EVALUATION OF THE GENETIC ARCHITECTURE OF SUDDEN CARDIAC ARREST

By
Rebecca N. Mitchell

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
September 2019

Abstract

Sudden cardiac arrest (SCA) is one of the leading causes of death in the United States. While several risk factors are observationally associated with SCA, the genetic architecture of SCA in the general population remains unknown. Furthermore, understanding which risk factors are causal may help target prevention strategies. Given that SCA is a complex disease with a heterogeneous makeup of risk factors and underlying causal diseases that may differ by both race and sex, studying the genetics of this disease has proved challenging. Here, we use genome-wide association studies (GWAS) to identify common genetic variants associated with SCA risk. We performed race-, sex-, and disease-stratified GWASs in an attempt to create a more homogenous phenotype in order to identify variants associated with SCA risk. However, we were not able to identify any common genetic variants associated with SCA.

We utilized Mendelian randomization to identify causal risk factors for SCA, identifying CAD status, BMI and QT interval as being causally associated. We further investigated the genetic association between prolonged QT interval and SCA risk in sex-stratified and disease-stratified (ischemic vs. non-ischemic disease) analyses. We tested for association between the top QT interval associated SNP, rs12143842 (in the *NOS1AP* locus), and SCA risk. We also tested for causal association of QT interval in the various subgroups. We found that non-ischemic individuals, particularly women with non-ischemic disease, showed the strongest association between rs12143842 and SCA risk and the strongest causal association. Ischemic SCA victims, irrespective of sex, did not

show an association between rs12143842 and SCA risk or a causal association for QT interval.

This work sought to further our understanding of the genetics of SCA and its underlying etiology. While we were not able to identify any common genetic variants associated with SCA, we did identity several causal risk factors for SCA and found these causal risk factors may differ by both underlying disease and sex. Furthering our knowledge of the genetics of SCA, its etiology, and its causes will ultimately lead to better identification of higher risk individuals.

Advisor: Dan E. Arking, Ph.D.

Reader: Marc Halushka, M.D., Ph.D.

Preface

This thesis is a culmination of several years of work and would not have been possible without the support and guidance of many different people. First and foremost, I want to thank my advisor, Dr. Dan Arking, for allowing me, someone with no coding experience, to join his lab and mentor me over these past four years. Dan has taught me so much about computational genetics and biostatistics with never-ending patience. I will always appreciate his ability to think critically and quickly about any question and/or problem I encountered. His love of science and his drive to answer scientific questions interesting to him are traits I hope to emulate in my career going forward. I would not be the scientist I am today without his guidance and teaching and for that I am most thankful.

I also want to thank other members of the Arking lab who have helped me throughout my graduate school career. The early lab members, including Foram Ashar, Shannon Ellis, Nathan Bihlmeyer and Anna Moes, were always so helpful and willing to answer any dumb coding or methods question I had when I first joined the lab. Ryan Longchamps, my MD GEM partner in crime, was immensely helpful in the lab and a good friend who was always willing to listen to my complaints and stresses. The newer members of the lab, including ThuyVy Duong and Stephanie Yang, have become such great friends and colleagues and immensely helped me through the ups and downs of graduate school. Finally, I would not be where am I today without Christina Castellani. We both joined the lab at about the same time and became fast friends. Her scientific mentorship and

friendship have played a crucial role in what I was able to accomplish in graduate school.

Outside of the Arking lab and Hopkins, I interacted with many different collaborators and scientists from all over the world that have contributed to my training. I want to thank in particular Nona Sotoodehnia, at the University of Washington, for being a great mentor and friend to me over the past few years. Her willingness to contribute to my research has extended beyond the traditional collaborator role and I have greatly appreciated the interest she's taken in my life and future career. I also wanted to thank Juhani Junttila, at the University of Oulu, for his patience and expertise over the past few years.

I also wanted to thank my thesis committee, including Dr. Marc Halushka, Dr. Rasika Mathias, and Dr. Ellseo Guallar, for their support and scientific mentorship. Their expertise helped shape my projects and they were always willing to assist me even outside of committee meetings. I also wanted to thank the Human Genetics program and Dr. Valle and Sandy for their support of my graduate training. I also wanted to thank the MD GEM program, particularly Priya Duggal and Jennifer Deal, for providing me the advanced training necessary for my graduate work in biostatistics and genetic epidemiology. I also appreciate their willingness to mentor me and provide advice in areas beyond science.

Outside of science, I have a fantastic support group of people in my life. My poker group, including Daniel, Diana, Dorothy, Kenny and Hannah, has been so supportive and we've had so much fun over the past few years becoming great friends and having babies! ☺ I am very excited we are all staying in Baltimore for

the foreseeable future and look forward to many more fun times! Also, my "mom" group of friends, whom I met through a moms book club, has provided so much support as a new mom and many, many glasses of wine! I wouldn't be the mom I am today without our group text message chain!

I also want to thank my family, in particular my parents, for their support through graduate school. They were always so encouraging and excited about my decision to attend graduate school. I wouldn't be here without their support! I also want to thank my other family, my in laws. The love and assistance you have provided me is irreplaceable. I am so happy to have married into such a wonderful family!

Finally, I want to thank my children, Ella and Lyla, for always loving me. I hope to make you proud and one day perhaps you will also be scientists! Lastly, but certainly not least, I want to thank my wonderful husband, Jason. I would not have survived without his rock-solid love and support. I cannot even begin to list everything he has done for me; he made so many sacrifices for me to attend graduate school, including moving to Baltimore and changing jobs. Thank you for taking such great care of our family and me during these last several years. We couldn't have made it without you!

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Sudden cardiac arrest (SCA) is a major public health concern, affecting ~300,000 individuals in the US every year.[1] While there are many different definitions of SCA, the American Heart Association defines SCA as "unexpected cessation of cardiac mechanical activity, as confirmed by the absence of signs of circulation."[1] SCA is commonly the result of ventricular fibrillation and is often the first manifestation of heart disease.[2,3] Currently, the biggest risk factor for SCA is reduced ejection fraction (EF). The ejection fraction is the percentage of blood that is pumped out of the left ventricle with each heartbeat. Normal EF ranges from 50-65% and reduced EF is generally considered to be less than 35%.[4] A meta-analysis of twenty different studies found that an EF <30% to <40% was associated with a relative risk of 4.3 for major arrhythmic events.[5] However, given the majority of individuals who have an SCA event have a normal ejection fraction, the utility of EF as a predictor for SCA is highly limited. In addition, observational studies have identified numerous other clinical and subclinical risk factors for SCA; understanding which of these associations are causal will help target prevention strategies. Identifying individuals at increased risk for SCA, as well as finding preventative and therapeutic measures, is necessary to reduce the impact of this largely fatal disease.

Several studies have identified family history of SCA as a strong predictor of SCA in the general population, suggesting that genetic variation may influence SCA risk. One study found that a family history of SCA (<65 years of age) was associated with 2.7-fold increased risk for SCA after adjusting for history of MI

and other CVD risk factors.[6] The Paris Prospective study of ~7,000 individuals found a history of one parent with an SCA event was associated with a relative risk of 1.89 of developing SCA and a history of both parents with an SCA event was associated with a relative risk of 9.44 of developing SCA.[7] Identifying genetic markers associated with SCA risk could provide additional risk stratification beyond the traditional risk factors mentioned above.

However, identifying genetic variation associated with SCA risk has proven to be difficult and the findings have been mixed. Studies first utilized candidate gene approaches, investigating SNPs within genes that are known to be associated with SCA in individuals with different inherited arrhythmia conditions such as Long QT Syndrome and Brugada Syndrome. Smaller studies have found both rare and common variants within these genes associated with SCA risk in the general population.[8–11] However, we recently published a large study (3,939 cases and 25,989 controls) in which we did not find evidence that common variation in Mendelian arrhythmia genes is associated with SCA risk in the general population.[12]

To expand the search for common genetic variants associated with SCA outside of known arrhythmia genes, our lab previously published the first genome-wide association study (GWAS) of individuals of European descent consisting of 1,283 SCA cases and >20,000 controls.[8] One SNP within the *BAZ2B* gene reached the genome-wide significance threshold ($5x10^{-8}$). However, we recently published a larger GWAS of European descent individuals, consisting of 3,939 SCA cases and 25,989 controls, which did not replicate this finding.[12]

Further, we failed to identify any variant that reached the genome-wide significance threshold ($P<5x10^{-8}$). In addition to the lack of power, the heterogeneity of the SCA phenotype may also cause the failure to replicate the previous findings and lead to the lack of additional findings.

SCA occurs as a result of multiple underlying disease pathologies, including heart diseases such as coronary artery disease (CAD) and cardiomyopathies, as well as primary electrical defects.[13] Men have a higher risk of SCA than women[14,15], and furthermore, the underlying cardiac pathology differs between the sexes. CAD, the common underlying cause of SCA, is more common in men than women. By contrast, non-ischemic pathology, such as primary myocardial fibrosis, valvular heart disease, and arrhythmogenic right ventricular cardiomyopathy, occurs more commonly in women with SCA compared to men with SCA.[16–18] Given these known differences, it can be hypothesized that risk factors, in addition to genomic markers, associated with SCA may differ between both the underlying causal disease and sex.

The results of our two genome-wide association studies may provide evidence for this hypothesis. The first GWAS for SCA of 1,283 cases and >20,000 controls included almost exclusively SCA individuals with underlying CAD from 5 different studies, given the majority of the studies explicitly excluded SCA cases with non-ischemic underlying disease. The second GWAS of 3,939 cases and 25,989 controls included SCA individuals from 9 different studies in which studies did not exclude cases with non-ischemic disease, such as valvular heart disease and cardiomyopathy. While the second GWAS had a larger

sample size, the additional cases of different underlying disease may be contributing to lack of findings if the associated genetic variation with SCA risk differs by underlying disease. Previously, stratifying cases in genome-wide analyses of complex diseases by subtype or risk factor have been successful in identifying genetic variants specific to the subgroups. Some of these stratified analyses include: ischemic stroke stratified by cases with different pathologies (cardioembolic, small vessel and large vessel)[19]; migraine stratified by cases with and without aura[20]; type 2 diabetes stratified by BMI of cases[21]; and rheumatoid arthritis stratified by the presence or absence of antibodies to citrullinated peptide antigens (ACPA)[22]. Therefore, we hypothesize that stratifying SCA cases by underlying disease may create a more homogenous SCA phenotype and increase our ability to detect genetic variants associated with SCA risk in these subgroups and the associated genetic variants may differ between the subgroups.

In addition to underlying disease and sex, SCA risk also differs by race. Studies have found African Americans are disproportionately affected by SCA in every age group.[23–25] There are likely several factors contributing to this increased risk. African American individuals have higher rates of traditional cardiac risk factors, including hypertension[26], left ventricular hypertrophy[27], heart failure, obesity and type 2 diabetes.[28] These higher rates of cardiac risk factors and SCA may also be a result of large racial disparities in health care; one study found that African Americans are more likely to not survive an in-hospital cardiac arrest compared to Caucasians, indicating the lower survival rate among African Americans may be due to the quality of care they receive at a specific medical

center.[29] African American individuals are also less likely to receive an implantable cardioverter defibrillator (ICD) for primary prevention of SCA.[30] While these factors all likely contribute to the racial differences in SCA risk, we wanted to examine whether genetics plays a role in the differences as well. Given no GWAS for SCA has ever been performed in African American individuals, it is unknown whether the genetic markers associated with SCA risk would differ between races.

All of these factors, race, sex and underlying disease, have contributed to the difficulty in studying the genetics of SCA. Given our null results from a GWAS of relatively large sample size, there is evidence that there are no common genetic variants of large effect associated with overall SCA risk in the European population. However, we hope to identify genetic variants of moderate or lower effect on SCA risk by increasing the sample size of the European ancestry GWAS. We also performed sex-stratified GWAS for individuals of European descent. In addition, we performed a GWAS for SCA in African American individuals for the first time, to identify any genetic markers associated with SCA. Finally, we utilized a large Finnish study of SCA individuals with post-mortem autopsy-confirmed underlying disease to study the disease-specific (ischemic vs. non-ischemic disease) genetic variants associated with SCA risk.

