Exploration of Single Nucleotide Polymorphisms (SNPs) within Four Genes Involved with Cardiac Function and Association with Sudden Unexplained Death

Tierney Mullaney, MSFS, OSU Center for Health Sciences, School of Forensic Sciences

Jane Pritchard, BS, MT(ASCP)MB, OSU Center for Health Sciences, School of Forensic Sciences

Mark Payton, PhD, Oklahoma State University, Department of Statistics

Jun Fu, PhD, OSU Center for Health Sciences, School of Forensic Sciences

Shen Di, PhD, OSU Center for Health Sciences, School of Forensic Sciences

Robert W. Allen, PhD, OSU Center for Health Sciences, School of Forensic Sciences

Abstract

Seventeen blood samples obtained at autopsy from victims of cardiac-related causes were subjected to DNA analysis to explore possible associations of single nucleotide variations in a panel of four genes known to be involved in cardiac function. DNA extracted from autopsy samples as well as from 56 presumably normal individuals was subjected to nucleotide sequencing at sites within the *EYA4*, *MYH6*, *TNNI3*, and *NEXN* genes involved with cardiac function. Five single nucleotide polymorphisms (SNPs) in the four genes were interrogated through DNA sequencing. Each SNP site harbored a nucleotide considered to represent a wild type allele or a nucleotide variant potentially affecting the function of the gene product. The purpose of the study was to determine if a greater number of variant alleles exist in the samples from the Medical Examiner than in the general population, and whether any particular SNP phenotypes or combinations of phenotypes are linked to cardiac-related death. Results showed individuals dying from cardiac-related causes had a significantly higher frequency of variant SNPs than did the Normal group. Moreover, some combinations of variant phenotypes were associated with cardiac death in the Medical Examiner group. Our results support the notion that death from cardiac-related causes can result from single nucleotide polymorphisms that, by themselves, are not lethal but result in less than optimal performance of their respective gene products. Variant SNPs in several genes may collectively compromise cardiac function sufficiently to elevate the risk of death when elevated demands are placed on cardiac function.

Introduction

Sudden cardiac death (SCD) defined, in the broadest sense, is an unexpected death that is caused by a sudden loss of cardiac function.[1-3] The World Health Organization defines sudden cardiac death as an "unexpected death due to a cardiac cause within one hour after the onset of symptoms in a person with known or unknown cardiovascular disease".[3] In individuals over 40 years of age, coronary artery disease and acute myocardial infarction are the most common causes of cardiac-related death. However, in individuals under 40 years of age,

cardiomyopathies and channelopathies, which can be linked to genetic causes, are the most common.[1-8] SCD can occur when the electrical system of the heart malfunctions causing arrhythmias and can also result from structural defects in heart muscle that compromise the contractile efficiency.[2] Both situations starve the tissues for oxygenated blood and can lead to unexpected death.

In the U.S., sudden cardiac death accounts for half of all heart disease deaths, killing around 325,000 adults each year.[1-6] This is the leading cause of natural death annually. Adults in their mid-30s are most likely to be affected by SCD, and the condition affects men at twice the rate of women.[2] The detection of SCD post-mortem can sometimes pose a significant challenge to the Medical Examiner (ME) because there is often not gross or microscopic evidence to explain why the heart stopped beating. In some cases, a diagnosis of SCD is made only tentatively when all other possible causes of death have been ruled out.[1,2,4] According to Deo and Albert and Tester et.al., approximately 5% of cases involving cardiac-related deaths fall into this category in which no discernable cause of death can be identified.[2,4]

Single nucleotide polymorphisms, or SNPs, are nucleotide variations that change a DNA sequence and, when they occur within genes, SNPs can alter the amino acid sequence of a gene product or even silence a gene altogether. SNPs exist throughout the human genome and those that exist within genes can have effects on the activity of a gene product ranging from little to no reduction in activity to a severe reduction or even the complete abolishment of activity. It is possible that a single mutation eliminating the critical functioning of a gene product essential for cardiac function could result in a sudden cardiac-related death. However, the literature suggests it is also possible that suboptimal functioning of the products of a constellation of genes associated with cardiac function can cause SCD.[1,9] In other words, any single SNP might not completely abolish the function of a given gene product but rather simply result in sub-optimal activity which would not be life threatening. SNPs causing the sub-optimal functioning of a collection of gene products however could be life threatening when extraordinary demands are made on the heart.

One of the outcomes of the Thousand Genomes Project has been the detection and mapping of millions of SNPs in the human genome and, in some cases, SNPs in genes have been linked to pathological conditions.[10,11] This is true for genes associated with cardiac function where panels of genes have been identified and can be easily sequenced using commercially available kits.[12] Such gene panels are useful for detection of SNPs that may affect the function of the heart and advance the possibility of developing a diagnostic strategy to reveal individuals at risk for SCD before an event occurs.[1,3,9,13] While technology is currently available to easily produce DNA sequence information from a panel composed of hundreds of genes in a single reaction, the relevance of the multitude of SNPs that will be revealed in the data to a given pathological condition can be difficult to assess. For example, a panel of 58 cardiac-related genes we sequenced using massive parallel DNA sequencing (MPS) detected an average of 200 nucleotide positions within the genes in each of four genomic DNA samples analyzed that differed from the consensus sequence of those genes present in the reference human sequence library available in GenBank (unpublished observations). Although there were an average of 200 SNPs detected in the four samples, they were not the same 200 SNPs that were shared. Two of the genomic DNA samples tested were obtained from individuals who died from cardiac-related causes. However,

two others died from drug overdoses and yet still revealed ~200 SNPs in the cardiac panel. Evaluating which SNPs from among the hundreds identified are associated with cardiac-related death thus represents a significant challenge. Similar results and challenges have been reported by others.[1,3,9,13]

