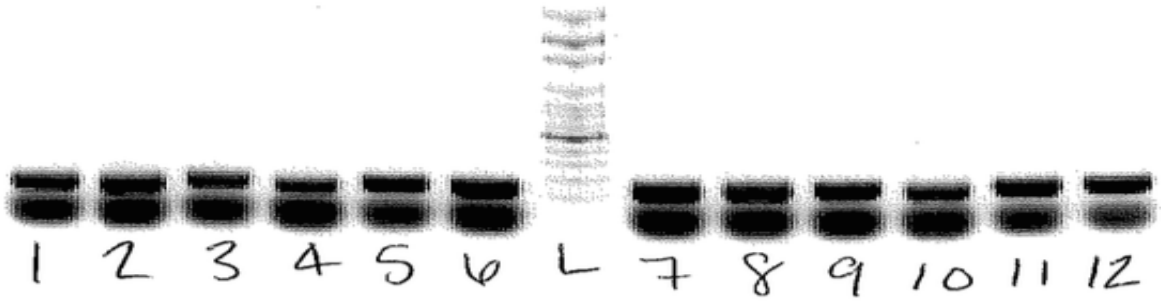


Figure 2. An example of agaraose gel electrophoresis results. The middle column labeled "L" is the size ladder. The other wells contain amplicons produced by PCR from genomic DNA samples. The two black bars visible are the template DNA (top bar) and the leftover components (bottom bar).

ExoSap-IT® PCR Product Clean Up

The ExoSap procedure (Affymetrix USB, Waltham, MA) is used to "clean up" the samples and remove the leftover primers and dNTPs with a one-step enzymatic reaction. The reaction degrades residual PCR primers from the first amplification step and also destroys dNTPs that remain, preventing these reagents from participating in the SNaPshot™ reaction. To remove unincorporated primers and dNTPs, 5 µL of PCR product was mixed with 2 µL of the ExoSAP-IT enzyme mixture. Each sample tube was labeled with the name, date, and type of sample, such as: 13170M 11/10/17 EXO. The tubes were briefly mixed and then briefly centrifuged. Sample tubes were then incubated at 37°C for 1 hour followed by a 15 minute incubation at 80°C to inactivate the enzyme mixture. Once this reaction was complete, the samples were ready for single nucleotide sequencing of the SNP site using the SNaPshot™ assay.

SNaPshot™ PCR

The ABI Prism® SNaPshot™ Multiplex Kit Protocol from Applied Biosystems, Inc. (Carlsbad, CA) was used to determine the nucleotide present at the SNP site located within each of the PCR amplicons produced in the step described above. The SNaPshot™ reaction can interrogate up to 10 individual SNPs in a single capillary electrophoresis injection (manual provided with the SNaPshot™ kit from Applied Biosystems, Inc., Carlsbad, CA). The SNaPshot™ kit (Thermo Fisher Scientific, Waltham, MA) contains a reaction mixture composed of buffer and dideoxynucleotide triphosphates like those used in Sanger DNA sequencing reactions. Each ddNTP is labeled with a different colored fluorescent dye. Being ddNTPs, they can become incorporated into a growing DNA chain, but the growth of the replicating DNA chain terminates once a ddNTP is incorporated. Because the primer used to interrogate a SNP site terminates one nucleotide upstream from the site, a single ddNTP will become incorporated at the SNP site and the color of the resulting DNA fragment will specify the complementary nucleotide present at the SNP site in the DNA template. Recall that SNaPshot™ primers were designed to be different lengths. Therefore, following completion of the SNaPshot™ reaction, products are separated using capillary electrophoresis and the various lengths of the primers allows an investigator to trace each fluorescent peak in an electropherogram to a SNP interrogated with the procedure; the color of the peak will identify that nucleotide at the SNP site. SNaPshot™ reactions were prepared by mixing 1 μL of each SNP primer (5 total) with 5 μL of nuclease free water to create 10 μL of a 10 μM mix of primers. Next, 1 μL of the 10 μM primers were mixed with 49 μL of nuclease free water to create a 50 μL primer mix at a concentration of 0.2

μM. Samples analyzed using SNaPshot™ included a positive control (supplied with the kit), and a negative control, as well as the products produced from the genomic DNA samples. The 0.2 mL micro-centrifuge tubes were labeled according to name of sample, type of sample, and the date. The mixtures used to prepare the samples are shown in Table 3 below.

Table 3. The components of SNaPshot™ reactions.

	SNaPshot™ RX Mix (μL)	Pooled SNaPshot™ primers (μL)	DNA Template (PCR products) (μL)	Nuclease free water (μL)	Total Amount (μL)
Sample	2.5	0.5	2	-	5
Negative Control	2.5	0.5	-	2	5
Positive Control	2.5	0.5	1	1	5

Once SNaPshot™ reactions were created, samples were amplified using PCR. The thermocycling program was as follows: 25 cycles of 96°C for 10 seconds, 50°C for 5 seconds, 60°C for 30 seconds, followed by a cycle of 4°C for an infinite amount of time until the samples were removed. Once amplification was complete, the samples were removed and treated with shrimp alkaline phosphatase (SAP) treatment to destroy any remaining unincorporated fluorescent ddNTPs that could interfere with data analysis.

SAP Treatment

The SAP treatment works much like the ExoSAP-IT; it helps to clean up the sample and prepare it for capillary electrophoresis. This procedure used the samples from the SNaPshot™ reactions that were treated with SAP in a Shrimp Alkaline Phosphatase kit (USB®, Waltham, MA). SAP reactions followed the Affymetrix USB procedure for dephosphorylation of ddNTPs: 5 µL of sample that resulted from the SNaPshot™ reaction mix was mixed with 1 µL SAP and 0.6 µL of SAP 10X Buffer. The tubes were mixed and incubated for 60 minutes at 37° C and then 15 minutes at 65°C to inactivate further SAP activity.

Capillary Electrophoresis and Data Analysis

Fluorescent SNaPshot™ products were separated by size using the 3130xl Genetic Analyzer (Applied Biosystems Inc., Foster City, CA), a capillary electrophoresis

Table 4. Parameters for 3130xl Genetic Analyzer from ABI sequencing.

Name of Plate	**Application**	**Owner**	**User**	**Plate Type**		
Plate Name	GeneMapper -ABI3130XL	OSU	TAM (user's initials)	96-Well		
Sample Name	**Sample Type**	**Size Standard**	**Panel**	**Analysis Method**	**Results**	**Instrument Profile**
Sample Name	Sample	Liz 120	Cardio SNPSs	Cardio	Snapshot	Snapshot_pop 7_25s

platform that separates DNA fragments by size and detects them using fluorescence.

Data were collected using data collection software supplied with the instrument (Version 3.0, ThermoFisher Scientific). Electrophoresis run parameters are listed in Table 4 above. Once the run was complete, GeneMapper analysis software (ver. 3.2, ABI Inc., Foster City, CA) was used to convert raw data into viewable results.