To identify causal traits of SCA, we performed Mendelian randomization, a method utilizing associated genetic variants with known SCA risk factors. Lastly, we utilized our large Finnish study of post-mortem autopsy-confirmed SCA individuals to stratify by underlying causal disease (ischemic vs. non-

ischemic disease) and sex to study the differences in causal risk factors with SCA between these different groups. Overall, our goal was to further our knowledge of the genetics of SCA and better understand it's underlying etiology.

Chapter 2: Using GWAS to identify loci associated with Sudden Cardiac Arrest

2.1 Introduction

Race appears to play a significant role in the development of SCA; studies have found individuals of African descent have a higher risk of SCA than individuals of European descent.[28,31] There are also large racial disparities in cardiac health care and higher rates of traditional cardiac risk factors among African Americans that could also lend to the observed increased rates of SCA.

While these health disparities do lead to the increased risk of SCA observed in the African American population, there is also some evidence that genetic variation may also lead to increased risk. Studies have found differences between ethnicities in allele frequency for encoding sodium and potassium channels which are associated with J-point elevation, Brugada syndrome, and Long QT syndrome.[28] However, there is also evidence that the specific genetic variants associated with SCA risk may be shared among the different ethnicities, although the allele frequency and effect size may differ. For example, a QT interval GWAS in African American individuals found the *NOS1AP* locus SNP rs12143842, a SNP also previously found to be associated with SCA risk, as the top QT interval-associated SNP.[32] This is the same top QT interval-associated SNP found in the QT interval GWAS in European individuals.[33] It is likely that there is both race-specific and non-race specific genetic variation associated with SCA risk.

Given that men have a higher risk of SCA than women[14,15] and the underlying causal disease differ by sex, in addition to race, genetic variation associated with SCA risk may also differ by underlying disease pathology and sex. Genome-wide association studies (GWAS) are a useful tool to identify common genetic variation associated with SCA risk. Identifying genomic risk markers for SCA will help classify those individuals at high risk in the general population. Previously, we performed a GWAS for SCA in European-descent individuals using 3,939 SCA cases and 25,989 controls to identify potential genetic variation associated with SCA risk.[12] However, no SNP reached the genome-wide significance threshold ($P<5x10^{-8}$). Here, we ran a GWAS for SCA using individuals of European descent with an additional 1,351 SCA cases and 1,008 controls (a total of 5,290 SCA cases and 26,997 controls), as well as a sex-stratified GWAS. We also ran a GWAS for SCA using African American individuals (974 SCA cases and 3,526 controls). Finally, we performed a trans-ethnic GWAS for SCA using European (5,290 SCA cases/26,997 controls), African American (974 SCA cases/3,526 controls), and Asian (152 SCA cases/176 controls) individuals for a total of 6,416 SCA cases and 30,699 controls.

In addition to perform race-specific GWASs, we also stratified by underlying disease. We used a large Finnish study of individuals who died from an SCA event with a post-mortem autopsy that confirmed the underlying causal disease (ischemic vs. non-ischemic) as well as Finnish population controls. We hypothesized that there are genetic differences between those individuals with

diagnosed CAD and individuals with CAD whom go on to have an SCA event. To test this hypothesis, we used two control cohorts of individuals with diagnosed CAD to perform the GWAS. Finally, we stratified by underlying disease pathology in order to create a more homogenous phenotype that may result in the improved ability to detect genetic variants associated with SCA risk.

2.2 Methods

2.2.1 Study Population and Phenotype Definition

2.2.1a Race-stratified and trans-ethnic analyses

We conducted a GWAS for individuals of European descent using 9 studies consisting of 5,290 SCA cases and 26,997 controls. We also conducted a GWAS for individuals of African descent using four different studies consisting of 974 SCA cases and 3,526 controls. Finally, we performed a trans-ethnic GWAS consisting of the studies of individuals of European descent, African American descent, as well as one study of individuals of Asian descent (152 SCA cases and 176 controls) for a total of 6,416 SCA cases and 30,699 controls. Study descriptions, study-specific SCA definitions and genotyping methods are detailed in **Table 2.1 and 2.2**.

2.2.1b Underlying causal disease stratified analyses

Fingesture

The Fingesture study, started in 1998, aimed to collect consecutive victims of out-of-hospital sudden death from a defined geographical area, Oulu University Hospital District in northern Finland. All victims of sudden death were autopsied at the Department of Forensic Medicine, University of Oulu, Oulu, Finland. SCA

victims were defined as those with a witnessed sudden death within 6 hours of the onset of the symptoms or within 24 hours of the time that the victim was last seen alive in a normal state of health. Individuals with age at SCA event <30 years old or >80 years old were excluded from analysis.

The underlying pathologies were divided into three categories: (1) ischemic, (2) non- ischemic, and (3) other disease. The ischemic SCA victims included individuals with evidence of a coronary complication, defined as a fresh intracoronary thrombus, plaque rupture or erosion, intraplaque hemorrhage, or critical coronary stenosis (>75%) in the main coronary artery. The non-ischemic SCA victims included individuals with the following conditions: hypertrophy due to hypertension; valve disease; cardiomyopathy due to alcohol use; dilated cardiomyopathy; hypertrophic obstructive cardiomyopathy; cardiomyopathy due to obesity; arrhythmogenic right ventricular cardiomyopathy; and primary myocardial fibrosis. Further definitions of these conditions have been previously described.[16] The "other" SCA victims included individuals with the following conditions: myocarditis, cardiac anomaly, and normal autopsy individuals (e.g. individuals with a channelopathy).

NFBC1966

The Northern Finland Birth Cohort (NFBC) study is the product of a project initiated in the 1960s to examine risk factors involved in pre-term birth and intrauterine growth retardation, and the consequences of these early adverse outcomes on subsequent morbidity. The NFBC1966 cohort comprised of 12,068 mothers and 12,231 children with an expected date of birth in 1966 within the

province of Oulu, Finland. Our study samples consisted of DNA extracted from the blood of the offspring at their 31-year follow-up visit.

Artemis

ARTEMIS is a prospective observational study (Innovation to Reduce Cardiovascular Complications of Diabetes at the Intersection; ClinicalTrials.gov identifier NCT01426685) that recruited patients with angiographically documented CAD, with or without type 2 diabetes. The study population was recruited from a series of patients enrolled in the coronary angiography registry at the Division of Cardiology, Oulu University Hospital, between August 1, 2007 and December 31, 2012. The initial examinations and determination of inclusion/exclusion status were conducted at least 3 months after coronary angiography and/or the last revascularization.[34]

MRFAT

The Multiple Risk Factor Analysis Trial (MRFAT) study population consisted of enrollees in a prospective post-MI study, enrolled between January 1996 and January 2000. The diagnosis of MI was based on the presence of at least two of three criteria from ICD-10: elevated troponin levels, ECG findings, and typical angina pectoris.

2.2.2 GWAS

For the European studies, genome-wide genotype data was imputed to either the HapMap2-CEU reference panel or the TOPMed Freeze 5 reference panel. Studies that provided sex-stratified data were utilized in the sex-specific GWAS meta-analysis as indicated in **Table 2.1**. Genome-wide genotype data for

the African American studies was imputed to the TOPMed Freeze 5 reference panel. Genome-wide genotype data for the study of individuals of Asian descent was imputed to 1000G Phase1v3. Each study performed regression analyses adjusting for sex and other study-specific covariates. Inverse variance meta-analysis using METAL[35] was performed for the European-only and African American-only meta-analyses. For the European-only GWAS, a SNP had to present in at least 4/10 of the studies to be included in the final analysis.

The trans-ethnic meta-analysis was performed using the software Metasoft. Metasoft uses three different meta-analysis methods: (1) fixed effect model; (2) random effects model; and (3) Han and Eskin's random effects model.[36] The third model, Han and Eskin's random effects model, differs from the traditional random effects model in that it assumes no heterogeneity under the null hypothesis. Therefore, in the presence of heterogeneity, this model is more powerful than the traditional random effects model. A SNP had to be present in at least 4/16 of the studies to be included in the final analysis.

The underlying disease stratified GWASs were performed using 1,171 Fingesture (614 ischemic and 557 non-ischemic) SCA individuals; 761 population controls (NFBC1966); and 1,015 ischemic controls (455 Artemis and 560 MRFAT). Samples were run on either the Illumina Infinium Global Screening Array (GSA) or the Affymetrix Genome-wide Human SNP Array 6.0 and imputed to the TOPMed Freeze 5 reference panel. Three different analyses were performed: ischemic SCA individuals vs. population controls; ischemic SCA

individuals vs. ischemic controls; and non-ischemic individuals vs. population controls. We performed logistic regression using FASTv2.4.[37]

2.3 Results

2.3.1 Race-specific GWAS

After meta-analysis of the nine studies of individuals of European descent (5,290 SCA cases and 26,997 controls), no SNP reached genome-wide significance ($5x10^{-8}$). In addition, no SNP reach genome-wide significance for the meta-analysis of the four studies of individuals of African American descent (974 SCA cases and 3,526 controls). The QQ plot and Manhattan plot for these race-specific meta-analyses are shown in **Figures 2.1 and 2.2**, respectively**.**

2.3.2 Trans-ethnic GWAS

There were no SNPs that reached the genome-wide significance threshold under the fixed effects and traditional random effects models. One SNP, rs10803352, reached genome-wide significance using the Han and Eskin's random effects model. The QQ plots and Manhattan plots for these meta-analyses are shown in **Figures 2.3 and 2.4**, respectively**.** The forest plot for the top SNP (rs10803352) is shown in **Figure 2.5.** However, while the SNP achieved genome-wide significance, it is clear the association is driven by one of the studies. Given this association is not replicated in the other studies, this is likely a false positive association, and therefore we concluded that no true positive SNP reached the genome-wide significance threshold for meta-analysis using the Han and Eskin's random effects model.

2.3.3 Sex-specific GWAS

Neither the men nor women-specific GWAS had SNPs that reached the genome-wide significance threshold. The QQ plots and Manhattan plots for these meta-analyses are shown in **Figures 2.6 and 2.7**.

2.3.4 Disease-specific GWAS

None of the disease-specific GWASs identified any SNP associated with SCA risk at the genome-wide significance threshold. QQ plots and Manhattan plots are shown in **Figures 2.8 and 2.9**. However, the top SNP from each GWAS (ischemic cases/population controls; ischemic cases/ischemic controls; non-ischemic cases/population controls) were unique to each GWAS. Regression results for each top SNP in all analyses are found in **Tables 2.3** (ischemic cases/population controls), **Table 2.4** (non-ischemic cases/population controls), and **Table 2.5** (ischemic cases/ischemic controls).

2.4 Discussion

SCA is a complex disease with a heterogeneous make-up of risk factors and underlying causal diseases. This heterogeneous nature contributes to the difficulty in identifying genetic markers associated with SCA risk. Our previous attempt to identify common genetic variation associated with SCA risk in individuals of European descent using the largest GWAS (3,939 cases and 25,989 controls) failed to identify any SNP of genome-wide significance.[38] Here, with the addition of new European SCA cases and controls, while we did identify SNPs of greater significance than the previous study, we still failed to identify any common genetic variation associated with SCA risk that exceeded the genome-

wide significance threshold in individuals of European descent. We also performed the largest GWAS to date for SCA risk using African American individuals. Similar to the European GWAS, we were unable to identify any SNPs that reach genome-wide significance.

In the trans-ethnic GWAS, even when utilizing different statistical models, we were unable to identify any plausible SNPs associated with SCA risk. While the Han and Eskin's random effects model found one SNP that exceeded the genome-wide significance threshold, it was largely driven by one study of African American individuals, and therefore likely a false positive finding. This SNP also did not replicate in the African American GWAS, lending additional evidence this is a false positive.