In this study, five SNPs residing in four genes involved with cardiac function were evaluated for an association with cardiac-related death. Each of the SNPs studied was listed in a publically available database of the National Center for Biotechnology Information (NCBI), a program of the National Institutes of Health, in an area of the database known as ClinVar which lists gene variations and their association with pathological conditions.[14] Five SNPs harbored within the four genes constituting the panel used for this study are referenced in the ClinVar database and variant nucleotides residing at these SNP sites are suggested to either represent "pathological" variants, likely benign variants, or variants of "uncertain" effect. It should be emphasized however, that a SNP denoted as "pathological" by the ClinVar database might be designated as non-pathological by other laboratories experienced with the interpretation of SNPs or other gene variations detected through DNA sequencing.[3,13]

In the study reported here, the nucleotide resident at each SNP site was determined in a cohort of 17 blood samples obtained at autopsy from victims of cardiac-related death (the ME population group) using a variation of Sanger DNA sequencing. This process was likewise performed with a cohort of 56 DNA samples from individuals reflecting the population at large (the Normal population group). The overall frequency of variant SNPs was significantly higher in the ME sample group than the Normal group and some pairwise combinations of SNP genotypes within the four genes showed a significant association with cardiac-related death.

Material and Methods

Genes and SNPs

Four genes were chosen for this study based upon results of preliminary DNA sequencing analyses performed by our laboratory. In addition, suggestions from the literature and the ClinVar database helped in choosing the genes for this panel.[14] The genes chosen and the characteristics of the SNPs harbored by them are summarized in Table 1.

Table 1. Gene/SNP Panel Characteristics

Gene	Chromosome	NCBI ID	SNP Ref	Wild Type>Variant (Effect)*
EYA4	6q23.2	NM_172105.3	rs3734279	C>T (3'UTR)
EYA4-2	6q23.2	NM_172105.3	rs9493627	G>A (Missense, Gly277Ser)
MYH6	14q11.2	NM_002471.3	rs140596256	G>A (Missense Glu98Lys)
TNNI3	19q13.4	NM_000363.4	rs3729711	G>T (Synonymous; Arg68)
NEXN	1p31.1	NM_144573.3	rs3767028	C>G (3'UTR) rev

*Data taken from the ClinVar database: https://www.ncbi.nlm.nih.gov/clinvar/. The designation of a SNP as wild type indicates the nucleotide at that position matches that recorded in GenBank and does not alter the coding sequence of the gene.

The *EYA4* gene is a member of the "eyes absent" family of proteins.[15,16] Putative functions for the gene include acting as a transcriptional activator necessary for eye development and function. Defects in the gene are also associated with dilated cardiomyopathy.[15,16] The *MYH6* gene encodes the heavy chain of cardiac muscle myosin.[16-18] Defects in the *MYH6* gene can cause familial hypertrophic cardiomyopathy and atrial septal defect 3.[16,17] The *TNNI3* gene encodes troponin I which is a member of the troponin complex that is exclusively expressed in cardiac muscle.[19,20] Mutations in this gene cause familial hypertrophic cardiomyopathy type 7 and familial restrictive cardiomyopathy.[16,18] It is noted that the SNP in *TNNI3* included in the panel is synonymous (i.e., doesn't change the amino acid sequence of the polypeptide). However, the ClinVar database information for this SNP states that it is tightly linked to another variation within the *TNNI3* gene that is likely pathological.[14] Therefore, detection of the variant allele in our panel would indirectly identify the pathological SNP at the other position in the *TNNI3* gene. Finally, the *NEXN* gene encodes an actin-binding, filamentous protein that may function in cell adhesion and migration.[21,22] Mutations in the *NEXN* gene have been associated with dilated cardiomyopathy.[16,21,22]

Samples Used for Testing

Blood samples, provided as blood stains, were obtained from 17 victims of cardiac-related death as determined through autopsy by the Office of the Chief Medical Examiner for the State of Oklahoma (ME). Nine blood samples were obtained from individuals <30 years old, three of which were less than one-year old. Six samples were from individuals over 40 but less than 60 years old. Two samples were from individuals just over 30 years old (32 and 34). Ten samples were obtained from males and seven from females. All individuals were determined to either have died from cardiac-related issues or from no discoverable cause. Bloodstains were created on bloodstain cards (Fitzco 705, Fitzco Inc., Spring Park, MN) by spotting 100 µl aliquots of blood into each of four pre-printed circles on the bloodstain card. Spotted cards were dried overnight at room temperature before being stored in paper envelopes at room temperature in the dark. These samples were designated the ME population group.