Reading the Electropherogram Results

An example of SNaPshot™ results produced using the Gene Mapper software is shown in Figure 3. The different colors define the ddNTP incorporated at each SNP site:

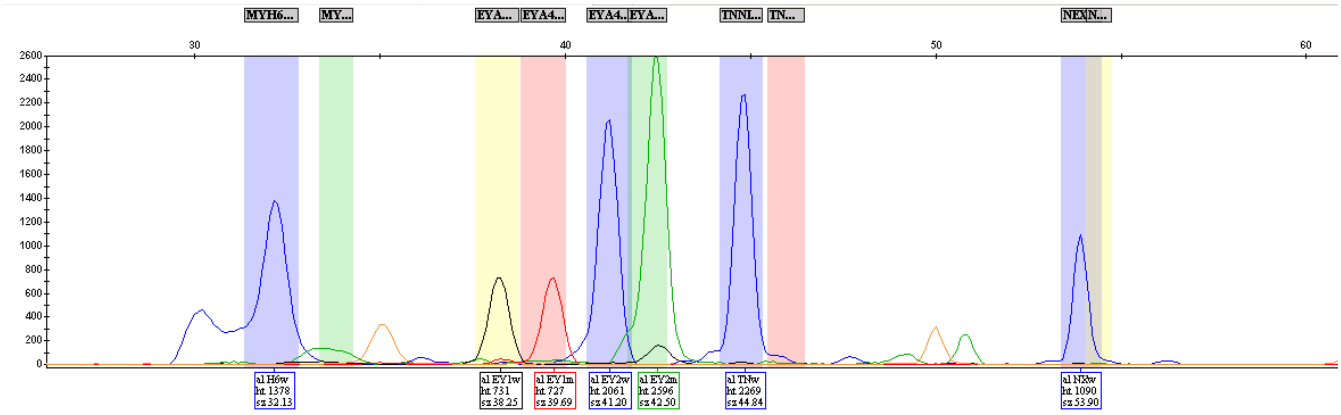


Figure 3. Example of electropherogram results.

Yellow (shown as black for visibility) represents an incorporated cytosine (C), blue represents incorporated guanine (G), red represents incorporated thymine (T), and green represents incorporated adenine (A). Note that for genes in which the SNP exhibits a heterozygous genotype, the labeled SNaPshot™ products migrate with slightly different apparent molecular weights. Because the size of the SNaPshot™ fragments is so small, the chemical characteristics of the different fluorescent dyes affects electrophoretic migration slightly during capillary electrophoresis and suggests the fragments may be different size. For example, the SNP site in the *EYA4-1* gene exhibits a red and a black

peak, which means that this SNP is heterozygous for T and C. The chemical characteristics of the red and yellow (yellow is shown as black) fluorescent dyes affect electrophoretic migration and the heights of the fluorescent peaks reflect the fluorescence incorporated into the primer and thus the efficiency of the SNaPshot™ reaction for that particular SNP. By knowing the nucleotide incorporated into the SNaPshot™ product one can deduce the nucleotide that is present at the SNP site in the template amplicon and assess its possible association with SCD.

Statistical Analysis

Statistical analysis for the genotype data collected asked whether or not any differences observed in the frequency of SNP genotypes in normal or ME samples were significant. A contingency table was produced to determine relationships between several variables. The variables being compared were the "wildtype" (i.e. non-pathogenic) verses "pathogenic" (i.e. perhaps associated with compromised gene function) genotype frequencies overall between the two population groups as well as between the individual SNP genotypes in individual genes. The p-value or probability value of <0.05 was chosen as the cut-off to test the significance, since that number is generally accepted as the boundary of scientific significance.[23] The SNP sites were deemed significantly different or not in the contingency table if the lowercase letter after the frequency result for a given SNP matches that of another SNP, then the two sites are NOT significantly different. If the two lowercase letters differ among SNP sites, then the proportion of wild type and pathogenic SNPs ARE significantly different in the normal versus ME population group. A contingency table was created to compare the

differences in frequencies between the wildtype nucleotides and pathogenic SNPs in the normal and ME population groups.

Population studies based upon the Hardy-Weinberg principle and linkage analysis were also applied to allele and genotype counts in the two population groups. To begin the process, an allele frequency database was created. The allele counts for the normal population were used to create an allele frequency databases for each SNP and these frequencies were used subsequently used to compare the expected versus observed genotype results in normal and ME samples to check for population equilibrium. The Hardy-Weinberg equation p2+2pq+q2=1 will allow genotype frequencies for the different genotypes possible for a genetic marker to be accurately estimated if the population is in equilibrium. The frequencies of the different allele forms of SNPs associated with EYA4, EYA4-2, TNNI3, and NEXN genes were sufficiently high to suggest these SNPs represent di-allelic genetic markers. If these di-allelic markers are in equilibrium in the normal population, deviations from equilibrium in the ME population group would suggest an association of a SNP with SCD. Moreover, if the different genes are unlinked in the normal population group, any sign of linkage in the ME group would be further support for an association of particular SNPs in our gene panel with SCD. Ultimately, chi square analysis was used to investigate any possible correlation of SNP genotypes with SCD. Consideration was also given to an analysis of linkage among three genes in the panel as a cause of death, but this was not pursued since the number of ME samples was small.

CHAPTER IV

RESULTS AND INTERPRETATION

Sudden Cardiac Death (SCD)

Sudden cardiac death (SCD) occurs when the body experiences a sudden loss of blood due to the heart failing.[1] While technological advances allow researchers to identify and examine variability in human DNA more efficiently than ever before, researching the hundreds of SNPs located within genes that have been linked to SCD in this study would have been prohibitively expensive. Therefore, 4 genes harboring 5 SNPs believed to be associated with compromised cardiac function were chosen for genotyping in a normal population and in a population of individuals thought to have died from cardiac episodes evaluated post-mortem by the Office of the Chief Medical Examiner for the State of Oklahoma. The genes studied included: EYA4, MYH6, TNNI3, and NEXN.

Five different SNPs were researched in this study in the four genes previously mentioned. The genes were chosen for this study in two ways: In a unpublished study performed by Dr. Di Shen (a visiting research scientist from Beijing, China), several DNA samples obtained from the Office of the Chief Medical Examiner were subjected to

massively parallel DNA sequencing (MPS) of a library of 58 genes known to be involved in cardiac function, some of which have been implicated in SCD. Results from the sequencing identified numerous SNPs contained within the 58 genes sequenced, some of which were chosen for this study (NEXN, EYA4, and TNNI3). Also considered in gene/SNP selection was the literature that exists concerning SCD and the confidence values reported for associations of particular SNPs with pathology that may be causative of SCD. The SNPs associated with the NEXN, EYA4, TNNI3, and MYH6 genes among those suggested in the literature good candidates for this.

The samples subjected to SNaPshot™ genotyping were acquired from the Office of the Chief Medical Examiner for the State of Oklahoma and from the Oklahoma State University Center for Health Sciences' Human Identity Testing laboratory. The twenty-eight medical examiner (ME) samples were provided as bloodstains from individuals who may have died due to SCD or other cardiac conditions. One sample provided by the ME's office was thought to have died from SCD with high confidence(Sample 1 in Table 7), since it was obtained from an infant who died suddenly for no apparent cause. This sample exhibited a genotype that contained pathogenic SNPs at each of the 5 SNP sites in the four genes that were used in our study (Table 5).

The 56 DNA samples from the OSU Human Identity Testing laboratory had already been extracted and thus, with rare exception, DNA extraction was not required. The OSU samples were from individuals past the age of 40. It was presumed that if the people did not have a cardiac episode before age 40, then they could legitimately be deemed at low to no risk for SCD since this diseases generally strikes relatively early in life. These samples were thus considered to represent the normal population group.

Genomic DNA samples obtained from normal donors and those obtained from the Medical Examiner were amplified using PCR to produce the template for the SNaPshot™ assay. However, prior to performing the SNaPshot™ assay, the amplicons produced were assessed for quality by agarose gel electrophoresis. Products suitable for further analysis were visualized in a stained gel as a single band of about 150 bp (basepairs) that migrated just behind the PCR primers used for amplification (Figure 4).