Failure to identify any associated SNPs could be due to several different factors. First, the SCA phenotype is broadly defined and each study classifies SCA somewhat differently. This phenotypic heterogeneity could potentially contribute to differences in the underlying genetics; therefore, any SNPs identified through this GWAS will have to be the shared risk variants across the different SCA phenotypes. Identifying the underlying disease of SCA individuals and analyzing the different diseases separately would provide a more homogenous phenotype and the ability to identify genetic variants associated with SCA risk within that phenotype. Second, the genetic susceptibility to SCA may be derived of multiple variants of moderate or low effect size and therefore our sample size is not large enough to detect these effect sizes. Given that the African American GWAS has fewer than 1,000 SCA cases, a lack of power is likely a major factor

in the failure to identify genome-wide significant SNPs for that particular racial group. Additional samples would assist in increasing power to detect these variants of moderate and lower effect sizes. Third, the causal SNP(s) may not be available in the earlier reference panels used in the imputation of most of the studies. Recently, the TOPMed reference panel became available for use with imputation. We were able to use it for imputation for 6 of the 16 studies used in the various meta-analyses. The TOPMed reference panel provides significantly more SNPs (>7 million post QC) than older reference panels (typically ~2 million SNPs). Updating the other studies to imputation using the TOPMed reference panel would provide analyses on a significantly larger number of SNPs, increasing the power to detect associations with these SNPs. All of these limitations likely contribute to the lack of genome-wide significant SNPs associated with SCA risk in our three GWASs.

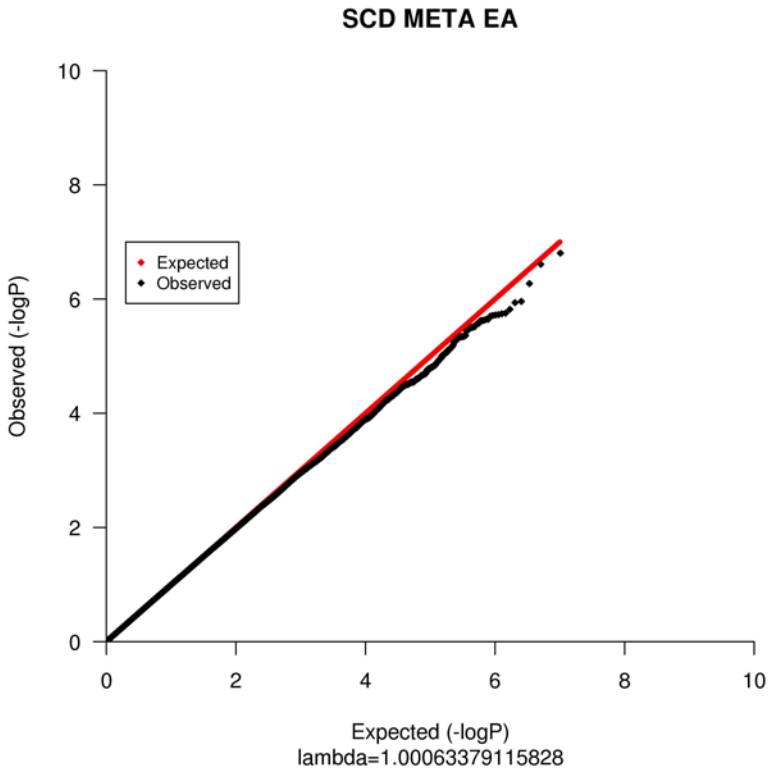
In conclusion, we did not identify any genome-wide significant SNPs in any of our analyses, likely due to lack of power and phenotypic heterogeneity. However, when stratifying by underlying causal disease, we did see more significant results and the top SNPs were unique to each disease subtype, indicating the genetic variation influencing SCA risk is likely specific to each underlying disease. Including new SCA cases, as well as identifying the underlying causal disease in order to create a more homogenous phenotype in the existing studies, and imputing all data to the TOPMed reference panel are all ways to improve potential detection of common genetic variants associated with SCA risk. Ultimately, identifying genetic risk markers for SCA will improve

detection of individuals at increased risk in the general population, regardless of race.

2.5 Figures

Figure 2.1: QQ plots for race-specific GWASs for SCA

A. European ancestry



B. African ancestry

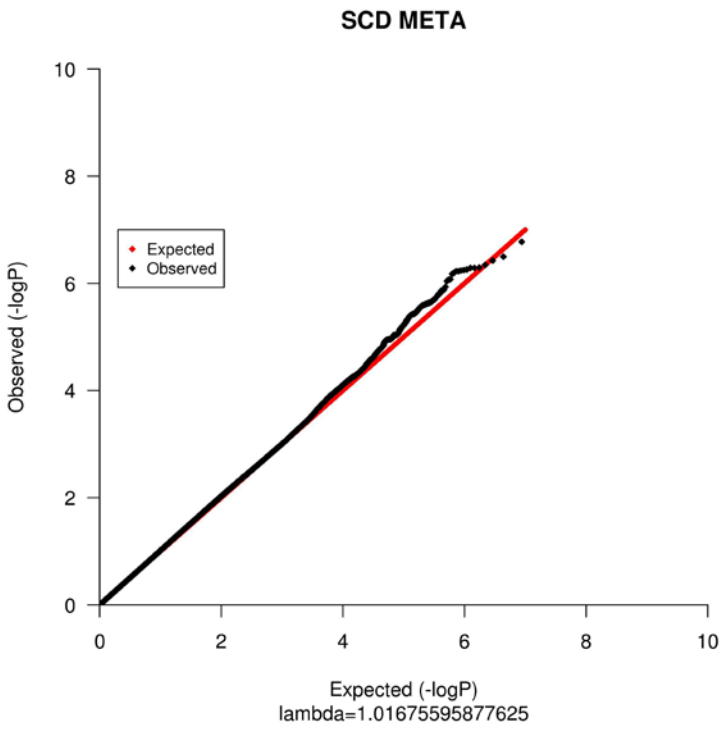
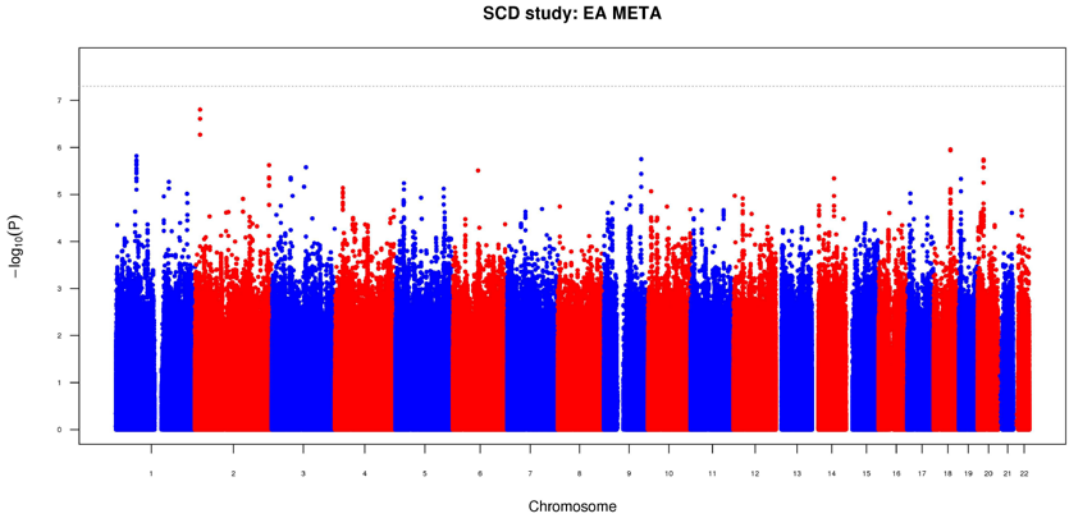


Figure 2.2: Manhattan plots showing results for race-specific GWASs for SCA

A. European ancestry



B. African ancestry

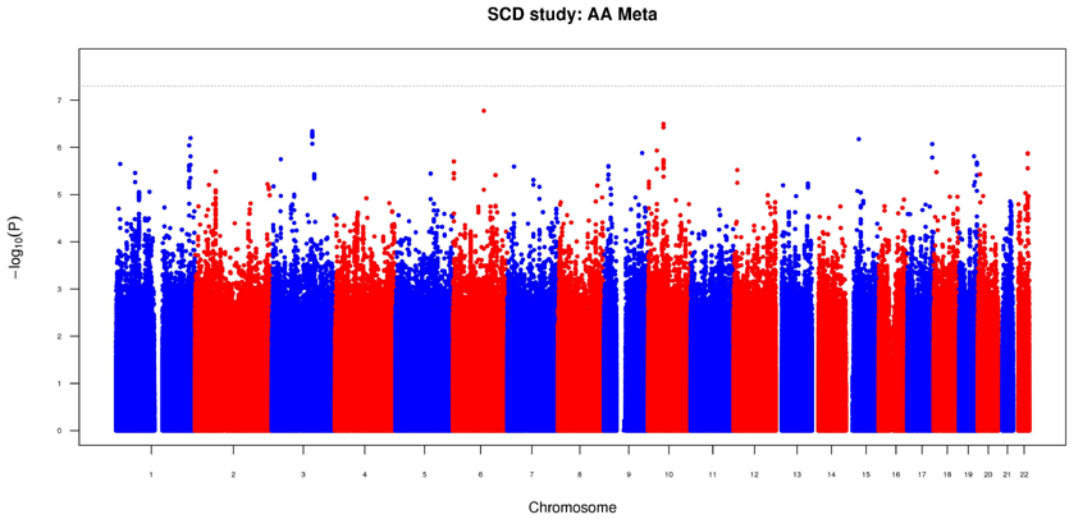
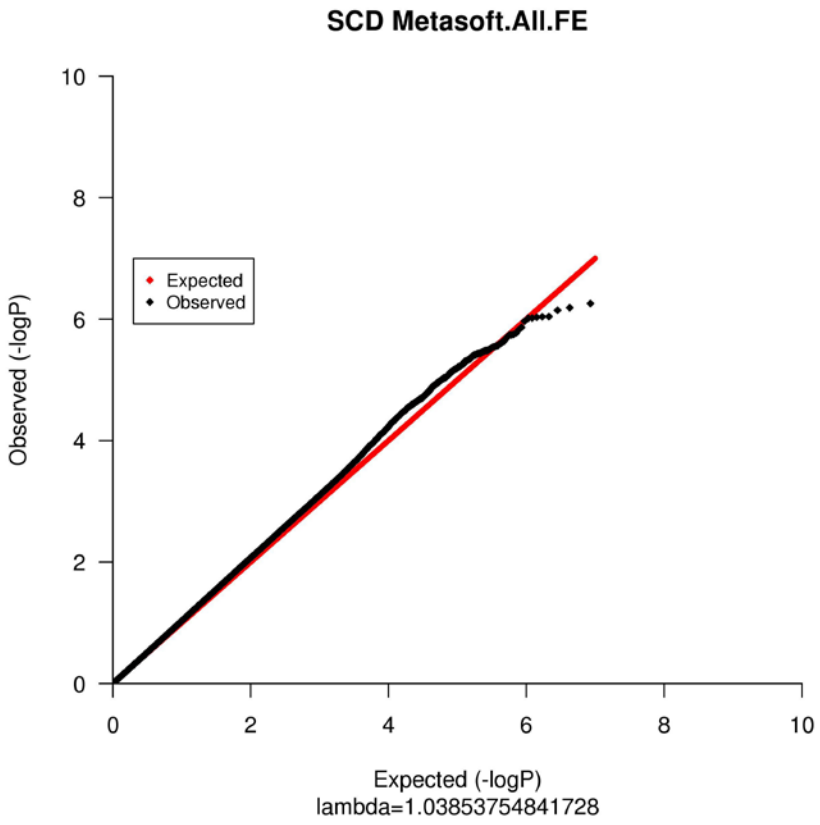
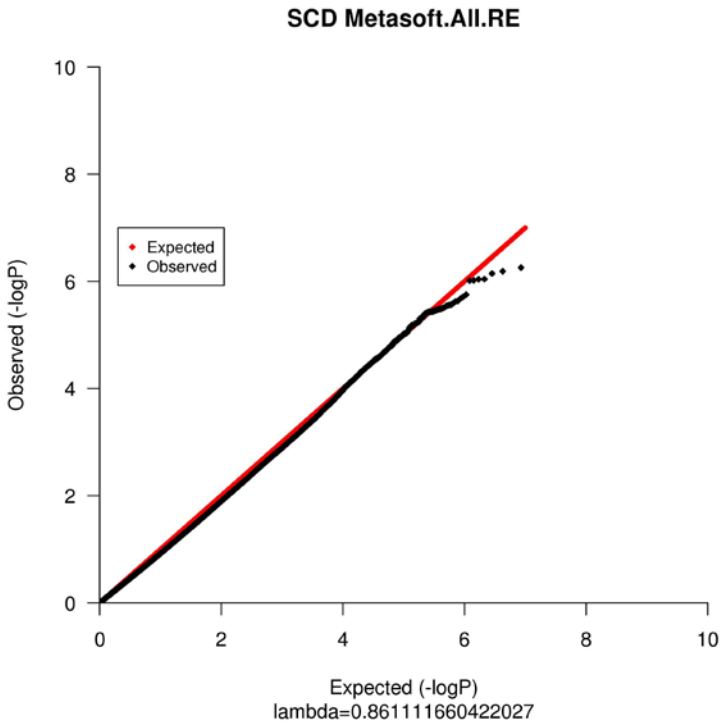