DNA samples were also retrieved from our archived DNA sample repository. A total of 49 previously extracted DNA samples were selected from the repository that originated from individuals between 40 and 60 years of age, were all alive, and represented the population at large. This group of samples was designated the Normal population group.

DNA Extraction

DNA was extracted from the ME group of samples using a procedure involving digestion of the stains with Proteinase K (Promega Corp., Madison, WI) in the presence of 0.5% (v/v) sodium dodecyl sulfate (SDS) detergent. Stains, representing 100 µL of dried blood, were cut from the stain cards and minced into small pieces. The pieces were transferred to a 0.6 mL microfuge tube to which was added 300 µL of TNE (10 mM Tris-Cl, pH 8.0 containing 0.2 M NaCl and 0.1 mM EDTA) containing 400 µg/mL of Proteinase K and 0.5% SDS. Samples were incubated at 65ºC for 1.5 hours and then were subjected to extraction with a mixture of phenol: $CHCl_3$:isoamyl alcohol (9:0.96:0.04; v/v) through vigorous vortex mixing to separate contaminating cellular constituents from genomic DNA. Organic extracts were centrifuged at 10,000 xg at room temperature for one minute to separate the organic and aqueous phases, and the aqueous phase (containing the DNA) was collected and subjected to extraction with

$CHCl_3$:isoamyl alcohol (24:1; v/v) to remove any residual phenol from the DNA contained within the aqueous phase. DNA was recovered from the aqueous phase of the organic solvent extract by binding it to silica spin columns in the presence of 5-6 M guanidine isothiocyanate (Clean and Concentrator kit, Zymo Research, Orange, CA) following procedures supplied by the vendor. The DNA concentration was estimated using spectrophotometry (Thermo Fisher, Waltham, MA).

DNA Sequencing Using SNaPShot™ Technology.

The SNaPShot™ kit (Thermo Fisher, Waltham, MA) contains all the reagents necessary to perform single nucleotide sequencing of DNA using chain termination technology. The method consists of two stages: in the first stage, a short PCR amplicon is produced that harbors the SNP, preferably in the middle of the amplicon. In this study, amplicons of 100-150 bp were produced that harbored the SNP site somewhere near the middle of the fragment.

In stage two, the amplicon is denatured and hybridized to an oligonucleotide primer whose sequence terminates one nucleotide upstream from the SNP site (i.e. to the 5' side of the SNP). If the oligonucleotide:single stranded template hybrid is then incubated with DNA polymerase and a mixture of di-deoxynucleotide triphosphates (ddNTPs), a single ddNMP will be incorporated at the SNP site that is complementary to the nucleotide resident there. Further, if the ddNTPs are labeled with one of four fluorescent dyes [i.e., ddATP = green, ddGTP = blue, ddCTP = yellow (shown as black), ddTTP = red], the color incorporated into the oligonucleotide primer will identify the nucleotide resident at the SNP site.

Thus, the incorporated ddNMP extends the oligonucleotide primer by one nucleotide and the extended primer is detectable as a small fluorescent DNA fragment migrating during capillary electrophoresis in a 3130XL genetic analyzer (Applied Biosystems, Foster City, CA). The length of the primers can be engineered such that multiple SNPs can be interrogated simultaneously and the collection of fluorescent products of the SNaPShot™ reaction can be identified in an electropherogram based upon the known lengths of the products. Characteristics of the primers used for this study are summarized in Table 2.

Table 2. Characteristics of SNaPShot™ Primers Used*

Gene	Amplicon Size (bp)	Direction	Sequence 5'-3'
EYA4	20	Forward	TCTCTCTCCCATCCCTCCTT
EYA4	20	Reverse	TGTAGAGCCAGAGGCATTGA
EYA4	34	SNP	AAAAAAAAAAAACTGTGTTCTTTAGCCGGA GATC
EYA4-2	22	Forward	TCAGAAACAAATGGGGCTGAGT
EYA4-2	20	Reverse	ACGCTCCATACGTTGATGCT
EYA4-2	40	SNP	AAAAAAAATTTCAGGATTATCCATCCTATA CAGCCTTT
MYH6	20	Forward	CCACCCAAGTTCGACAAGAT
MYH6	20	Reverse	GTCTCTCCCCCTCTTCTTGG
MYH6	29	SNP	AAAAAAAAAACATGCTGACCTTCCTGCAC
TNNI3	21	Forward	GGTCTTTATCCTGAAGCCCCG
TNNI3	20	Reverse	AGAAACCTCGCATCCTTGGG
TNNI3	46	SNP	AAAAAAAAAAAAAAAAAAAAAAAAAAAA AAGAGAGGCGGAGGAGCG
NEXN	24	Forward	AGCAGTCAACAATAAAGGATCTGC
NEXN	21	Reverse	CCGCCACAGAAAAGATGGATG
NEXN	53	SNP (antisense)	ATCAGCTAAAAAAGAGAAGAAGAAAAGTG ATTTTAAGGAAAAAAAAATTAATA

*The forward and reverse primers are used to amplify the short piece of genomic DNA that will serve as the template for the SNaPShot™ sequencing reaction. The primers labeled "SNP" are those that terminate one nucleotide upstream from the SNP site. Note the length of these primers differs and determines where the fluorescent products will migrate during capillary electrophoresis. In some cases, multiple adenosine residues are added to the SNP primer to manipulate the overall size of the SNaPShot™ product.