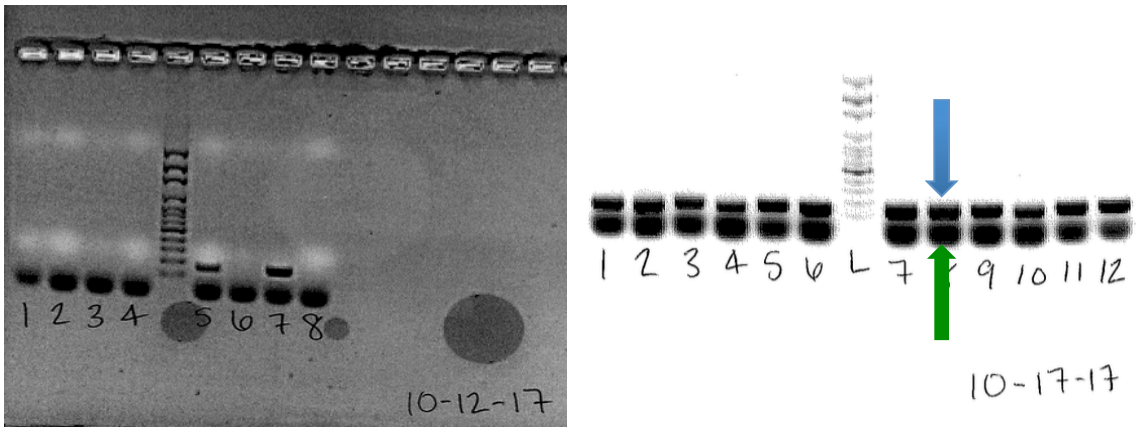


Figure 4. Amplicon produced by PCR using 5 sets of primers to amplify templates for SNaPshot™ reactions. Products were separated by electrophoresis in 0.7% agarose gel electrophoresis, with GelRed and 1xTAE buffer. These PCR amplicon products served as template for the SNaPshot™ assay. Molecular weight standard models are shown in between the 4/5 lanes and the lane labeled "L". The first black band is desirable DNA product (example shown by blue arrow) and the second (green arrow) is unincorporated nucleotides and/or primers.

SNaPshot™ Results

The single nucleotide polymorphisms (SNP) were characterized using SNaPshot™ methodology. SNP sites were researched using the National Center for Biotechnology Information (NCBI) database. This database provides the gene identity and description, the chromosomal location, any aliases, and other useful information about any particular gene and polymorphisms or mutations within that gene. The wild type and nucleotide variants for the five SNPs chosen for this study are listed in Table 5. below. NCBI lists all five of the variant genotypes as potentially pathogenic. A SNP, although changing the DNA sequence, is not always harmful to the host. However, if a SNP is defined as pathogenic, then it is linked to a disease and has the potential to cause or contribute to said disease. A SNP may alter the amino acid sequence of a gene product and result in the loss of function for that protein. Such a SNP would therefore be designated pathogenic and could contribute to SCD.

Table 5. The wildtypes and variants for the five loci tested in this research.

Locus	Pos GRCh37	dbSNP ID	Wildtype	Variant
EYA4	133849966	rs3734279	C	T
EYA4-2	133789728	rs9493627	G	A
MYH6	23874889	rs140596256	G	A
TNNI3	55667647	rs3729711	G	T
NEXN	78408536	rs3767028	G	C

SNaPshot™ results are visualized using capillary electrophoresis (CE). Following the SNaPshotTM reaction, the oligonucleotides that have been extended by

one base are separated from one another electrophoretically using CE with the 3130XL Genetic Analyzer. A successful SNaPshot™ reaction was scored if there was adequate fluorescence associated with an amplicon (i.e. the fluorescence met or exceeded a threshold value) and if the fluorescent peak was of expected size. Heterozygosity was determined if peaks of two colors were detected at a particular SNP site and if the peak heights were balanced, had a characteristic triangular peak shape, and if fluorescence exceeded 50 rfu. Examples of true and false SNP genotypes are shown in Figures 5a and 5b respectively. The arrow in Figure 5a identifies the "A" peak for the SNP associated with the gene MYH6. We would conclude that this is not a true A-allele (i.e. the incorporation of adenosine at this SNP site is an artifact) since the peak does not have a true, characteristic peak shape and also the peak height is severely imbalanced with the G-allele to the left of the artifact. In Figure 5a, all SNPs among the 5 genes analyzed exhibit homozygosisty. Figure 5b shows results in which the EYA4-2 gene exhibits heterozygosity with an A residue incorporated to complement a T residue on one chromosome whereas the other chromosome harbors a C residue (because of the incorporation of a G residue in the SNaPshot™ assay). Peak heights are balanced reasonably well and each peak shows the expected characteristics shape.

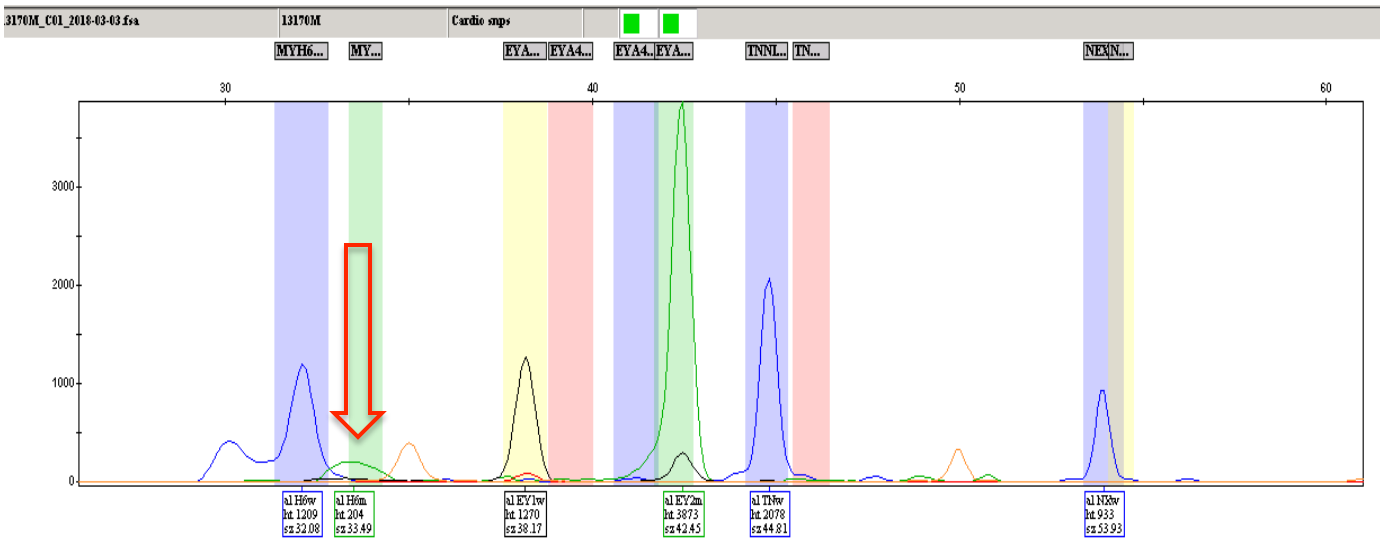


Figure5a. An electropherogram of a normal sample shown from program GeneMapper using ABI 3130 for sequencing. SNP sites MYH6, EYA4, EYA4-2, TNNIE3, and NEXN are all on the panel. MYH6 does not have a true peak for nucleotide A due to its unusual shape, as shown by the red arrow.