Figure 2.3: QQ plots for trans-ethnic GWAS for SCA

A. Fixed effects model



B. Random effects model



C. Han and Eskin's random effects model

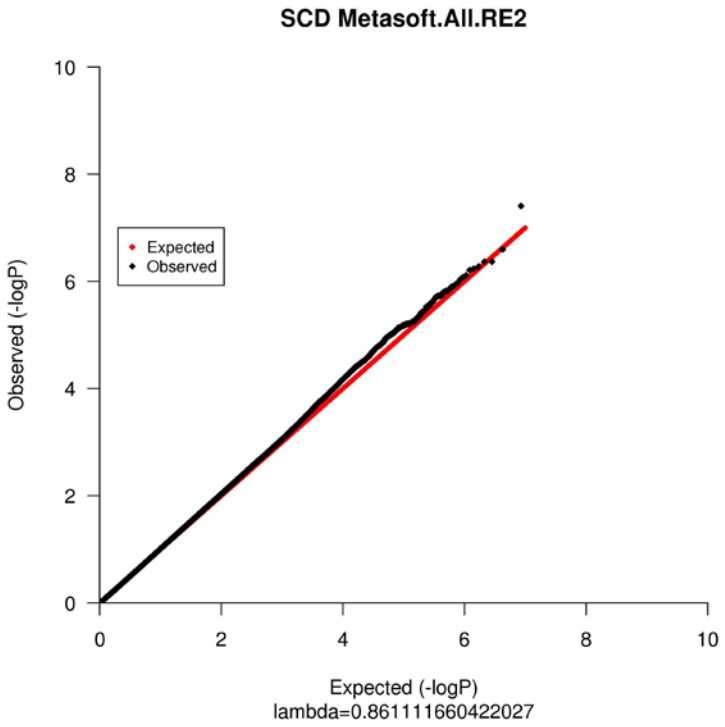
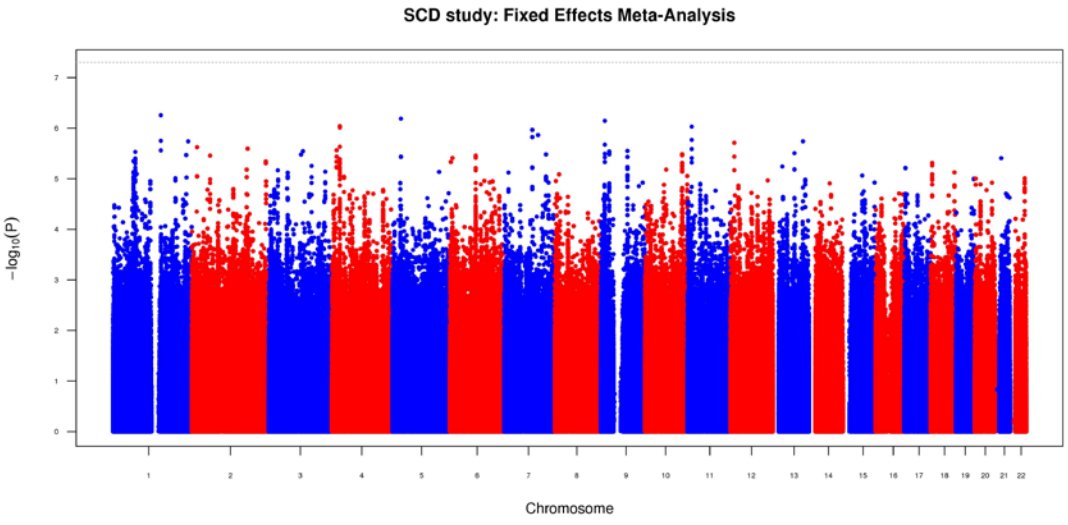
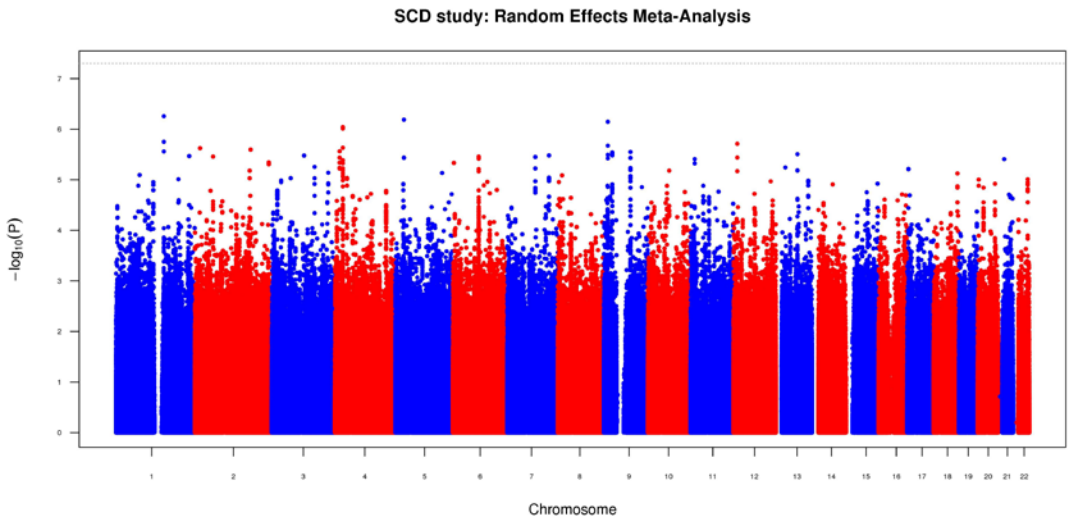


Figure 2.4: Manhattan plots showing results for trans-ethnic GWAS for SCA

A. Fixed effects model



B. Random effects model



C. Han and Eskin's random effects model

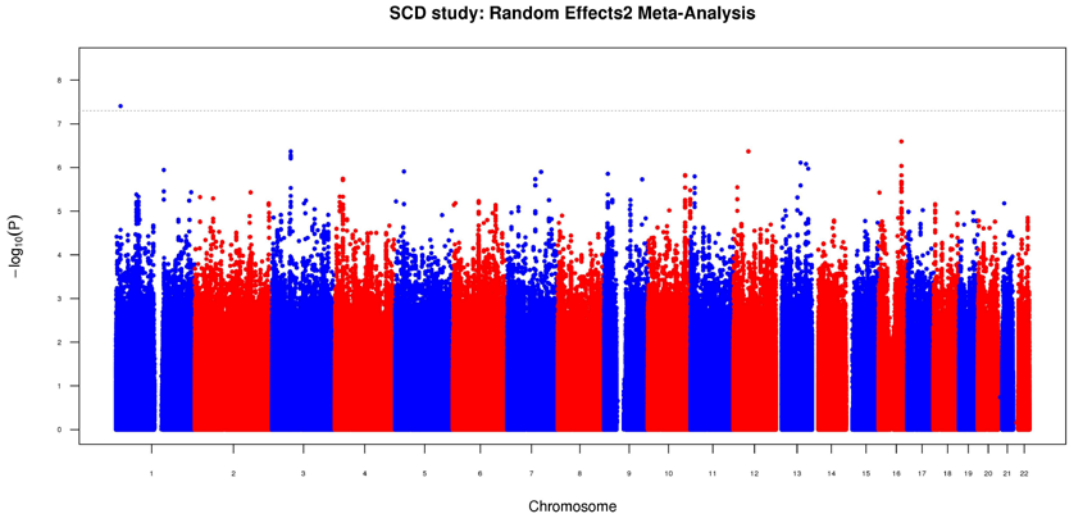
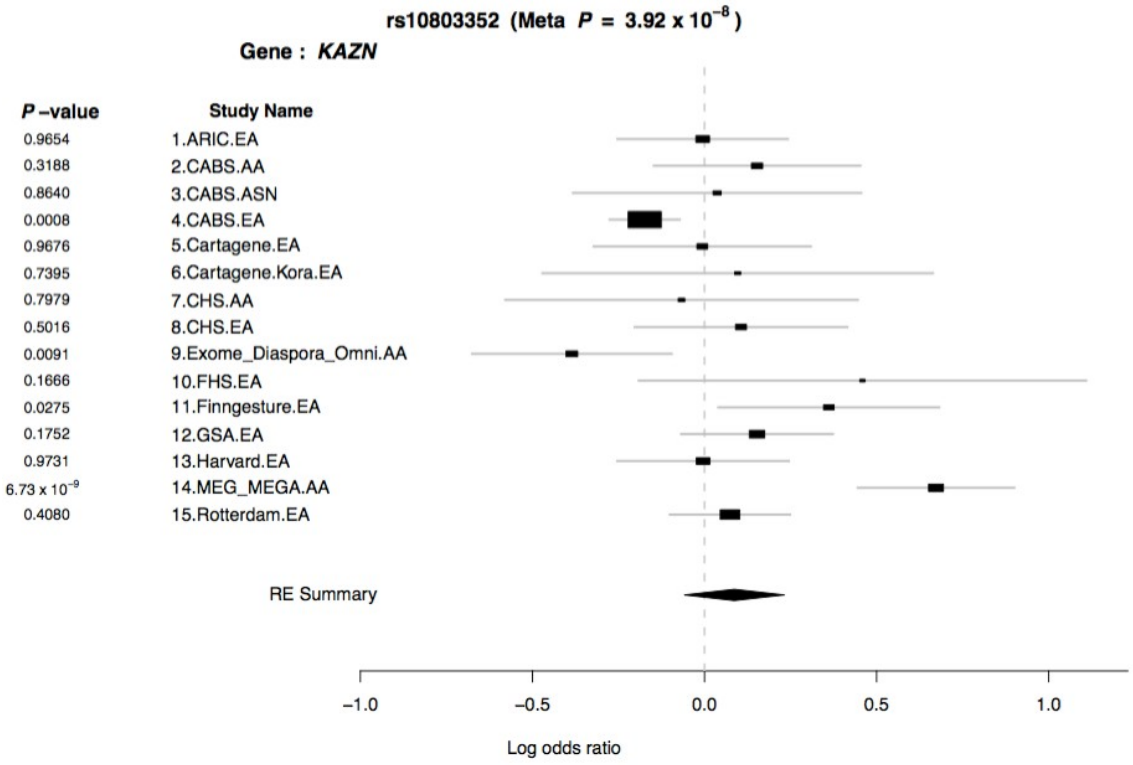


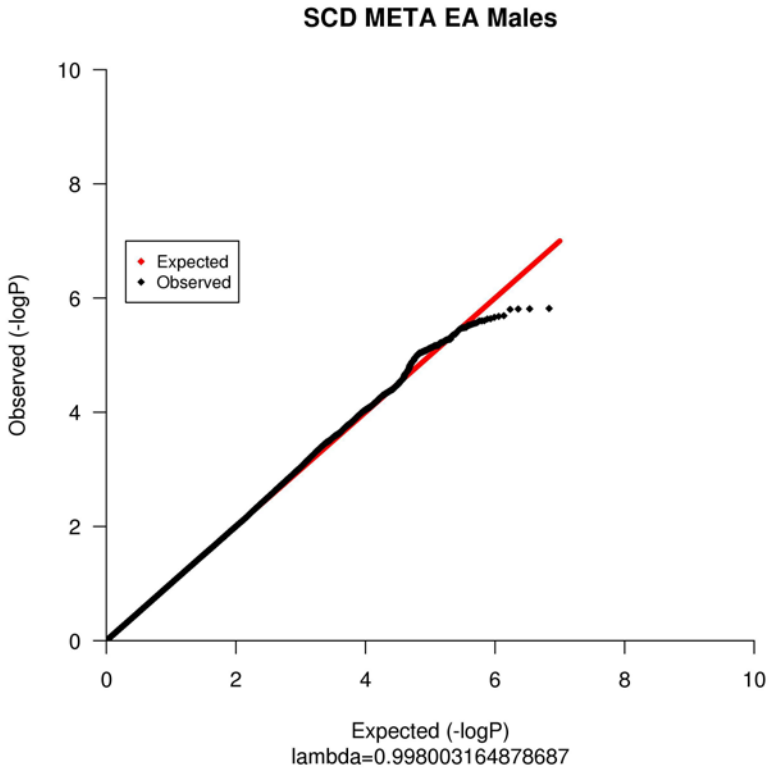
Figure 2.5: Forest plot for rs10803352



The forest plot for the single SNP (rs10803352) that reached the genome-wide significance threshold using the Han and Eskin's random effects model shows the significance of the SNP is largely driven by a single cohort (MEG_MEGA.AA). Given the effect is not replicated in the other cohorts, this SNP is likely a false positive finding.