Data Analysis

The hypothesis tested in this study was that the ME sample group would harbor a higher number of variant SNPs within the four genes known to be involved with cardiac function than would the Normal sample group if the variant alleles are associated with cardiac-related death. The elevated number of variant SNPs would also be expected to produce differences in the frequencies of SNP genotypes for individuals in the ME and Normal population groups that might also be associated with cardiac-related death.

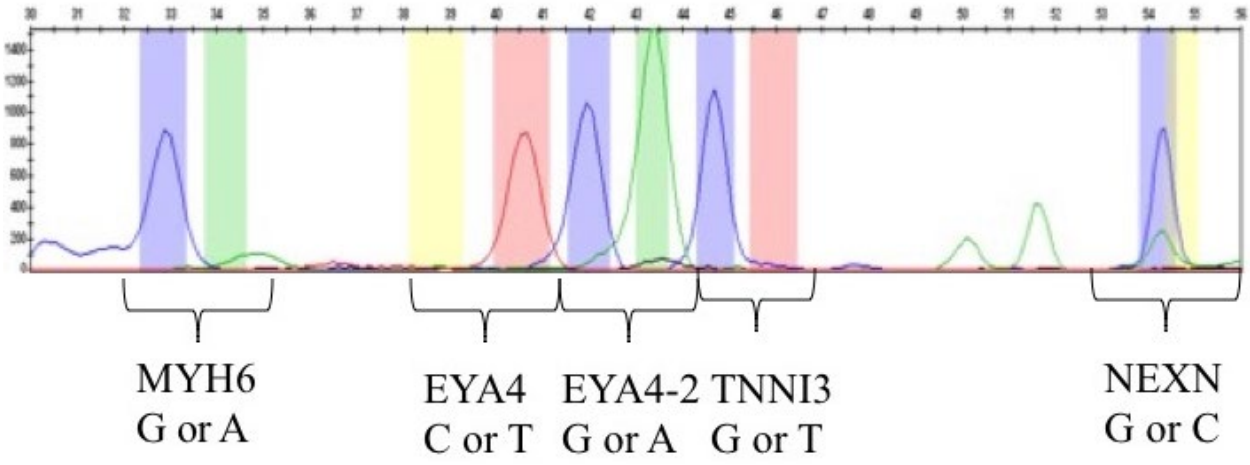
The frequencies of variant versus wild type SNPs in the two sample populations was determined from SNaPShot™ genotyping results. Wild type and variant SNPs in the Normal population group exhibited frequencies suggesting they could be treated as if each SNP represented a di-allelic genetic marker. This was true for four of the five SNPs studied. Therefore, it was possible to use statistical analysis methods often used in population genetics to ask if there was an association of particular genotypes, or combinations of genotypes, with the common phenotype of cardiac-related death in the ME population group. The significance of any

associations could be evaluated using Chi square analysis of expected versus observed genotype frequencies or the frequency of genotype pairs.

Results

Representative SNaPShot™ results for the SNP markers resident in the gene panel are shown in Figure 1. Primers used to interrogate each SNP are engineered to be different lengths so that the chain terminated oligonucleotides will migrate to different positions during capillary electrophoresis. Since the length of the primer is known, the different SNPs can be discriminated and the colors of the fluorescent peaks in the electropherogram will identify the nucleotide present at each site.

Figure 1. Electropherogram of Typical SNaPShot™ Results*



*The electropherogram was produced using a 3130XL capillary electrophoresis instrument. Note that even though a single oligonucleotide primer is used to type each SNP, the migration of primers differs slightly depending upon the nucleotide incorporated because of the chemical characteristics of the dyes.

In the profile shown in figure 1, the individual tested is homozygous for the G, wild type, allele at the *MYH6* locus, homozygous for the variant T allele at the *EYA4* locus, heterozygous for the *EYA4-2* SNP with one chromosome harboring the wild type G allele and the other harboring the variant A allele. For both *TNNI3* and *NEXN*, the sample donor is homozygous for the wild type G allele.

SNaPShot™ genotyping was performed for the DNA samples obtained from the Normal and ME population groups and the resulting genotypes are summarized in Tables 3 and 4.