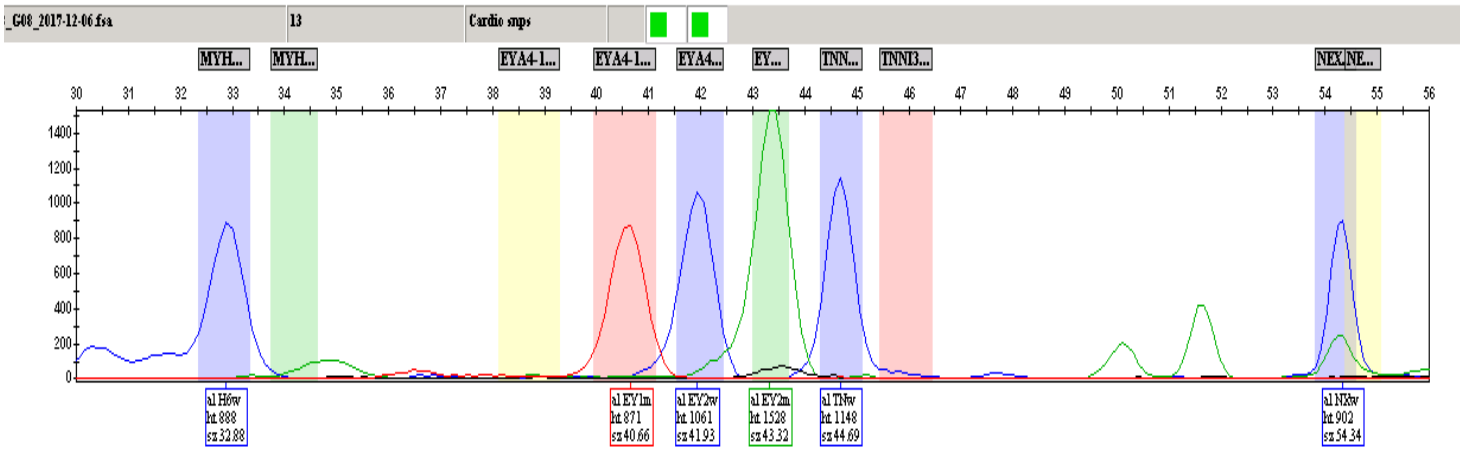


Figure 5b. An example of a successful electropherogram. The adenosine peak for MYH6 is not called due to its imbalance and unusual shape. The genotype of the EYA4-2 locus is a valid heterozygote because of the reasonably balanced blue and green peaks observed.

SNaPshot™ Genotyping Results

SNaPshot™ genotypes for the five genes among the normal and ME samples are summarized in Tables 6 and 7. Some samples yielded results that were questionable for

various reasons and DNA had to be re-extracted and the SNaPshotTM reaction repeated. In total, there were 94 repetitions of SNaPshot™ assays and the results shown in Tables 6 and 7 were all-reproducible and represent the consensus genotypes. The presence of a "wildtype" nucleotide at a SNP position is denoted in black typeface whereas the presence of a "variant" nucleotide at that position is denoted in red typeface. Variants may represent true mutations that are pathologic with respect to the functioning of the gene in question. However, variants by themselves may not seriously affect gene function and may simply represent alternative alleles for a di-allelic genetic marker.

Table 6. The SNP results for “Normal Samples”. The red colored nucleotides represent variant nucleotides while those in black represent the wild type.

Sample Name	MYH6	EYA4	EYA4-2	TNNI3	NEXN
	G to A	C to T	G to A	G to T	G to C
SAM	GG	CC	GG	GG	GG
13170M	GG	CC	AA	GG	GG
13493AV	GG	CC	GG	GG	CC
13531AF	GG	CT	GG	GG	GG
13532M	GG	TT	AA	GG	GG
13534AF	GG	CT	GA	GG	GG
13550M	GG	CC	AA	GG	GG
13607AF	GG	CC	GG	GG	GG
13624M	GG	CT	GG	GG	GC
13649M	GG	TT	GA	GG	GG
13654AF	GG	CT	GG	GG	GG
13682M	GG	CT	GA	GG	GC
13695M	GG	CC	GA	GG	GG
13697M	GG	CC	GG	GG	GG
13701AF	GG	CT	GA	GG	GG
13729M	GG	CT	GG	GG	GG
13505M	GG	CT	GG	GG	GG
13492AF	GG	CT	GA	GG	GG
13493M	GG	CC	GG	GG	GC
13505M	GG	CT	GG	GG	GG
13533M	GG	CC	GA	GG	GG
13534AF	GG	CT	GA	GG	GG
13540AF	GG	CC	GG	GG	GG
13546AF	GG	CC	GG	GG	GG
13547M	GG	CT	GA	GG	GG
13550AF	GG	CC	GA	GG	GG
13551AF	GG	CT	GA	GG	GG
13559AF	GG	TT	AA	GG	GG
13563AF	GG	TT	AA	GG	GG
13602M	GG	CC	GA	GG	GG
13622S3	GG	CT	GA	GG	GG
13624M	GG	CT	GG	GG	GC
13624M	GG	CT	GA	GG	GC
13636APA	GG	CT	GA	GG	GC

13636M	GG	TT	GG	GG	GC
13647AF	GG	CT	AA	GT	GG
13648AF	GG	CC	GG	GG	CC
13648M	GG	CT	GA	GG	GC
13649M	GG	TT	GA	GG	GG
13654AF	GG	TT	GA	GG	GG
13664AF	GG	CT	AA	GG	GG
13668AF	GG	CC	GG	GG	CC
13673AF	GG	CC	GG	GG	GG
13677M	GG	TT	GA	GG	GG
13682AF	GG	CT	GG	GG	GG
13689AF	GG	TT	GA	GG	GG
13696AF	GG	CT	GA	GG	GC
13712AF	GG	CC	GA	GT	GC
13720S2	GG	TT	AA	GG	GG
13726AF	GG	CC	GG	GG	GC
13729M	GG	CT	GG	GG	GG
13731AP	GG	CT	GA	GG	GC
13731M	GG	CC	GA	GG	GG
13745APA	GG	CC	GG	GG	GC
13745M	GG	CT	GA	GG	GC
13639AF	GG	CC	GA	GG	GG

Table 7. The SNP results for the medical examiner sample results. The variant nucleotides are marked in red while the wild type nucleotides are shown in black.

Sample Name	MYH6	EYA4	EYA4-2	TNNI3	NEXN
	G to A	C to T	G to A	G to T	G to C
1	GA	TT	GA	GT	GC
5	GG	CT	GA	GT	GG
6	GG	CC	GG	GG	GG
7	GG	TT	GA	GG	GC
9	GG	CC	AA	GG	GG
10	GG	CT	GG	GG	GC
11	GG	CT	GA	GG	GG
12	GG	CC	GG	GG	GG
13	GG	TT	GG	GG	GG
14	GG	TT	GG	GG	GG
19	GG	CC	GA	GG	GG
21	GG	CT	GA	GT	GG
24	GG	CC	AA	GG	GG
26	GG	TT	GG	GG	GG
27	GG	CT	AA	GG	GG
28	GG	CT	GA	GG	GG
29	GG	CT	GG	GG	GG
30	GG	CT	AA	GG	GG
32	GG	CT	AA	GG	GG
33	GG	CT	AA	GG	GG
34	GG	CT	AA	GG	GG
35	GG	CT	GG	GG	GG
36	GG	TT	AA	GG	GG
37	GG	CT	GG	GG	GG
38	GG	CT	GA	NONE	CC
39	GG	CT	GG	GG	GG
40	GG	CT	GG	GG	GG
41	GG	TT	GA	GG	GC

After all SNP genotypes were determined, a count was made to define and compare the number of pathologic versus non-pathologic SNPs among the 5-gene panel in the two population groups. The numbers can be seen in Table 8 below. From this data, the

percentages of pathogenic and wildtype nucleotides found in each population group is shown in Table 9.

Table 8. The counts for varied and wildtype nucleotides found in "normal" and ME sample groups.

Normal	MYH6	EYA4	EYA4-2	TNNI3	NEXN
Path	0	45	42	2	20
NonPath	112	67	70	110	92
Totals	112	112	112	112	112
ME	**MYH6**	**EYA4**	**EYA4-2**	**TNNI3**	**NEXN**
Path	1	30	25	3	6
NonPath	55	26	31	51	50
Totals	80	80	80	78	80

Table 9. The percentages of mutation/variations and wildtype mutations found in each SNP site according to sample group.