Figure 2.6: QQ plots for sex-specific GWASs of European individuals

A. Males



B. Females

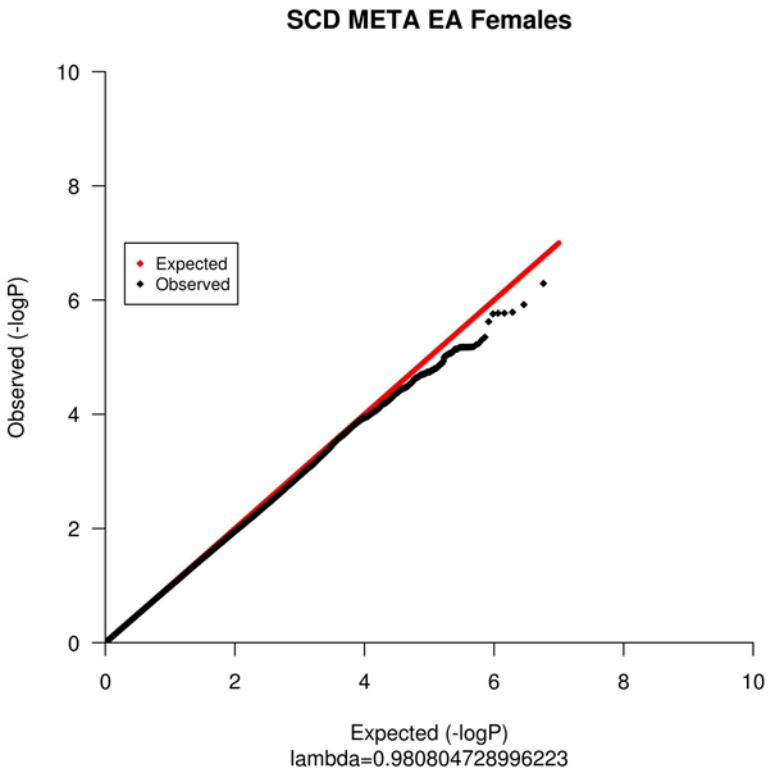
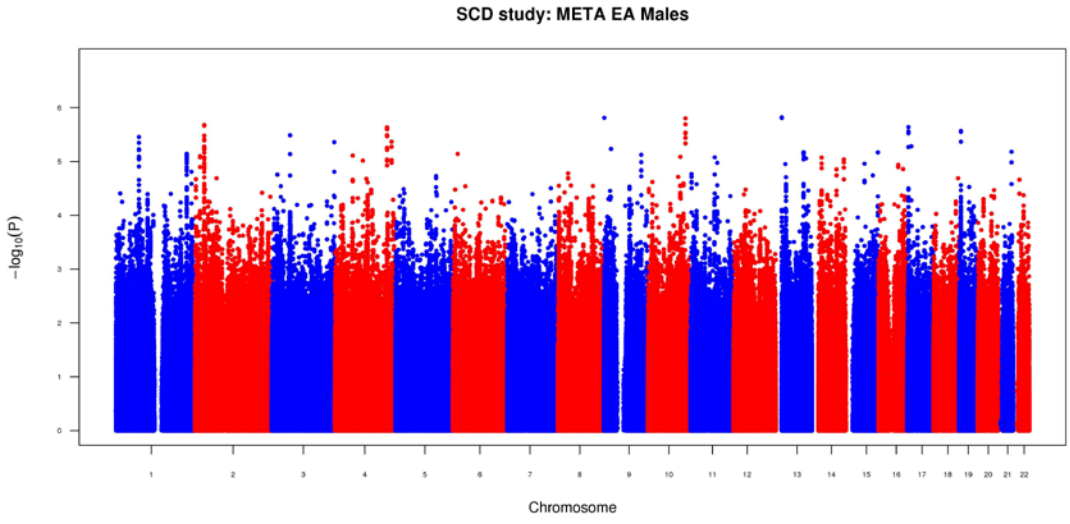


Figure 2.7: Manhattan plots for sex-specific GWASs of European individuals

A. Males



B. Females

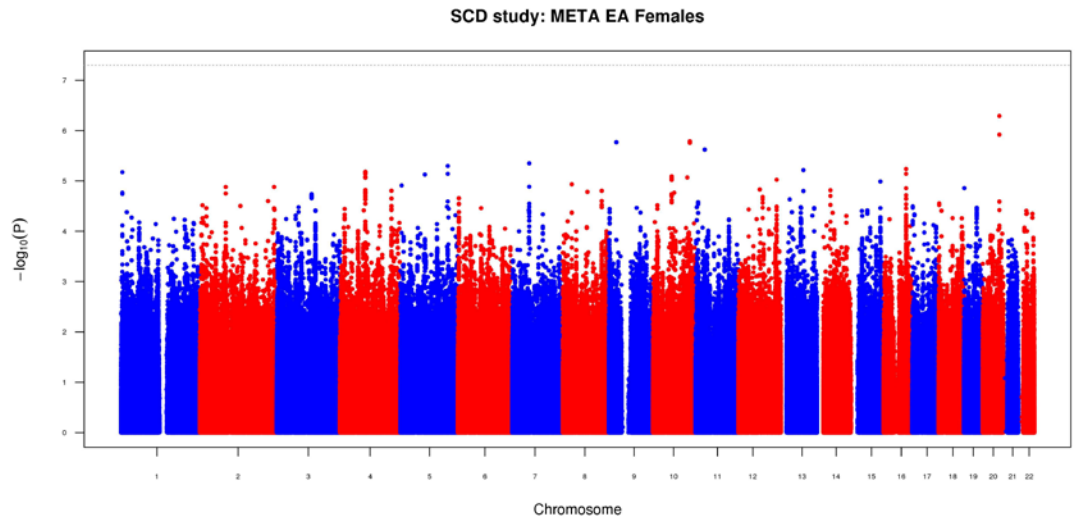
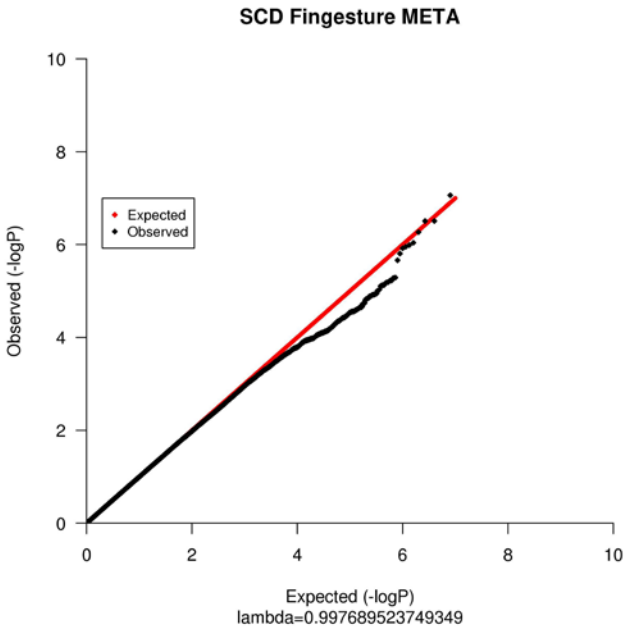
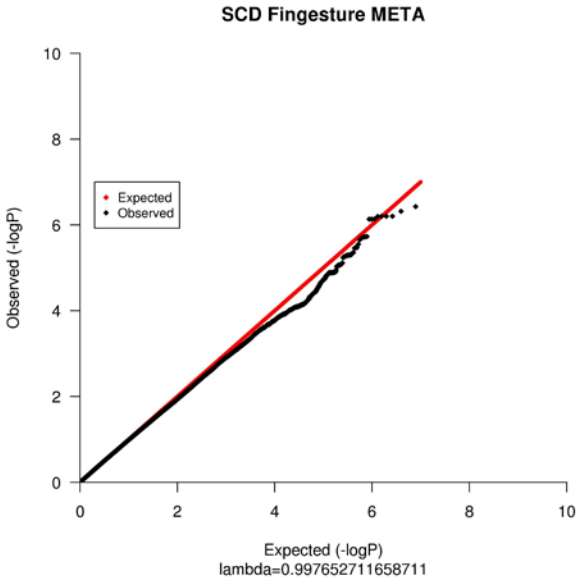