Table 3. SNP Genotypes* Produced from the Normal Population Group

Sample Number	MYH6	EYA4	EYA4-2	TNNI3	NEXN
	G to A	C to T	G to A	G to T	G to C
1	GG	CC	GG	GG	GG
2	GG	CC	AA	GG	GG
3	GG	CT	GA	GG	GG
4	GG	CC	GG	GG	CC
5	GG	CC	GG	GG	GC
6	GG	CT	GG	GG	GG
7	GG	CT	GG	GG	GG
8	GG	TT	AA	GG	GG
9	GG	CC	GA	GG	GG
10	GG	CT	GA	GG	GG
11	GG	CC	GG	GG	GG
12	GG	CC	GG	GG	GG
13	GG	CT	GA	GG	GG
14	GG	CC	GA	GG	GG
15	GG	CC	AA	GG	GG
16	GG	CT	GA	GG	GG
17	GG	TT	AA	GG	GG
18	GG	TT	AA	GG	GG
19	GG	CC	GA	GG	GG
20	GG	CC	GG	GG	GG
21	GG	CT	GA	GG	GG
22	GG	CT	GG	GG	GC
23	GG	CT	GA	GG	GC
24	GG	TT	GG	GG	GC
25	GG	CC	GA	GG	GG
26	GG	CT	AA	GT	GG
27	GG	CC	GG	GG	CC
28	GG	CT	GA	GG	GC
29	GG	TT	GA	GG	GG
30	GG	CT	GG	GG	GG
31	GG	CT	AA	GG	GG
32	GG	CC	GG	GG	CC
33	GG	CC	GG	GG	GG
34	GG	TT	GG	GG	GG
35	GG	CT	GA	GG	GG
36	GG	CT	GA	GG	GC
37	GG	TT	GA	GG	GG
38	GG	CC	GA	GG	GG
39	GG	CT	GA	GG	GC
40	GG	CC	AA	GG	GG
41	GG	CT	GA	GG	GG
42	GG	CC	GA	GT	GC
43	GG	TT	AA	GG	GG
44	GG	CC	GG	GG	GC
45	GG	CT	GG	GG	GG
46	GG	CT	GG	GG	GC
47	GG	CC	GA	GG	GG
48	GG	CC	GG	GG	GC
49	GG	CT	GA	GG	GC

*Nucleotides in red font represent SNPs that are variants and may be associated with reduced cardiac function.

Table 4. SNP Genotypes* Produced from the ME Sample Group

Sample Number	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
6	GG	CC	GG	GG	GG
7	GG	TT	GA	GG	GC
13	GG	TT	GG	GG	GG
14	GG	TT	GA	GG	GG
19	GG	CC	GA	GG	GG
21	GG	CT	GA	GT	GG
24	GG	CC	AA	GG	GG
27	GG	CT	AA	GG	GG
28	GG	CT	GA	GG	GG
29	GG	CT	GG	GG	GG
30	GG	CC	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
40	GG	CT	GG	GG	GG

*Nucleotides in red font represent SNPs that are variants and may be associated with reduced cardiac function.

Table 5 summarizes the number of wild type and variant alleles in both population groups.

Table 5. Wild type and variant allele frequencies in the Normal and ME population groups*

Gene	Normal Grp. - Wt (%)	Normal Grp. Var. (%)	ME Grp - Wt (%)	ME Grp - Var. (%)
EYA4	62 (63.27%)	36 (36.73%)	17 (50.00%)	17 (50.00%)
EYA4-2	59 (60.20%)	39 (39.80%)	16 (47.06%)	18 (52.94%)
TNNI3	96 (97.96%)	2 (2.04%)	32 (94.12%)	2 (5.88%)
NEXN	80 (81.63%)	18 (18.37%)	32 (94.12%)	2 (5.88%)
MYH6	98 (100%)	0 (0%)	33 (97.06%)	1 (2.94%)

*Wt symbolizes wild type alleles whereas Var. represents variant alleles that may be associated with reduced cardiac function.

Results from Tables 3 and 4 are summarized in Table 5. 19.39% of all SNPs in the Normal population group represented variant alleles as opposed to 23.53% of all alleles in the ME group. Overall, there was a statistically significant difference in the frequency of wild type and variant

alleles in the Normal and ME population groups ($p < 0.001$) and thus there appears to be a significant association of variant alleles in the ME population group with death from cardiac-related causes.

The relative frequencies of wild type and variant nucleotides at SNP sites in the *EYA4, EYA4-2, NEXN*, and *TNNI3* genes are characteristic of di-allelic genetic markers that exist throughout the human genome. The working definition of a genetic marker (as opposed to a mutation) is that the frequency of the rarest allele must exceed 1% in the population.[23] Thus, all of the SNPs examined in this study, with the possible exception of the *MYH6* SNP, would qualify as di-allelic genetic markers rather than mutations. If the SNPs in the panel are treated as genetic markers, allele frequency databases can be created and used in calculations to predict genotype frequencies for individual as well as paired SNPs (Table 6).

Table 6. Allele frequency databases for alternate forms of each SNP in each gene

	Allele frequencies; Normal Population Group									
SNPs	***MYH6***		***EYA4***		***EYA4-2***		***TNNI3***		***NEXN***	
Wild type	G	0.9987	C	0.6327	G	0.6020	G	0.9796	G	0.8163
Variant	A*	0.0014	T	0.3673	A	0.3980	T	0.0204	C	0.1837

*The frequency of the A allele for the *MYH6* gene was obtained from the dbSNP database of NCBI since we didn't observe any variants in our limited population sampling (i.e. 49 individuals in the Normal population group).