Normal	**MYH6**	**EYA4**	**EYA4-2**	**TNNI3**	**NEXN**
Mutation	0%	40.18%	37.50%	1.79%	17.86%
Wildtype	100%	59.82%	62.50%	98.21%	82.14%
ME	**MYH6**	**EYA4**	**EYA4-2**	**TNNI3**	**NEXN**
Mutation	1.79%	53.57%	44.64%	5.56%	10.71%
Wildtype	98.21%	46.43%	55.36%	96.30%	89.29%
Total Percentages:		Normal Mutation: 19.46%		Normal Wildtype: 80.54%	
		ME Mutation: 23.38%		ME Wildtype: 76.62%	

The data from Table 8 from the normal population group can also be used as the population frequencies for the alternate forms of the SNP alleles for the di-allelic genetic markers. The designation of a nucleotide sequence variant as an allele for a genetic marker (as opposed to a mutation) is generally dependent upon the frequency of the variant in the population, which must exceed a frequency of 1%. The single "A" residue

detected in an ME sample for the MYH6 SNP may therefore represent a true mutation inasmuch this one pathogenic allele was observed in both population groups, even though its frequency exceeds 1% (i.e. 1.78%). For the remaining SNPs in the different genes, the possible variants look more like the alleles of a di-allele genetic marker (Table 8). Regardless of whether a SNP is an allele for a di-allelic system or a mutation, according to the NCBI, variant SNPs for the genes analyzed here are considered potentially "pathologic" for gene function and therefore an individual harboring a variant may be considered at elevated risk for cardiac events.

Considering all SNPs as a group among the normal and ME populations, the percentage of wild type alleles is higher than the percentage of pathogenic alleles in both population groups as seen in Table 9. Considering each gene individually, the numbers of wildtype SNPs was higher than the variants except for the EYA4 gene in the ME group (Table 9).

Table 10. The gene frequencies for the alleles in normal samples and medical examiner samples.

Normal Sample Gene Frequencies										
SNPs	**MYH6 (G-A)**		**EYA4 (C-T)**		**EYA4-2 (G-A)**		**TNNI3 (G-T)**		**NEXN (G-C)**	
Non Path	G	1	C	0.598	G	0.625	G	0.982	G	0.821
Path	A	0	T	0.402	A	0.375	T	0.018	C	0.179

ME Sample Gene Frequencies										
SNPs	**MYH6 (G-A)**		**EYA4 (C-T)**		**EYA4-2 (G-A)**		**TNNI3 (G-T)**		**NEXN (G-C)**	
Non Path	G	0.982	C	0.464	G	0.554	G	0.944	G	0.893
Path	A	0.018	T	0.536	A	0.446	T	0.056	C	0.107

Statistical Analysis of SNaPshot™ Genotypes

A contingency table was created for the allele counts to see if the frequency of wildtype and variant SNPs differed between the normal population group and the group represented in medical examiner samples (Table 10). Contingency tables examine relationships between multiple variables. The variables analyzed in this table are the frequency of wild type versus variant SNPs in the medical examiner samples and in the normal sample group. The SNP sites in the different genes were also compared with one another. The chi square results at the bottom of the tables are both <0.0001 for the ME and normal sample groups (Table 11). Since the p-value is less than 0.05, the difference in the frequency of wildtype and variant SNPs are significantly different between the normal and ME sample group. Thus, the higher frequency of pathogenic alleles in the

ME group is statistically different from their frequency in the normal group suggesting a relationship between the pathogenic SNPs detected and cardiac function defects contributing to death. Comparisons that are NOT significantly different from each other within each group are labeled with the same lowercase letter following the p-value whereas SNP sites that ARE significantly different have a different lowercase letter following the p-value (Table 11). Thus, in the ME sample group, pathogenic allele frequencies for SNPs at the EYA4 and EYA-4 loci are not significantly different from each other, both being denoted by the letter "a". NEXN, TNNI3, and MYH6 are also not significantly different from each other in terms of the frequency of pathogenic alleles. In the normal sample group, the frequencies of wild type and pathogenic alleles at the EYA4 and EYA4-2 loci again are not significantly different from each other. Interestingly however, although the frequency of pathogenic and wild type SNPs residing at the TNNI3 and MYH6 loci are not significantly different from each other, the frequency of pathogenic and wild type SNPs at the NEXN locus is significantly different from all other SNP sites.

Table 11. The contingency table comparing the relative frequencies of wildtype and pathologic alleles in the normal population and in the ME group of samples. The colors help to denote which SNP sites are not significantly different from the other. If the colors are the same, the sites are NOT significantly different.

ME Sample Group						
	EYA4	**EYA4-2**	**MYH6**	**NEXN**	**TNNI3**	**Total**
Nonpathogenic	26	31	55	50	51	213
Pathogenic	30	25	1	6	3	65
Frequency	53.57a	44.64a	1.79b	10.71b	5.56b	
Total	56	56	56	56	54	
Normal Sample Group						
	EYA4	**EYA4-2**	**MYH6**	**NEXN**	**TNNI3**	**Total**
Nonpathogenic	67	70	112	92	110	451
Pathogenic	45	42	0	20	2	109
Frequency	40.18a	37.50a	0.00c	17.86b	1.79c	
Total	112	112	112	112	112	
					Chi Square	<0.0001

Other Population Analyses

The Hardy-Weinberg principle can be used to predict genotype frequencies for a genetic marker in equilibrium in the population using allele frequencies for the predictions.[24] The predicted genotype frequencies for the 5 SNP markers were computed using allele frequencies obtained from the normal population group (Table 10). The predicted genotype frequencies were then compared with the observed genotypes for both the normal and ME population groups to determine if the 5 SNP markers are in equilibrium. Results of the H-W analysis of the SNP markers showed that both the normal population and ME populations appear to be in equilibrium for all 5 SNP markers

(Table 12). The EYA4-2 SNP comes close to being in disequilibrium in the ME group, however the p-value is still greater that 0.05 and therefore we must conclude that the evidence does not support this marker being in disequilibrium within the population.

Table 12. The expected versus observed chi square results for the normal and ME sample population. Since the p-values are all greater than 0.05, the difference between the expected and observed is not significant. However, the SNP site EYA4-2 for the ME population is very close to being not in equilibrium.

Chi Square Results for Expected vs Observed		
SNP	**Normal Population**	**ME Population**
MYH6 (G-A)	N/A	N/A
EYA4 (C-T)	0.867	0.113
EYA4-2 (G-A)	0.998	0.064
TNNI3 (G-T)	0.991	0.099
NEXN (G-C)	0.542	0.216

In an extension of the H-W analysis, it was also possible to determine if the frequencies of paired genotypes for two SNP markers were significantly different in the normal and ME population groups. Since SCD almost certainly can involve gene defects in multiple genes, a statistical analysis of paired SNPs might support a multi-gene cause of the underlying cardiac defects leading to death in the ME sample group. The observed results versus the expected results were examined, and the p-value was determined for all SNP site di-allelic comparisons.

Table 13. The Hardy-Weinberg phenotype predictions using allele frequencies from the normal population.

	MYH6		**EYA4**		**EYA4-2**		**TNNI3**		**NEXN**	
p	G	1	C	0.598	G	0.625	G	0.982	G	0.821
q	A	0	T	0.402	A	0.375	T	0.018	C	0.179

Results from the statistical analysis of expected versus observed genotypes from pairwise combinations of SNP markers are shown in Table 14a and 14b for EYA4 (blue) versus MYH6 (red).