Figure 2.8: QQ plots for the disease-stratified GWASs

A. Ischemic cases/Population controls



B. Non-ischemic cases/Population controls



C. Ischemic cases/Ischemic controls

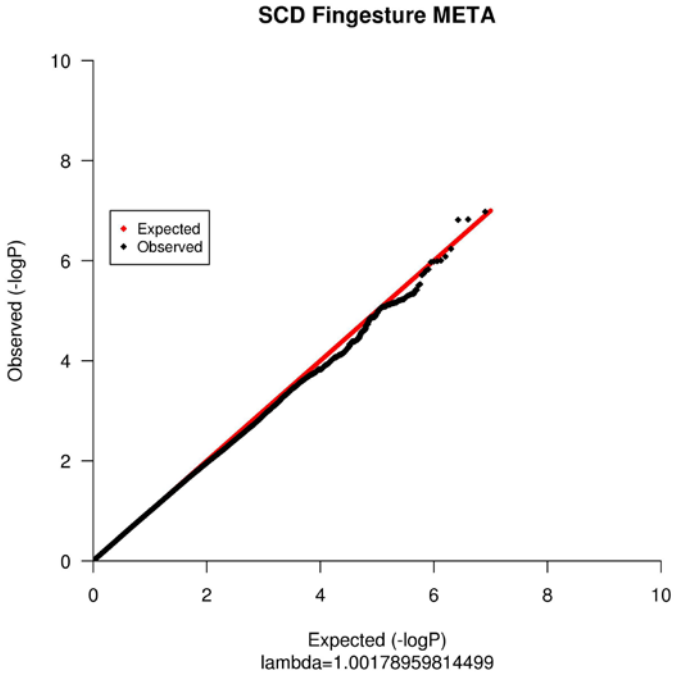
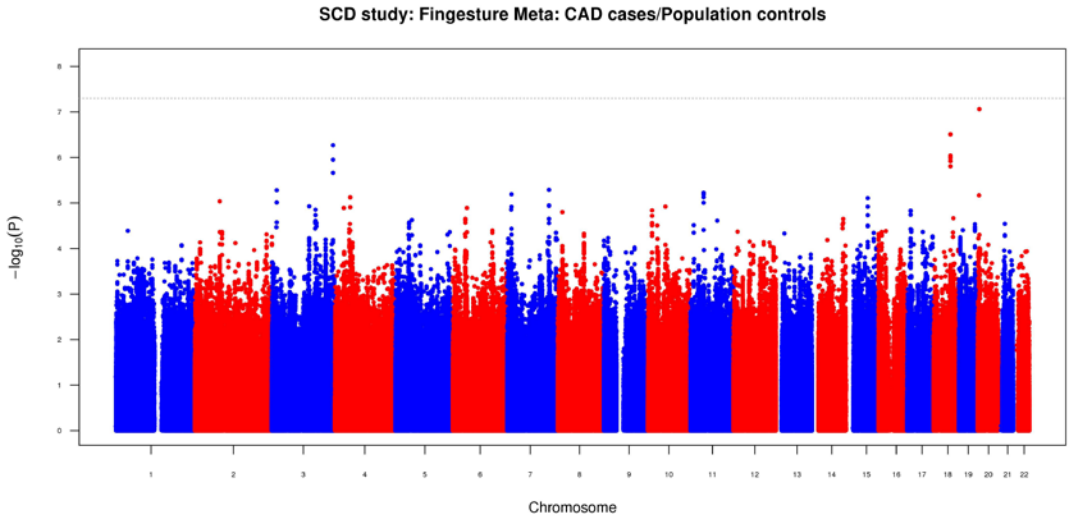
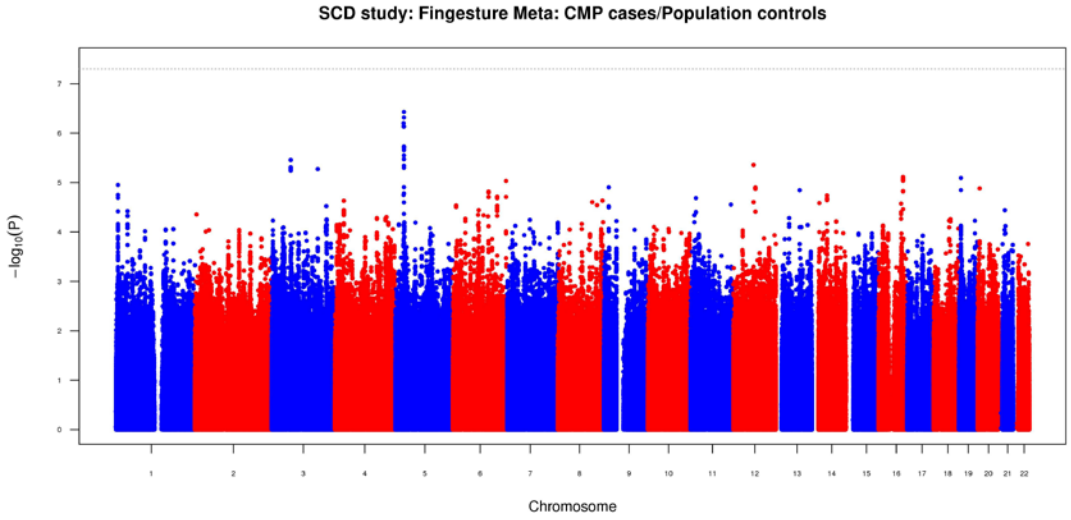


Figure 2.9: Manhattan plots for the disease-specific GWASs

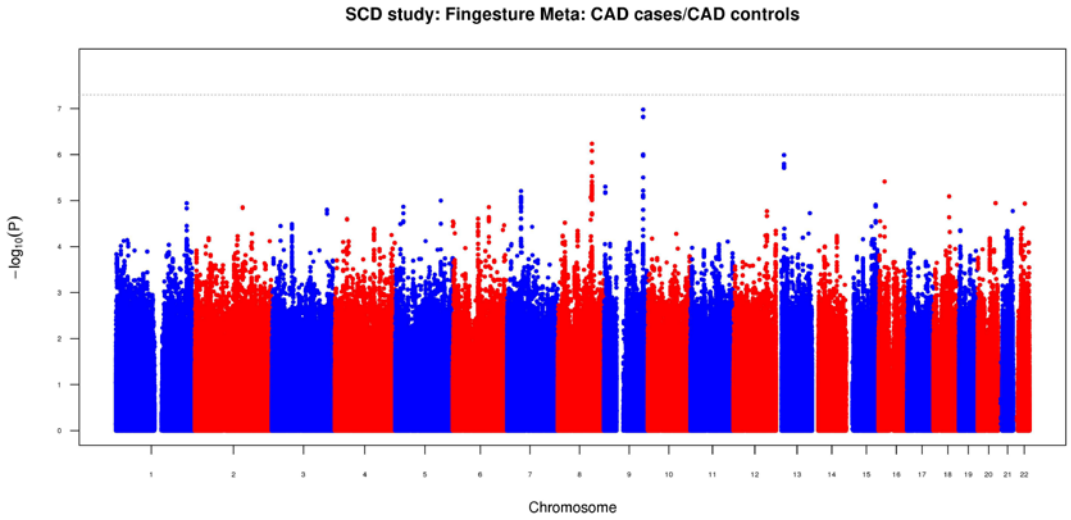
A. Ischemic cases/Population controls



B. Non-ischemic cases/Population controls



C. Ischemic cases/Ischemic controls



2.6 Tables

Table2.1 Sample Characteristics of European Cohorts

Cohort	ARIC	CABS	CARTAGENE
N, number of cases with genotype data	204	2165	166
N, number of controls with genotype data	8,682	2430	241
QC criteria, per sample	Sex-check, Removed duplicates, checks for cryptic relatedness and genetic outliers from PCA	Sex-check, Removed duplicates, checks for cryptic relatedness and genetic outliers from PCA	Sex-check, Removed duplicates, checks for genetic outliers from PCA
Genotyping platform	Affymetrix 6.0	Affymetrix Axiom	Illumina Human660K
Genotype calling algorithm	Birdseed	apt-probeset-genotype	Illumina beadstudio
Inclusion criteria--MAF	>1%	>1%	>=1%
Inclusion criteria--Call Rate per SNP	>95%	>95%	>=95%
Inclusion criteria--pvalue HWE	>1E-4	>1E-5	>=5E-6
Autosomal SNPs after QC	7,300,831	2,543,888	2,599,339
Imputation Reference Panel	TOPMed Freeze5	Hapmap.v2	Hapmap.v2
Imputation Software	Minimac3	Beagle	Impute v1.0.0
Meta analysis Imputation quality filter	>0.75	<1.1	<1.1
Meta analysis minor allele count (MAC) in cases filter	>5	no filter	no filter
Genomic inflation factor for meta-analysis	No adjustment	1.049	1.012
Sex, number of women among cases	53	496	30
Sex, number of women among controls	4655	537	50
Age, mean age at baseline among cases	55.78	67.55	56.2
Age, age-range at baseline among cases	45-65	20-101	22-77
Age, mean age at time of SCD among cases	68.29	66.65	56.6
Age, age-range at time of SCD among cases	51.29-79.95	23-96	28-86
Average time to SCD, yrs (for prospective studies)	12.5		
Mean followup time, yrs (for prospective studies)	20.3		
Study design	Prospective	Case-control	Case-control

Cohort	CHS	FHS	Fingesture	CARTAGENE/ KORA F3
N, number of cases with genotype data	138	32	340	169
N, number of controls with genotype data	3157	4358	570	338
QC criteria, per sample	Array call rate <95%, sex check		Sex-check, Removed duplicates, checks for cryptic relatedness and genetic outliers from PCA, genotyping call rate > 90%	Sex-check, Removed duplicates, checks for genetic outliers from PCA
Genotyping platform	Illumina CNV370	Affymetrix 500K+ 50K Human Gene Focused Panel	Affymetrix 6.0	Illumina HumanOmniExpress+HumanOmni25
Genotype calling algorithm	Illumina beadstudio	BRLMM	Birdseed	Illumina Genomestudio
Inclusion criteria--MAF	>=0%	>1%	>1%	>=0.1%
Inclusion criteria--Call Rate per SNP	>=97%	>95%	>95%	>=98%
Inclusion criteria--pvalue HWE	>1E-5	>1E-6	>1E-6	NA
Autosomal SNPs after QC	2,613,506	2,010,513	7,088,524	560,568
Imputation Reference Panel	Hapmap.v2	Hapmap.v2	TOPMed Freeze5	not imputed
Imputation Software	BimBam	Mach	Minimac3	not imputed
Meta analysis Imputation quality filter	<1.1	<1.1 & >=0.4	<1.1	not imputed
Meta analysis minor allele count (MAC) in cases filter	>5	>5	>10	>5
Genomic inflation factor for meta-analysis	1	1.024	1.029	No adjustment
Sex, number of women among cases	70		52	29
Sex, number of women among controls	1935		135	58
Age, mean age at baseline among cases	72.34		63.85	58
Age, age-range at baseline among cases	64-98		35-92	19-76
Age, mean age at time of SCD among cases	74.09		61.22	59.5
Age, age-range at time of SCD among cases	65-94		28-83	35-84
Average time to SCD, yrs (for prospective studies)	9.17	3.78		
Mean followup time, yrs (for prospective studies)	12.9	5.56		
Study design	Prospective	Prospective	Case-control	Case-control

Cohort	Harvard	Rotterdam study	Fingesture
N, number of cases with genotype data	420	385	1,171
N, number of controls with genotype data	424	5589	1,208
QC criteria, per sample	Checks for cryptic relatedness and genetic outliers from PCA, genotyping call rate > 95%	Sex-check, Removed duplicates, checks for cryptic relatedness and genetic outliers from PCA	Sex-check, Removed duplicates, checks for cryptic relatedness and genetic outliers from PCA, genotyping call rate > 95%
Genotyping platform	Affymetrix 6.0	Illumina HumanHap610	Illumina Global Screening Array
Genotype calling algorithm	Birdseed	Beadstudio Genecall	Illumina beadstudio
Inclusion criteria--MAF	>1%	>1%	>1%
Inclusion criteria--Call Rate per SNP	>95%	>95%	>95%
Inclusion criteria--pvalue HWE	>1E-6	>1E-6	>1E-4
Autosomal SNPs after QC	2,402,071	2,494,839	9,888,815
Imputation Reference Panel	Hapmap.v2	Hapmap.v2	TOPMed Freeze5
Imputation Software	Mach	Mach	Minimac3
Meta analysis Imputation quality filter	<1.1	<1.1 & >0.5	NA
Meta analysis minor allele count (MAC) in cases filter	no filter	>5	>10
Genomic inflation factor for meta-analysis	1.005	1.029	no adjustment
Sex, number of women among cases	127	203	420
Sex, number of women among controls	134	3344	1920
Age, mean age at baseline among cases	64.3	71.7	
Age, age-range at baseline among cases	40.3-91.9	55.2-95.5	
Age, mean age at time of SCD among cases	64.2	69.3	61.23
Age, age-range at time of SCD among cases	48.1-96.6	54.5-99.5	35-80
Average time to SCD, yrs (for prospective studies)	7.49	9.2	
Mean followup time, yrs (for prospective studies)	11.02	13.2	
Study design	Case-control from prospective studies and clinical trials	Prospective	Case-control

Cohort	ARIC	CABS	CARTAGENE	CARTAGENE/ KORA F3
SCD definition/Ascertainment	Sudden, pulseless condition from a cardiac origin in a previously stable individual, review of death and medical records	Sudden, pulseless condition from a cardiac origin in a previously stable individual, review of death and medical records. Pt in VF or asystole (NO PEA)	Sudden, pulseless condition from a cardiac origin in a previously stable individual, review of death and medical records.	Sudden, pulseless condition from a cardiac origin in a previously stable individual, review of death and medical records.
Control definition	Population based	Population based	French registry of Acute ST elevation or non-ST-elevation Myocardial Infarction (FastMI)	Population based
Software used for GWAS statistical analysis	FASTv2.4	R	snptest v2.1.1	PLINK v1.07
Model with covariates	Cox proportional hazards, with age, sex, and PCs as covariates	age, sex	age, sex	age, sex, PCs
Included in sex-stratified meta-analysis	Yes	Yes	No	No