Allele frequency databases for each SNP were created by simply counting the wild type and variant SNPs at each site using the Normal population group to represent the population at large. There were no variant alleles observed in the Normal population group for the *MYH6* SNP and, therefore, the minimal allele frequency for the A allele was obtained from the dbSNP catalogue of SNP frequencies available from the 1000 genome repository.[10,11] Allele frequencies can be used in conjunction with the Hardy-Weinberg (H-W) principle to predict genotype frequencies for each SNP in members of the population.[24] For a di-allelic genetic marker in equilibrium within the population (i.e., equilibrium as used here refers to a lack of association of any particular genotype for a marker with either a positive or negative effect on survival to reproductive age), the formula $p^2 + 2pq + q^2$ predicts the frequency of homozygous (i.e., p^2 and q^2) and heterozygous (2pq) phenotypes using the allele frequencies in the database. If the predicted genotype frequencies for a population agree with the observed genotype frequencies, the population can be considered to be in equilibrium. In contrast, if the predicted and observed genotype frequencies do not agree, the population is not in equilibrium and it can mean that particular genotypes may be associated with a selective advantage (or disadvantage) for survival of that genotype in the population. In these experiments we used allele frequencies derived from the Normal population group to predict genotype frequencies in the ME sample group. Chi square analysis comparing predicted and observed genotype frequencies for the individual SNPs in the Normal and ME population groups was performed and results are shown in Table 7.

Table 7. Chi square Analysis of Expected versus Observed Genotype Frequencies for SNPs in the Cardiac Gene Panel

SNP site	Expected v Observed (Normal) (p-values)	Expected v Observed (ME) (p-values)
MYH6 (G-A)	0.934	0.00007
EYA4 (C-T)	0.695	0.275
EYA4-2 (G-A)	0.761	0.089
TNNI3 (G-T)	0.989	0.262
NEXN (G-C)	0.439	0.161

Results of this analysis showed there is no statistically significant difference in the expected versus observed phenotype frequencies for any of the SNPs in the Normal population group but there is a deviation from population equilibrium for the *MYH6* SNP in the ME population group due to the detection of a single variant allele in one of the ME population samples (Table 4). Aside from this one deviation, the remaining SNP markers appear to be in equilibrium and there is no other apparent association of individual SNP phenotypes with cardiac-related death among members of the ME population group.

Our laboratory extended our population genetic analysis to include asking if there was an association of pairs of SNP genotypes with cardiac-related death in the ME population group. If two genetic markers are unlinked (i.e., the markers are independently inherited), then the frequency of one genotype can be multiplied by the frequency of the other to estimate the frequency of encountering both genotypes in a single individual. If two markers are linked or associated with one another, the predicted combined phenotype frequency will differ significantly from the observed frequency of the genotype pairs. Expected versus observed frequencies of all pairwise comparisons of SNP genotypes were evaluated using Chi square analysis (Table 8).

Table 8. Chi square Analysis of Expected versus Observed Paired Genotype Frequencies for SNPs in the Cardiac Panel

Paired genotypes	Population	p-Value
MYH6 / EYA4	Normal	0.833
	ME	0.288
MYH6 / EYA4-2	Normal	0.877
	ME	0.018
MYH6 / TNNI3	Normal	0.938
	ME	0.130
MYH6 / NEXN	Normal	0.504
	ME	0.094
EYA4 / EYA4-2	Normal	0.482
	ME	0.132
EYA4 / TNNI3	Normal	0.830
	ME	0.528
EYA4 / NEXN	Normal	0.722
	ME	0.020
EYA4-2 / TNNI3	Normal	0.908
	ME	0.012
EYA4-2 / NEXN	Normal	0.605
	ME	0.457
TNNI3 / NEXN	Normal	0.650
	ME	0.114

Expected versus observed genotype frequencies were not significantly different for all paired analyses among the Normal population group. However, within the ME group, three pairs of genotypes (*MYH6/EYA4-2, EYA4/NEXN, and EYA4-2/TNNI3)* were significantly different from expectations suggesting that SNP genotypes at these loci could be associated with cardiac-related death in the ME group. In addition, the *MYH6/NEXN* SNP combination was close to reaching a level of significance. It is possible that this SNP pair might have achieved significance if the number of individuals in the ME group had been higher. All SNP pair combinations that achieved significance involved the *EYA4-2* SNP even though it will be recalled that this SNP alone did not achieve a significant association with cardiac-related death (Table 7).

Discussion

The role of genetics in assessing the risk for cardiac-related episodes capable of leading to death, while generally acknowledged by the scientific community, is complex and has not yet led to a sufficiently clear understanding of how the many DNA sequence variants found in genes associated with cardiac function may underlie sudden cardiac death.[1-3,9,13] In a large study reported by Van Driest et.al., over 2000 individuals for whom complete medical records were available were tested for variant nucleotides in two genes, *SCN5A* and *KCNH2*, known to be involved with electrical signaling of the heart to maintain a normal rhythm.[13] Genetic testing

was performed by three independent laboratories with experience working with ion channel disorders. *SCN5A* and *KCNH2* genotypes were compared with the electronic medical records that included EKG records and other notations of cardiac issues. Results of the study identified over 100 nucleotide sequence variants in the two genes in about 10% of the study cohort.[13]However, there was low concordance among the three testing laboratories as well as the NIH database of nucleotide sequence variants (i.e., ClinVar) as to which variants were "pathogenic". Findings like these and others raise concerns about the value of genetic testing as a predictive tool to assess risk of SCD events in individuals and their surviving family members.