Table 14a. The results for phenotype combinations between EYA4 and MYH6 for the normal sample group. EYA4 phenotypes are shown in blue while the MYH6 phenotypes are in red. The allele frequencies were determined from the normal sample group by using the Hardy-Weinberg

Expected Combination in EYA4xMYH6				
CCxGG	0.358	CTxGA	0	CTxAA
TTxAA	0	CCxGA	0	0
CCxAA	0	TTxGA	0	**Total:**
TTxGG	0.161	CTxGG	0.481	1
Expected Combination in Population in EYA4xMYH6				
CCxGG	20.040	CTxGA	0	CTxAA
TTxAA	0	CCxGA	0	0
CCxAA	0	TTxGA	0	**Total:**
TTxGG	9.040	CTxGG	26.920	56
Observed Combination in Population in EYA4xMYH6				
CCxGG	21	CTxGA	0	CTtxAA
TTxAA	0	CCxGA	0	0
CCxAA	0	TTxGA	0	**Total:**
TTxGG	10	CTxGG	25	56

Table 14b. The chi square results for the pairwise combinations of phenotypes for EYA4 and MYH6 in the normal samples.

EYA4xMYH6	Expected	Observed	Difference	Chi Square
CCxGG	20.040	21	0.960	0.867
TTxGG	9.040	10	0.960	>0.05
CTxGG	26.920	25	1.920	Not Significant
CTxAA	0	0	0	
TTxAA	0	0	0	
CCxAA	0	0	0	
CTxGA	0	0	0	
CCxGA	0	0	0	
TTxGA	0	0	0	

If the combinations of genotypes harboring pathogenic SNPs are suggested to correlate with cardiac defects, the p-value calculated with chi-square analysis would be <0.05. The chi-square result produced from a comparison of expected versus observed SNP phenotypes for EYA4 combined with MYH6 in the normal population group is 0.867. Therefore, this combination of genotypes appears to be unlinked and phenotypes randomly associate depending, as predicted by multiplying the individual genotype frequencies. Table 15 shows the similar analysis for all the pairwise genotype combinations in the normal and ME population groups. In the normal population group, each pair of SNP markers is observed to occur with the expected genotype frequency. Therefore, the 5 SNPs appear to all be unlinked in the normal population group.

The ME samples, however, did show a significant difference in the expected versus observed frequencies of pairwise genotype combinations (Table 15). For every pairwise combination except MYH6 X TNNI3, p-values were less than 0.05, and some combinations were extremely significant (EYA4-2 X TNNI3 for example with $p=1.13 \times 10^{-7}$). These results suggest that, among the group of DNA samples from victims of cardiac related deaths, there is a significant association of pairs of genotypes containing pathogenic alleles. Such genotype combinations may act synergistically to elevate the risk for cardiac defects leading to death. These results also support the notion that SCD is a heterogenetic condition.

The smaller a p-value (shown in "P-Value" column in Table 15) is, the stronger the suggested association. Therefore, p values produced from the pairwise comparisons shown in Table 15 might predict the "strength" of a given combination of pathogenic alleles in elevating the risk for cardiac related death. Under this hypothesis, the EYA4-2/TNNI3 SNP combination, which has a p value of 1.13 X 10-7, would represent a combination of pathogenic alleles that carries the highest risk for cardiac defects leading to death. Table 15 lists the gene SNP combinations in order of most correlated (lowest p-value) to least correlated (highest p-value) for the ME comparisons.

Performing the same exercise comparing expected and observe genotypes for three SNP markers was not undertaken because of the small size of the populations from which DNA samples were obtained.

Table 15. The chi square results for the di-allelic combinations in the normal sample groups and the ME sample groups. The combinations did not correlate with SCD if the p-value result was higher than 0.05, and the combinations were correlated to SCD if the p-value result was lower than 0.05.

Combination	Sample Type	P-Value
EYA4-2/TNNI3	Normal	0.9488
	ME	1.1344E-07
EYA4-2/NEXN	Normal	0.1260
	ME	2.6728E-05
EYA4/EYA4-2	Normal	0.1472
	ME	0.0003
EYA4/TNNI3	Normal	0.9991
	ME	0.0014
MYH6/EYA4-2	Normal	0.9057
	ME	0.0020
EYA4/NEXN	Normal	0.1017
	ME	0.0025
TNNI3/NEXN	Normal	0.9811
	ME	0.0237
MYH6/EYA4	Normal	0.8673
	ME	0.0308
MYH6/NEXN	Normal	0.5423
	ME	0.0346
MYH6/TNNI3	Normal	0.9908
	ME	0.2655

CHAPTER V

DISCUSSION AND CONCLUSIONS

Sudden cardiac death (SCD) is the name given to an unexpected death with no obvious outward cause that occurs in one who has no history of cardiac problems. SCD can result from conditions affecting the heart muscle such as cardiomyopathy, and often SCD results from a defect in the electrical system controlling heart muscles contraction creating irregularities in the heartbeats. The sudden stop in cardiac function causes the victim to lose consciousness, which makes it difficult for first responders to understand a victim's situation and provide emergency care. The main concern for a victim of SCD is the loss of blood flow to the brain, which can cause severe brain damage.

SCD is rooted in defects in one or more members of a collection of genes that are all associated with cardiac function. Nucleotide sequence variations in one of these genes may only marginally affect gene function or a sequence variation may silence gene expression altogether. Nucleotide sequence variations in individual genes that are not lethal by themselves may contribute to an overall increase in the risk of SCD when a collection of genes with variant nucleotide sequences is considered as a group. The spectrum of effects resulting from nucleotide sequence variations in the large number of genes associated with cardiac function is likely responsible for heterogeneity in SCD in terms of the age of a victim and also for the difficulty in detecting particular anomalies in

cardiac function that would be revealed in typical physical exams. The one consistent feature of SCD is the premature death of young people.[1]

Although limited physical or histological evidence of SCD is available from victims to use diagnostically, a lot is known about genes involved with cardiac function and the study of such genes may facilitate identifying gene variants that are correlated with defects in cardiac function. Cardiomyopathies are a common type of heart defect and, of the four principal types, hypertrophic cardiomyopathy (HCM) has been implicated in a high proportion of athletic deaths.[11] All four types of cardiomyopathy have been associated with variations in the nucleotide sequence of members of a panel of genes associated with cardiac function.[11]

Most SCD diagnoses are made post-mortem since many of the clues associated with SCD are missed during routine or even detailed physical exams. As stated previously, the lack of physical symptoms also makes it difficult for medical examiners or pathologists to determine cause of death and in some cases SCD is listed as the cause of death when no other cause can be identified. Molecular autopsies have gained popularity for this reason. Post-mortem genetic studies can help to confirm cause of death by analyzing the victim's DNA, searching for nucleotide sequence variations that could have affected cardiac function in a lethal way. Besides being able to provide a cause of death, diagnosis of SCD is important in order to warn surviving family members of their possible risk for SCD events in the future.