Cohort	CHS	FHS	Fingesture
SCD definition/Ascertainment	Sudden pulseless condition presumed due to a cardiac arrhythmia, without evidence for a non-cardiac condition as a cause of the arrest, in an otherwise stable patient, after review of events surrounding arrest / death and medical records.	Coronary heart disease death within one hour of onset of symptoms adjudiated by panel of physicians.	Sudden, pulseless condition from a cardiac origin in a previously stable individual, review of death and medical records.
Control definition	Population based		MI survivors
Software used for GWAS statistical analysis	R		FASTv2.4
Model with covariates	age, sex, clinic	age,sex	sex, 10 PC
Included in sex-stratified meta-analysis	Yes	No	Yes

Cohort	Harvard	Rotterdam study	Fingesture
SCD definition/Ascertainment	a cardiac death is considered a definite SCD if the death or cardiac arrest that precipitated death occurred within one hour of symptom onset as documented by medical records or next of-kin reports or had an autopsy consistent with SCD (i.e. acute coronary thrombosis or severe coronary artery disease without myocardial necrosis or other pathologic findings to explain death)	Death <1 hour of cardiovascular symptoms or found dead and seen <24 hours earliers in stable medical condition. Based on review of medical records.	Sudden, pulseless condition from a cardiac origin in a previously stable individual, review of death and medical records.
Control definition	Controls from population studies and clinical trials matched on on study cohort, sex, age (+/−1 year), ethnicity, smoking status (current, never, past), time and date of blood sampling, fasting status, and presence or absence of cardiovascular disease (MI, angina, CABG, or stroke) prior to death.	Population based	MI survivors and population based
Software used for GWAS statistical analysis	Plink/Eigenstrat	ProbABEL	FASTv2.4
Model with covariates	20 PCs, cohort	age, sex, PCs	sex, 10 PC
Included in sex-stratified meta-analysis	Yes	Yes	Yes

Table 2.2 Sample Characteristics of African American & Asian

Cohort	ARIC	CABS	CHS
N, number of cases with genotype data	127	196	29
N, number of controls with genotype data	1,685	194	686
QC criteria, per sample	Sex-check, Removed duplicates, checks for cryptic relatedness and genetic outliers from PCA	Sex-check, Removed duplicates, checks for cryptic relatedness and genetic outliers from PCA	Sex-check, Removed duplicates, checks for cryptic relatedness and genetic outliers from PCA
Genotyping platform	Affymetrix 6.0		
Genotype calling algorithm	Birdseed		
Inclusion criteria--MAF	>1%	>1%	>1%
Inclusion criteria--Call Rate per SNP	>95%	>95%	>95%
Inclusion criteria--pvalue HWE	>1E-4		
Autosomal SNPs after QC	6,951,301	8,676,110	7,693,386
Imputation Reference Panel	TOPMed Freeze5	TOPMed Freeze5	TOPMed Freeze5
Imputation Software	Minimac3	Minimac3	Minimac3
Meta analysis Imputation quality filter	>0.75		
Meta analysis minor allele count (MAC) in cases filter	>5	>15	>5
Genomic inflation factor for meta-analysis	No adjustment	1.045	0.894
Sex, number of women among cases	63		
Sex, number of women among controls	1076		
Age, mean age at baseline among cases	55.83		
Age, age-range at baseline among cases	46-64		
Age, mean age at time of SCD among cases	67.6		
Age, age-range at time of SCD among cases	48.98-79.84		
Average time to SCD, yrs (for prospective studies)	11.76		
Mean followup time, yrs (for prospective studies)	19.78		
Study design	Prospective	Case-Control	Prospective

Cohort	CVPath/ARIC	CABS/CVPath/ARIC	CABS-Asian Americans
N, number of cases with genotype data	187	435	152
N, number of controls with genotype data	367	593	176
QC criteria, per sample	Sex-check, Removed duplicates, checks for cryptic relatedness and genetic outliers from PCA	Sex-check, Removed duplicates, checks for cryptic relatedness and genetic outliers from PCA	sex check, removed duplicates, checks for cryptic relatedness and genetic outliers from PCA
Genotyping platform	Omni/Exome chip/diaspora	Illumina Multi-Ethnic Global array	Affymetrix Axiom
Genotype calling algorithm		Illumina Genomestudio	apt-probeset-genotype
Inclusion criteria--MAF	>1%	>1%	>1%
Inclusion criteria--Call Rate per SNP	>95%	>95%	>95%
Inclusion criteria--pvalue HWE	>1E-4	>1E-4	>1E-5
Autosomal SNPs after QC	13,651,803	12,828,867	16820556
Imputation Reference Panel	TOPMed Freeze 5	TOPMed Freeze 5	1000G PhaseIv3
Imputation Software	Minimac3	Minimac3	minimac
Meta analysis Imputation quality filter	>0.75	>0.75	no filter
Meta analysis minor allele count (MAC) in cases filter	>5	>10	>10
Genomic inflation factor for meta-analysis			
Sex, number of women among cases	54	129	70
Sex, number of women among controls	133	306	91
Age, mean age at baseline among cases			63.44
Age, age-range at baseline among cases			20-102
Age, mean age at time of SCD among cases			61.59
Age, age-range at time of SCD among cases			32-89
Average time to SCD, yrs (for prospective studies)			
Mean followup time, yrs (for prospective studies)			
Study design	Case-Control	Case-Control	Case-control

Cohort	ARIC	CABS	CHS
SCD definition/Ascertainment	Sudden, pulseless condition from a cardiac origin in a previously stable individual, review of death and medical records	Sudden, pulseless condition from a cardiac origin in a previously stable individual, review of death and medical records	Sudden, pulseless condition from a cardiac origin in a previously stable individual, review of death and medical records
Control definition	Population based	Population based	Population based
Software used for GWAS statistical analysis	FASTv2.4		
Model with covariates	Cox proportional hazards, with age, sex, and PCs as covariates	Logistic score test with age as covariate	Cox proportional hazards, with age, sex, clinic, and PCs as covariates

Cohort	CVPath/ARIC	CABS/CVPath/ARIC	CABS-Asian Americans
SCD definition/Ascertainment			Sudden, pulseless condition from a cardiac origin in a previously stable individual, review of death and medical records. Pt in VF or asystole (NO PEA)
Control definition	Population based	Population based	Population based
Software used for GWAS statistical analysis	FASTv2.4	FASTv2.4	R
Model with covariates	Logistic regression with sex, PCs1-10 as covariates	Logistic regression with sex, PCs1-10 as covariates	age, sex, 2 PCs

Table 2.3 Top SNP in ischemic cases/population controls: rs7269951 (C/G)

	All		
Dataset	**cases/controls**	**Beta (SE)**	**P-value**
All cases/all controls	1481/1692	-0.301 (0.110)	0.006
non-ischemic cases/population controls	557/761	-0.508 (0.187)	0.006
ischemic cases/ischemic controls	1011/952	-0.138 (0.139)	0.32
ischemic cases/population controls	701/1169	-1.01 (0.187)	8.60E-08

	Males		
Dataset	**cases/controls**	**Beta (SE)**	**P-value**
All cases/all controls	1368/1250	-0.201 (0.128)	0.10
non-ischemic cases/population controls	422/354	-0.557 (0.236)	0.02
ischemic cases/ischemic controls	791/699	-0.063 (0.157)	0.69
ischemic cases/population controls	531/354	1.03 (0.228)	5.86E-06

	Females		
Dataset	**cases/controls**	**Beta (SE)**	**P-value**
All cases/all controls	299/871	-0.504 (0.219)	0.02
non-ischemic cases/population controls	135/407	0.460 (0.302)	0.13
ischemic cases/ischemic controls	133/232	-0.402 (0.354)	0.26
ischemic cases/population controls	83/407	-0.706 (0.348)	0.04

Table 2.4 Top SNP in non-ischemic cases/population controls: rs145160360 (C/T)

	All		
Dataset	**cases/controls**	**Beta (SE)**	**P-value**
All cases/all controls	1481/1692	0.547 (0.121)	5.77E-06
non-ischemic cases/population controls	557/761	0.927 (0.182)	3.74E-07
ischemic cases/ischemic controls	1011/952	0.300 (0.165)	6.90E-02
ischemic cases/population controls	701/1169	0.425 (0.202)	0.035

	Males		
Dataset	**cases/controls**	**Beta (SE)**	**P-value**
All cases/all controls	1218/1053	0.521 (0.150)	5.10E-04
non-ischemic cases/population controls	422/354	1.08 (0.267)	4.90E-05
ischemic cases/ischemic controls	791/699	0.220 (0.186)	0.88
ischemic cases/population controls	531/354	0.768 (0.261)	0.003

	Females		
Dataset	**cases/controls**	**Beta (SE)**	**P-value**
All cases/all controls	268/639	0.601 (0.211)	0.004
non-ischemic cases/population controls	135/407	0.779 (0.252)	0.002
ischemic cases/ischemic controls	133/232	0.532 (0.417)	0.32
ischemic cases/population controls	83/407	-0.351 (0.438)	0.95

Table 2.5 Top SNP in ischemic cases/ischemic controls: rs58823851 (A/G)

	All		
Dataset	**cases/controls**	**Beta (SE)**	**P-value**
All cases/all controls	1481/1692	0.224 (0.083)	0.007
non-ischemic cases/population controls	557/761	-0.252 (0.131)	0.055
ischemic cases/ischemic controls	1011/952	0.614 (0.116)	1.05E-07
ischemic cases/population controls	701/1169	0.095 (0.125)	0.45

	Males		
Dataset	**cases/controls**	**Beta (SE)**	**P-value**
All cases/all controls	1368/1250	0.241 (0.098)	0.013
non-ischemic cases/population controls	422/354	-0.248 (0.157)	0.12
ischemic cases/ischemic controls	791/699	0.565 (0.131)	1.66E-05
ischemic cases/population controls	531/354	0.158 (0.140)	0.26

	Females		
Dataset	**cases/controls**	**Beta (SE)**	**P-value**
All cases/all controls	299/871	0.133 (0.172)	0.44
non-ischemic cases/population controls	135/407	-0.251 (0.228)	0.27
ischemic cases/ischemic controls	133/232	0.800 (0.287)	0.005
ischemic cases/population controls	83/407	-0.166 (0.290)	0.57

Chapter 3: Identifying causal risk factors for SCA using Mendelian Randomization in European Individuals

3.1 Introduction

Sudden cardiac arrest (SCA) is a major cause of cardiac mortality, affecting over 300,000 people in the US every year.[39] Despite recent increases in SCA survival rates[40], survival remains low. Exploring the genomic architecture of SCA allows us to assess causal relationships of clinical and subclinical risk factors with SCA. Mendelian randomization methods exploit the fact that genetic variants are determined at conception and randomly distributed in populations, to determine whether an exposure may be causally associated with the outcome, and to estimate the effect size of that causal association[41–43].

Here, we use a multi-SNP genetic risk score association (GRSA) model to compare genetic associations of 18 known SCA risk factors to genetic associations with SCA as an effective way to understand the potential underlying causal pathways and processes that modulate SCA risk. While the majority of the results reported in this chapter are using the most recently published EA GWAS of 3,939 SCA cases and 25,989 controls, we also report the Mendelian randomization results for the largest EA GWAS for SCA of 5,290 SCA cases and 26,997 controls described in the previous chapter, and discuss the discrepancies between the two analyses.

3.2 Methods

3.2.1 Mendelian Randomization Instrument

Observational studies examine association of an exposure (e.g., body mass index, or BMI) with an outcome (e.g., SCA) but cannot assess causality. Unobserved variables affecting both exposure and outcome may confound these associations and lead to biased estimates of association. Mendelian randomization is based on the assumption that because genetic variants are determined at conception and are randomly distributed in large populations, they are unassociated with potential confounders. Therefore, under certain assumptions such as the absence of genetic pleiotropy, genetic variants used as instrumental variables can determine whether an exposure is potentially causally associated with the outcome, and estimate the size of that association. Here we use a multi-SNP genetic risk score association (GRSA) model to compare genetic associations with SCA with those of known SCA risk factors as an effective way to understand the underlying causal pathways and processes that influence SCA risk.