The answer to developing a reliable genetic test to predict risk for SCD may not lie in sequencing additional genes involved with cardiac function. Gene panels are presently available commercially from a number of providers that sequence 92 genes (Thermo Fisher, Waltham, MA), 172 genes (Illumina Corp., San Diego, CA), or even the entire exome (i.e., all of the coding sequences in the human genome).[9,17] Although sequencing more genes may seem a wise approach, in reality single nucleotide variants are common in the genome and the challenge still remains identifying which variants are causative for defects in cardiac function.

Some investigators have suggested that, like other disease conditions, SCD results from defects in a number of genes.[2,9] In fact, it is possible that when examining sequence variants in a collection of genes, each variant reduces the activity of the gene product only very little and therefore is not lethal, but multiple variants may together tip the scale of risk for an SCD event, especially during times of high demand on the heart. This possibility is supported by the data produced here in which genotypes composed for each of the five SNPs studied individually largely showed no significant association of genotype with cardiac-related death, even though there was a significantly greater number of variant alleles associated with those genotypes in the ME population group. The exception was the *MYH6* SNP for which a single variant nucleotide was detected in one of the ME population samples (out of 34 alleles total from 17 individuals) while no variant alleles were detected among 98 alleles in 49 individuals from the Normal population group. The frequency of the variant SNP in the *MYH6* gene in the population as a whole, as indicated by several SNP databases that are publically available, suggests a frequency of less than 1%. Therefore, the MYH6 marker would not qualify as a di-allelic genetic marker which might explain the deviation of the ME population group from population equilibrium for the *MYH6* SNP.

While the population analyses on most of the SNP markers individually did not suggest an association with cardiac-related death, the genotypes for some pairs of SNP markers did show a significant deviation from equilibrium in the ME population group suggesting an association of paired genotypes for the *MYH6/EYA4-2, EYA4-2/TNNI3,* and *EYA4/NEXN* markers. It would appear that the SNPs associated with the *EYA4* gene are important, they appear to be important when paired with the SNPs in *MYH6, TNNI3,* or *NEXN* genes as the *EYA4-2* SNP alone or paired with other genes does not show a significant association with cardiac-related death (Tables 5 and 6). If the number of individuals in the ME population group could be increased, it is possible that the association of other pairs of SNPs with cardiac-related death might become significant as several SNP pairs produced p-values close to 0.05. It might also be helpful to look for possible associations of three genes with cardiac-related death but the size of the ME population group was too small to make such a comparison meaningful.

The principal motivation for undertaking this study was to explore the possibility of developing a molecular assay that could be useful for the Medical Examiner to assist with identifying a cause of death for that percentage of victims for whom cause of death cannot be determined though cardiac-related causes are suspected. This is especially important for those deaths caused by defects in the electrical signaling of the heart (i.e.,channelopathies) since there is often no anatomic or histological evidence that such an event occurred.[4,9] The panel of markers chosen for this study all reside in genes involved with heart muscle contraction and therefore defects in these genes are more associated with cardiomyopathy which can often be diagnosed at autopsy. However, the detection of variant SNPs such as those studied here could nonetheless be informative for surviving family members who may unwittingly be at risk for similar, future cardiac problems. The choice for these markers rather than markers suggested to be involved with electrical signaling was based primarily on the autopsy findings for those individuals that ultimately became the ME population group, most of which were diagnosed with cardiovascular disease. Nonetheless, the findings presented here suggest an approach that could be used as part of the autopsy process to link associations of SNPs individually or as a group with particular disease states in ways that could be evaluated statistically for significance.

References

1. Semsarian C., Ingles J., and Wilde A., Sudden cardiac death in the young: the molecular autopsy and a practical approach to surviving relatives. *Eur. Heart J*. 2015;36:1290-1296.

2. Deo, R., Albert C.M., Epidemiology and genetics of sudden cardiac death. *Circulation*. 2012;125:620-637.

3. Towards European recommendations integrating genetic testing into multidisciplinary management of sudden cardiac death. Draft summary of an expert workshop of the European Society for Human Genetics held in Geneva, Switzerland, November 23-25, 2016. Published March 2018. https://www.eshg.org/fileadmin/eshg/consultations/Draft_Recommendations_SCD_March_2018_for_consultation.pdf

4. Tester, D.J., Ackerman, M.J., Postmortem long QT syndrome genetic testing for sudden unexplained death in the young. *J. Am Coll. Cardiol*. 2007;49:240-246.

5. Schwartz P.J., Crotti L., Insolia R., Long QT syndrome: from genetics to management. Circ. Arrhythm. *Electrophysiol*. 2012;5:868-877.

6. Cerrone M., Priori S.G., Genetics of sudden death: Focus on inherited channelopathies. *Eur. Heart J*. 2011;32:2109-2120.

7. Pfeufer A., Sanna S., Arking D.E., Muller M., Gateva V., Fuchsberger C., Ehret G.B., Orru M., Pattano C., Kottgen A., et.al. Common variants at ten loci modulate the QT interval duration in the QTSCD study. *Nat. Genet*. 2009;41:407-414.