There are many genes that may contribute to elevated risk of SCD events. Commercial vendors of DNA typing reagents market next generation DNA sequencing

kits that interrogate SNP sites in hundreds of genes simultaneously (Illumina, Inc., San Diego, CA; Applied Biosystems Inc., Carlsbad, CA; Qiagen Inc., Mansfield, MA). Interrogating a large panel of genes, while yielding a complete genetic picture of the constellation of SNPs present in the genome, may not result in a conclusive diagnosis of the underlying cause of SCD for a given victim because of the interplay of SNPs in numerous genes that collectively elevate risk. Moreover, the use of next generation DNA sequencing, while providing in depth coverage of the genes potentially involved, is expensive and data management and storage are extensive. Our strategy for investigating a possible correlation of SCD or other cardiac conditions leading to death with gene defects was to use a simple molecular assay interrogating 5 SNPs residing in 4 genes known to be associated with cardiac function. The experimental plan was to screen genomic DNA isolated from blood samples from a cohort of individuals investigated for cause of death by the Medical Examiner as well as a group of presumably normal individuals. The four genes chosen (EYA4, MYH6, TNNI3, and NEXN) have been found by other researchers to play a role in the normal functioning of cardiac muscle and SNPs residing within these genes have been implicated with dilated cardiomyopathy and hypertrophic cardiomyopathy.[11] The five SNPs interrogated in our panel all result in nucleotide changes within the genes: the SNPs within EYA4 and NEXN all change a nucleotide within an untranslated region within the mRNA whereas the SNPs associated with EYA4-2, MYH6, and TNNI3 represent missense changes that alter the amino acid sequence of the protein.[4] While an untranslated region may seem like it should have no consequence or protein function, enhancers are known to exist within the UTR and so it is possible gene expression effects could accompany SNPs residing within the UTR. The

gene EYA4 has two SNP sites and variant alleles at both sites have been suggested by others to have pathogenic effects on gene function.[19] The EYA4 gene has been associated with cardiac muscle function and some forms of the gene are associated with cardiomyopathies that cause impaired contracting of the ventricles in the heart.

MYH6 was chosen for our panel due to its causation of premature deaths from cardiac failure due to dilated cardiomyopathies. TNNI3 is a gene that encodes a protein in the heart called troponin I. Troponin I regulates heart contractions and nucleotide sequence variations in TNNI3 has been found associated with many cardiomyopathies. Nucleotide sequence variations in the NEXN gene have been associated with hereditary dilated and hypertrophic cardiomyopathies that can cause the heart to collapse after intense exercise.[20] Due to the genes' associations with cardiomyopathies, the SNPs they possess are good candidates for exploring the association of SNPs with lethal defects in cardiac function. The SNP's have been named after the gene they are located in for convenience: the two EYA4 SNPs are EYA4 and EYA4-2, MYH6's SNP is MYH6, TNNI3's SNP is TNNI3, and NEXN's SNP is NEXN. These genes and their nucleotide sequence variants were chosen for our studies due to published work by Millat et al (2014) Brion et al. (2015) and from previous next generation DNA sequencing performed at OSU by Dr. Shen Di. In her work, a limited number of bloodstain samples were obtained from the Oklahoma Office of the Chief Medical Examiner and the DNA extracted from the stains was subjected to massively parallel DNA sequencing (MPS). The library used in MPS consisted of 58 genes associated with SCD. The SNPs in NEXN, EYA4, and TNNI3 were all chosen from this study and then the possible role of these genes in SCD was confirmed from the literature and from the NCBI database.

SNaPshot™ Results

The results of our genotyping produced 112 alleles (56 individuals) for each of the SNP sites from the normal population group and 56 alleles (28 individuals) for each of the SNP sites from the Medical Examiner (ME) population group. Analysis of the data began with a simple analysis of whether there were more pathogenic variants at the SNP sites in the ME group when compared with the normal group. Results from that analysis showed that indeed the ME population group harbored more pathogenic SNPs overall than did the normal group although pathogenic alleles did exist in the normal population group. Given the frequency of the wild type and variant SNPs for all genes (except MYH6), it appears these SNPs represent di-alleles rather than mutations. Thus, SNP-alleles for each site were counted in the normal population to establish an allele frequency database. It should be noted here that although most SNPs exhibited characteristics of di-allelic genetic markers (because the frequencies of the alternate nucleotide forms of the SNP exceeded by 1% by a large margin), the SNP associated with MYH6 exhibited characteristics more akin to a mutation in that only a single SNP variant was found in the MYH6 gene in an ME sample and no variants were found in the normal population group. We also performed population genetic analysis of the genotyping results and found that each SNP-gene was in equilibrium within both the normal and ME population groups. The EYA4-2 SNP was close to not being in equilibrium and perhaps if the number of ME samples was higher, an association of EYA4-2 with cardiac death might have been revealed (Table 12).

The differences in overall frequency of wild type and pathogenic nucleotides at the individual SNP sites in the normal samples and in the medical examiner samples were evaluated statistically using a contingency table. The difference in frequency of wild type and pathogenic alleles is significant considering all SNPs as a whole ($p < 0.0001$). Thus, the frequency of pathogenic alleles is significantly higher in the ME group than the normal group. Moreover, within each population group, the frequencies of wild type and pathogenic alleles at the EYA4 and EYA4-2 loci are NOT significantly different from each other. This might be expected since both SNPs are found in the same gene. NEXN, TNNI3, and MYH6 are also seen as NOT significantly different from each other as denoted by the letter "b". While NEXN and TNNI3 are not different from each other, they are significantly different from EYA4 and EYA4-2. Within the normal population group, wild type and pathogenic SNPs at EYA4 and EYA4-2 are also NOT significantly different in the normal sample group, as was the case for the ME sample group. However, the EYA4 SNPs are significantly different from all other SNPs. MYH6 and TNNI3 are this time seen as NOT significantly different from each other and, but NEXN *is* significantly different from all other SNPs. With the exception of NEXN, pathogenic alleles were more common in the ME sample group.

Given the heterogeneous causes of SCD, it was also important to ask whether particular combinations of nucleotide variants in different genes were significantly different in the ME and normal population groups. To address this question, we performed analyses of SNP genotypes from pairs of genes in the two population groups, asking whether the expected and observed frequencies at two SNP sites were significantly different. If two genetic markers are in population equilibrium, it is possible

to use the population genetic theory to predict the frequency of genotypes for each marker, given the allele frequencies. If two genetic markers are in linkage equilibrium, it is possible to multiply each genotype frequency by the other to produce the frequency of a pair of genotypes that can be found in a given individual within the population. If combinations of pathogenic SNPs are "linked" or significantly "correlated" with cardiac function anomalies that led to the death of members of the ME population group, we would expect the observed genotypes to vary significantly from expectations. When this experiment was performed we found that the observed pairwise comparisons of SNPs did not differ significantly from expectations in the normal population group, but did differ significantly in the ME population group. These results support the general consensus of opinion that cardiac disease, and especially cardiac disease leading to unexpected death, results from the interplay of more than one SNP. Among the pairwise comparisons p-values ranged over multiple orders of magnitude suggesting some SNP combinations could be more significant in elevating risk for cardiac risk than others.

Limitations of the Study

There are several limitations in our study. One key limitation is the size, age, and type of the group of samples that were tested. Large control samples are often used to make comparisons to results produced from an affected group. While a large number of control and SCD samples would have been extremely helpful and possibly strengthened our conclusions, it was difficult to obtain large numbers of samples from victims known to have died from SCD. SCD is difficult to diagnose and until a compelling diagnostic

test is widely available, collecting a large cohort of affected DNA samples will be difficult. Studies performed by Skinner et al. and Brion et al. focused on young sudden unexplained deaths (SUD) rather than expanding their number of samples to include deaths resulting for all cardiac causes. While a definitive diagnosis for SCD will likely await extensive DNA analysis using some of the new technologies available today, interpretation of the massive amount of data produced and the high frequency of SNP variations in human genomic DNA, deciphering the SNPs that are causative of SCD from those that are not associated with SCD represents a complex challenge that will take time.

Many studies (including Brion et al. 2015) use large gene-panels to investigate possible genetic causes of SCD. However, as our results showed, presumably pathogenic alleles can be detected in unaffected individuals who have advanced in age beyond that considered at an elevated risk for SCD. Therefore, even massively parallel DNA sequencing faces challenges in designing a compelling diagnostic molecular assay that is effective for the diagnosis of SCD. In contrast, a simple molecular assay, while perhaps missing gene defects that could result in a diagnosis of SCD nonetheless was effective in a limited population of victims of cardiac related death of identifying pathogenic SNPs that were highly correlated with defects in cardiac function that could have explained the death. It would be easily possible to expand the number of SNPs subjected to SNaPshot analysis to increase the sensitivity of the assay for diagnostic purposes.

Conclusions and Possible Future Studies

SNP polymorphisms occur within the human genome with a frequency of about 1 per 500-1000 nucleotides of genomic sequence. Hence variations in nucleotide sequence are common both in genes and between genes and yet have no apparent effect on survival and normal life. The results of our study would be enhanced by increasing the size of the two population groups used as a source of DNA for study. In addition, the sensitivity of our approach-using SNaPshot for the diagnosis of SCD-would be improved if additional SNPs in other genes associated with cardiac function were included in our limited gene panel. While sequencing large panels composed of hundreds of genes could be the most informative for diagnosing SCD, it seems there needs to be more of a consensus on the criteria that define SCD and there will need to be bioinformatics solutions to analyzing the massive amount of DNA sequence produced with this approach. Our results suggest that a more limited approach using simpler molecular technology that could be used by the medical examiner would improve their ability to identify cause and manner of death for those victims whose autopsy results do not identify the cause of death. If pathogenic alleles are identified in a decedent, the ME could then alert surviving family members to follow up with their own healthcare providers to help ensure their continued health in the future. And, if molecular analysis could be conducted on those who are at significant risk for SCD, such as young athletes, perhaps surprising deaths during strenuous exercise could be reduced.

REFERENCES

1. Sudden Cardiac Death (Sudden Cardiac Arrest). Cleveland Clinic. https://my.clevelandclinic.org/health/articles/sudden-cardiac-death. Accessed September 5, 2017.

2. About Arrhythmia. http://www.heart.org/HEARTORG/Conditions/Arrhythmia/AboutArrhythmia/About-Arrhythmia_UCM_002010_Article.jsp#.WzUAYa2ZMxc. Accessed June 28, 2018.

3. Brion M, Sobrino B, Martinez M, Blanco-Verea A, Carracedo A. Massive parallel sequencing applied to the molecular autopsy in sudden cardiac death in the young. *Forensic Sci Int Genet*. 2015;18:160-170. doi:10.1016/j.fsigen.2015.07.010

4. Information NC for B, Pike USNL of M 8600 R, MD B, Usa 20894. National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/. Accessed August 31, 2017.

5. Arking DE, Sotoodehnia N. The Genetics of Sudden Cardiac Death. *Annu Rev Genomics Hum Genet*. 2012;13:223-239. doi:10.1146/annurev-genom-090711-163841

6. Zipes DP, Wellens HJJ. Sudden Cardiac Death. *Circulation*. 1998;98(21):2334-2351. doi:10.1161/01.CIR.98.21.2334

7. Ackerman M, Atkins DL, Triedman JK. Sudden Cardiac Death in the Young. *Circulation*. 2016;133(10):1006-1026. doi:10.1161/CIRCULATIONAHA.115.020254

8. Finocchiaro G, Papadakis M, Sharma S, Sheppard M. Sudden Cardiac Death. *Eur Heart J*. 2017;38(17):1280-1282. doi:10.1093/eurheartj/ehx194

9. Klabunde R. *Cardiovascular Physiology Concepts*. 2nd ed. Baltimore, MD: Lippincott Williams & Wilkins; 2012.

10. Skinner JR, Morrow PL. Cardiac genetic investigation of sudden cardiac death: advances and remaining limitations. Research and Reports in Forensic Medical Science. doi:10.2147/RRFMS.S72063

11. Cruickshank RN S. Cardiomyopathy. *Nurs Stand*. 2003;18(23):46-522.

12. Gupta AK, Gupta UD. Chapter 19 - Next Generation Sequencing and Its Applications. In: Verma AS, Singh A, eds. *Animal Biotechnology*. San Diego: Academic Press; 2014:345-367. doi:10.1016/B978-0-12-416002-6.00019-5

13. Kakoi H, Tozaki T, Nagata S, Gawahara H, Kijima-Suda I. Development of a method for simultaneously genotyping multiple horse coat colour loci and genetic investigation of basic colour variation in Thoroughbred and Misaki horses in Japan. *J Anim Breed Genet*. 126(6):425-431. doi:10.1111/j.1439-0388.2009.00841.x

14. Carvajal-Zarrabal O, Hayward-Jones PM, Nolasco-Hipolito C, Barradas-Dermitz DM, Calderón-Garcidueñas AL, López-Amador N. Use of Cardiac Injury Markers in the Postmortem Diagnosis of Sudden Cardiac Death. *J Forensic Sci*. 2017;62(5):1332-1335. doi:10.1111/1556-4029.13397

15. Lahrouchi N, Raju H, Lodder EM, et al. Utility of Post-Mortem Genetic Testing in Cases of Sudden Arrhythmic Death Syndrome. *J Am Coll Cardiol*. 2017;69(17):2134-2145. doi:10.1016/j.jacc.2017.02.046

16. Millat G, Chanavat V, Rousson R. Evaluation of a new NGS method based on a custom AmpliSeq library and Ion Torrent PGM sequencing for the fast detection of genetic variations in cardiomyopathies. *Clin Chim Acta*. 2014;433:266-271. doi:10.1016/j.cca.2014.03.032

17. Baars HF, Doevendans PAFM, Houweling AC, Tintelen JP van. *Clinical Cardiogenetics*. Springer; 2016.

18. Albert CM, MacRae CA, Chasman DI, et al. Common Variants in Cardiac Ion Channel Genes are Associated with Sudden Cardiac Death. *Circ Arrhythm Electrophysiol*. 2010;3(3):222-229. doi:10.1161/CIRCEP.110.944934

19. Paneto GG, Careta F de P. Designing Primers for SNaPshot Technique. In: *PCR Primer Design*. Methods in Molecular Biology. Humana Press, New York, NY; 2015:165-172. doi:10.1007/978-1-4939-2365-6_12

20. National Library of Medicine - National Institutes of Health. https://www.nlm.nih.gov/. Accessed November 16, 2017.

21. Mayo Clinic. Dilated Cardiomyopathy. 2017.

22. Makielski JC. 98 - Sudden Infant Death Syndrome. In: Zipes DP, Jalife J, eds. *Cardiac Electrophysiology: From Cell to Bedside (Sixth Edition)*. Philadelphia: W.B. Saunders; 2014:975-979. doi:10.1016/B978-1-4557-2856-5.00098-4

23. Mathur R, ed. Statistical significance. *Work*. 2005;Spring 2005: Institute for Work & Health, Toronto(40). https://www.iwh.on.ca/what-researchers-mean-by/statistical-significance. Accessed June 4, 2018.

24. Hardy-Weinberg Law. Merriam-Webster.com.

VITA

Tierney Mullaney

Candidate for the Degree of

Master of Science

Thesis: MUATATIONS IN GENOMIC DNA ASSOCIATED WITH SUDDEN CARDIAC DEATH

Major Field: Forensic Sciences

Biographical:

Education:

Completed the requirements for the Master of Science in Forensic Sciences with specialization in Forensic Biology/DNA at Oklahoma State University, Center for Health Sciences, Tulsa, Oklahoma in July 2018.

Completed the requirements for the Bachelor of Science in Biology at South Dakota State University, Brookings, SD/USA in 2016.

Experience: Worked as a Research Assistant at Oklahoma State University-Center for Health Sciences. Worked as a Forensic Pathologist Technician at the Office of the Chief Medical Examiner of Oklahoma in Tulsa.

Professional Memberships: Member of the American Academy of Forensic Science. Member of the International Association for Identification. Member of South Western Association of Forensic Scientists.