8. Newton-Cheh C., Eijgelshelm M., Rice K., de Bakker P.I.W., Yin X., Estrada K., Bis, J., Marciante K, Rivadeneira F., Noseworthy P.A., et.al., Common variants at ten loci influence myocardial repolarization: the QTGEN consortium. *Nat. Genet*. 2009;4:399-406.

9. Schwartz P.J., Ackerman M.J., George A.L., Wilde A.A.M, Impact of genetics on the clinical management of channelopathies. *J. Am. Coll. Cardiol*. 2013;62:169-180.

10. Neubauer J., Lecca M.R., Russo G., Bartsch C., Medeiros-Domingo A., Berger W., Haas C., Post-mortem whole-exome analysis in a large sudden infant death syndrome cohort with a focus on cardiovascular and metabolic genetic diseases. *Eur. J. Human Genet*. 2017;25:404-409.

11. Sudmant P.H., Rausch T., Gardner E.J., Handsaker R.E., Abyzov A., Huddleston J., Zhang Y., et.al. An integrated map of structural variation in 2504 human genomes. *Nature*. 2015;526:75-81.

12. Auton A., Abecasis G.R., A global reference for human genetic variation. *Nature*. 2015;526:68-74.

13. Giudicessi J.R., Iftikhar J.K., Ackerman M.J., Precision cardiovascular medicine: state of genetic testing. *Mayo Clinic Proc*. 2017;92:642-662.

14. Van Driest SL, Wells Q.S., Stallings S., Bush W.S., Gordon A., Nickerson D.A., Kim J.H., Crosslin D.R., Jarvik G.P., et.al., Association of arrhythmia-related genetic variants with phenotypes documented in electronic medical records. *JAMA*, 2016;5:47-57.

15. Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Hoover J, Jang W, Katz K, Ovetsky M, Riley G, Sethi A, Tully R, Villamarin-Salomon R, Rubinstein W, Maglott DR. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res* . 2016 Jan 4;44(D1):D862-8. doi: 10.1093/nar/gkv1222. PubMed PMID:26582918

16. Schonberger J., Wang L., Shin J.T., Kim S.D., Depreux F.F., Zhy H., Zon L., Pizard A., Kim J.B., MacCrae C.A. et.al. Mutation in the transcriptional activator *EYA4* causes dilated cardiomyopathy and sensorineural hearing loss. *Nat. Genet*. 2005;37:418-422.

17. Stelzer G, Rosen R, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, Iny Stein T, Nudel R, Lieder I, Mazor Y, Kaplan S, Dahary D, Warshawsky D, Guan - Golan Y, Kohn A, Rappaport N, Safran M, and Lancet D. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analysis. *Current Protocols in Bioinformatics.* 2016;54:1.30.1 - 1.30.33.doi: 10.1002 / cpbi.5. Karki

18. Razmara E., Garshaski M., Whole-exome sequencing identifies R1279X of *MYH6* to be associated with congenital heart disease. Cardiovasc. Disord. 2018;18:137.

19. Posch M.G., Waldmuller S., Muller M., Scheffold T., Fournier D., Andrade-Navarro M.A., De Geeter B., Guillaumont S., Dauphin C., Yousseff D., et.al. Cardiac alpha myosin (*MYH6*) is the predominant sarcomeric disease gene for familial atrial septal defects. *PLoS One*. 2011;6:e28872, doi: 10.137/journal.pone.0028872.

20. Chen, Y., Yang S., Li J., Wang G., Cao K., Pediatric restrictive cardiomyopathy due to a heterozygous mutation of the *TNNI3* gene. *J. Biomed. Res*. 2014;28:59-63.

21. Nunez L., Gimero-Blanes J.R. Rodriguez-Garcia M.I., Monserrat L., Zorio E., Coats C, McGregor C.G., Hernandez del Rincon J.P., Castro-Bieras A., Hermida-Prieto M., Somatic *MYH7, MYBPC3, TPM1, TNNT2*, and *TNNI3* mutations in sporatic hypertropic cardiomyopathy. *Circ. J.* 2013;77:2358-2365.

22. Wang H., Li Z., Sun K., Cu, Q., Song L., Zou Y., Wang X., Liu X., Hul R., Fan Y., Mutations in *NEXN*, a Z-disc gene, are associated with hypertropic cardiomyopathy. *Am. J. Hum. Genet*. 2010;87:687-693.

23. Hassel D., Dahme T., Erdmann J., Meder B., Huge A., Stoll M., Just S., Hess A., Ehlermann P., Weichenhan D., Grimmier M., et.al. Nexillin mutations destabilize cardiac Z-disks and lead to dilated cardiomyopathy. *Nat. Med*. 2009;15:1281-1288.

24. R., Pandya D., Elston R.C., Ferlini C., Defining “mutation” and “polymorphism” in the era of personal genomics. *BMC Medical Genomics*. 2015;8:37-44.

25. Hartl D.L., Clarke A.G. *Principles of population genetics*. Sinauer, Sunderland, MA, c2007

Acknowledgements
The authors wish to thank the Office of the Chief Medical Examiner and Dr. Josh Lanter in particular for providing the blood stains representing the ME population group and for critical review of the manuscript.