3.2.2 Genetic Risk Score Association (GRSA)

We estimated a separate GRSA utilizing the recently published SCA GWAS of 3,939 cases and 25,989 controls[12] for each of the following: (1) CAD and traditional CAD risk factors, including type 2 diabetes (T2D), fasting glucose adjusted for BMI (FGadjBMI), fasting insulin adjusted for BMI (FIadjBMI), diastolic blood pressure (DBP), systolic blood pressure (SBP), total cholesterol (TCH), and triglycerides (TG); (2) cardiac electrophysiologic factors, including atrial fibrillation (AF), heart rate (HR), QRS interval (QRS), and QT interval (QT); and (3) anthropometric traits, including BMI, waist circumference adjusted

for BMI (WCadjBMI), waist to hip ratio adjusted for BMI (WHRadBMI), and height. **Table 3.1** details the 18 traits, and the source published GWAS used to construct the GRSA models for these traits.

To estimate GRSAs for each putative SCA risk factor, we examined genome-wide SNPs associated with the risk trait following stringent LD-pruning. For LD pruning, the 'clump' function implemented in PLINK v1.9[44,45] was applied on summary statistics from GWAS for risk factors to identify independent common variants (minor allele frequency >1%) associated with the respective traits. Briefly, the 'clump' function utilizes a P-value aware linkage disequilibrium (LD)-pruning approach that was used to identify index SNPs at the two alpha cutoffs, and exclude variants with $R^2 \geq 0.01$ within 1Mb of the index variant. For LD pruning, we used imputed genotype data from 9,747 European individuals from the ARIC study. For index variants with *P*-value $<5x10^{-8}$, we used a more stringent approach of excluding all variants within a 1Mb radius to exclude any possibility of spurious association due to weak LD with the index SNP. A feature of this LD-pruning process is that by iteratively selecting the most significant SNP in a 2 MB window, followed by removing all SNPs with $R^2 \geq 0.01$, even using the custom *P*-value cut-off, different traits are unlikely to contain a significant number of overlapping SNPs in the absence of significant pleiotropy. SNPs were also removed if not present in the CABS study, which contained over half of all the SCA cases.

The associations of these SNPs with the risk factors and the SCA outcome are used to calculate an inverse-variance weighted multi-SNP GRSA as

implemented in the R-package 'MendelianRandomization'.[46] This GRSA can be interpreted as an inverse-variance weighted, meta-analyzed (over SNPs) estimate of the causal log odds ratio for SCA associated with a one SD higher value of the risk factor from a Mendelian randomization analysis.[47] It is computationally equivalent to the slope estimate from a zero-intercept linear regression with log odds ratio for the association of an additional variant allele in SNPs with SCA (β_{SCA}) as the dependent variable and the mean difference associated with one additional variant allele in SNPs on the risk factor trait (β_{trait}) as the independent variable, weighted by the standard error of the β_{SCA} squared (SE_{SCA}^{2}) (**Figure 3.1**). We evaluated the use of other MR methods, including MR-Egger, simple median, and median-weighted. However, we found while these produced similar GRSA estimates as the inverse-weighted (IVW) method, these other methods had lower power (**Table 3.2**). We therefore only report the results from the IVW method. We also used the intercept test from the MR-Egger method to evaluate the presence of pleiotropy in our analyses (**Table 3.2**).

The validity of this analysis requires that SNPs included can only affect the outcome through their effects on the risk factor (i.e. no horizontal pleiotropy). If there is no pleiotropy, the SNPs contributing the GRSA estimate should all estimate the same magnitude causal association between risk factor and SCA. We use the HEIDI-outlier method from the 'gsmr' R package to detect and remove potentially pleiotropic SNPs.[48] Note that we report GRSA estimates from analyses only including SNPs that meet a stringent genome-wide significant (GWS) *P*-value cut-off ($P<5x10^{-8}$), $GRSA_{GWS}$, as SNPs at this significance level likely are

true positives and reliable instruments. However, the power for Mendelian randomization is dependent on the variance explained by the SNPs included in the GRSA, and for complex traits, the majority of the true signals may lie in SNPs that do not meet genome-wide significance. Therefore, we identified a somewhat arbitrary *P*-value cut-off based on visual inspection of the variance explained plots that largely maximizes variance explained while minimizing the number of SNPs (**Figure 3.2**). We found that all the traits fell between 0.2-0.4 *P*-value cutoff, but the results within a trait were robust to cutoffs chosen between 0.2 and 0.4. We use a GRSA constructed with this custom *P*-value cut-off ($GRSA_{max}$) to assess only the significance of the GRSA (P_{max}), as this model has the greatest power to assess the significance of an association. P_{max} is determined by permutation due to inflated test statistics (**Figure 3.3).** At less stringent *P*-values, false-positive SNPs may be included resulting in a bias of the estimate toward the confounded association level. Therefore, we do not use the $GRSA_{max}$ to determine the magnitude of the GRSA association, only its direction and significance. We performed two analyses, one using $GRSA_{GWS}$ to evaluate significance and effect size, and secondarily using the $GRSA_{max}$ to evaluate potential associations and directions of effect at maximal power (P_{max}). We performed multiple-testing adjustment on all resulting *P*-values (P_{GWS} and P_{max}) using a false discovery rate (FDR) cutoff of FDR<0.05.

We similarly computed risk factor GRSAs on the outcome of CAD. We use a 1-degree of freedom Wald test to test for difference in $GRSA_{GWS}$ magnitudes between SCA and CAD. We also performed sex-specific SCA GWAS analyses to

construct trait GRSAs separately by sex. GRSAs were constructed from the same set of LD-pruned SNPs used for overall $GRSA_{GWS}$ analyses. *P*-values for difference in $GRSA_{GWS}$ between sexes were obtained from 1-degree of freedom Wald test.

3.3 Results

3.3.1 CAD and CAD risk factors

Prevalent CAD is an important SCA risk factor with ~80% of male SCA survivors having underlying CAD[49]. From $GRSA_{GWS}$ analysis we show that the difference in CAD status is causally associated with SCA (odds ratio in SCA risk per log odds difference in CAD, 1.36; 95% CI, 1.19-1.55; $P_{GW} = 9.29 \times 10^{-5}$) (**Figure 3.4** and **Table 3.3**). While traditional CAD risk factors (blood pressure, lipids and diabetes) were not significantly associated with SCA at the more restrictive $GRSA_{GWS}$ threshold, using $GRSA_{max}$ to maximize power, several additional associations were detected, including type 2 diabetes ($P_{max}<0.001$), LDL ($P_{max=}0.005$), total cholesterol ($P_{max}<0.001$), triglycerides ($P_{max}<0.001$), diastolic blood pressure ($P_{max}=0.0170$), and systolic blood pressure ($P_{max}=0.0230$) (**Table 3.4**). In the $GRSA_{max}$ analysis, variants associated with higher diabetes risk, higher cholesterol and triglyceride levels, and higher systolic and diastolic blood pressure were all associated with higher SCA risk.

3.3.2 Cardiac electrophysiologic factors

To explore the influence of cardiac electrophysiology on SCA, we examined genetics of electrophysiologic traits associated with SCA: (1) atrial fibrillation, (2) QT interval (ventricular repolarization), (3) QRS interval

(ventricular conduction), and (4) heart rate. In the $GRSA_{GWS}$ analysis, we show that longer QT interval, a risk factor for SCA in the general population, is significantly associated with SCA (odds ratio in SCA risk per SD increase in QT, 1.44; 95% CI, 1.13-1.83; P_{GWS}=0.018) (**Figure 3.4** and **Table 3.3**).[50] Using $GRSA_{max}$, in addition to QT, we also identified a significant association of AF with SCA (P_{max}<0.001 for both QT and AF) (**Table 3.4**). Variants associated with longer QT interval and higher AF risk were associated with higher SCA risk. By contrast, no significant association was seen with QRS or heart rate, even at the more permissive and statistically powerful $GRSA_{max.}$

3.3.3 Anthropometric Measures

The BMI $GRSA_{GWS}$ was significantly associated with SCA (odds ratio for SCA risk per SD higher BMI, 1.63; 95% CI, 1.23-2.15; P_{GWS}=0.005) (**Figure 3.4** and **Table 3.3**). Using $GRSA_{max}$, we found a significant negative association between height and SCA (P_{max}<0.001) (**Table 3.4**). Variants associated with greater height are associated with lower CAD risk[51], and we correspondingly observed a negative GRSA between SCA and height. No significant association was seen with GRSAs composed of variants associated with measures of central/abdominal adiposity, such as waist-to-hip ratio or waist circumference.

3.3.4 Contrasting SCA and CAD GRSAs

Given the strong association of CAD with SCA, we compared the magnitudes of risk factor $GRSA_{GWS}$ on the outcomes of SCA (**Figure 3.4**) and CAD (**Figure 3.5**) to identify traits where risk factors may be more strongly causally associated with SCA than CAD. While the $GRSA_{GWS}$ for traditional CAD

risk factors (blood pressure and lipid traits) are larger for CAD risk than SCA risk, we find that $GRSA_{GWS}$ for electrophysiologic traits of QT interval (0.34 for SCA vs. 0.096 for CAD, *P* for difference = 0.06) and AF (0.097 for SCA vs. -0.029 for CAD, *P* for difference=0.017), there was a suggestion of a larger association with SCA than CAD risk (**Figure 3.6** and **Table 3.3**).

3.3.5 Sex differences

Sex differences in SCA incidence, underlying SCA pathophysiology, and prevalence of certain risk factors have been well documented[52], yet little is known about whether the effect of risk factors on SCA differs by sex. Among $GRSAs_{GWS}$ where a main effect association was identified, we found a nominally significant difference in association between women and men for diabetes (0.240 for women vs. 0.021 for men, *P* for difference = 0.05) and HDL (-0.417 for women vs. 0.026 for men, *P* for difference = 0.04) (**Table 3.5**).

3.4 Discussion

This study demonstrates that while SCA is a complex disease with multiple risk factors, a comprehensive genetic approach can shed light on causal versus correlational associations. Using Mendelian randomization with the European SCA GWAS (3,939 cases and 25,989 controls), we establish that differences in CAD, BMI, and QT interval are causally associated with SCA. Secondary analyses further implicate type 2 diabetes, additional traditional CAD risk factors such as lipids and blood pressure, as well as height and atrial fibrillation. Of particular interest, the GRSA estimates of phenotypes associated with electrical instability (AF and QT) are causally associated with SCA risk,

more so than they are causally associated with CAD. This confirms our understanding of the pathophysiology of SCA—SCA is not simply fatal CAD, but rather, electrical instability also plays a prominent role in influencing SCA risk.

Intriguingly, not all electrophysiologic phenotypes observationally linked to SCA are causally associated with SCA in our analyses. QRS interval and heart rate, two traits observationally associated with SCA[53,54], failed to show significant evidence of a shared genetic basis with SCA. This lack of association may be due to inadequate power to identify more modest correlations. Alternatively, it may be that the associations from observational studies are confounded by other factors, and not causative (**Figure 3.1B-C**). For instance, underlying CAD can lead to both longer QRS interval and increased SCA risk; thus, while observational studies show an association between SCA and both traits (CAD and QRS interval), the association between SCA and QRS interval may not be causal. Similarly, the observational association of higher heart rate with SCA risk may be confounded by higher adrenergic state due to underlying heart disease and not itself be causal. Thus, the GRSA approach to examining observational risk factors assists in differentiating causative factors from confounded associations.

CAD is the most common underlying pathologic substrate for SCA. It is reassuring, therefore, that we find significant estimated causal associations with SCA risk using GRSA models constructed from CAD and traditional CAD risk factors, including blood pressure, diabetes and cholesterol traits.

Anthropometric measures appear to be causally associated with SCA. Shorter stature is associated with increased SCA risk in observational studies; our

findings support the conclusion that this observational association is causal. Observational data on BMI and SCA risk have been conflicting, perhaps due to confounding from smoking status and frailty. Previously[55], we have shown that increased BMI is associated with increased SCA risk in non-smokers, but not smokers. In this study, we find that differences in BMI, but not central/abdominal obesity, were causally associated with SCA risk. This finding is especially interesting in the context of recent data that imply different biological process underlying BMI and central obesity.[56,57]

Finally, of the traits associated with SCA, we found that GRSAs for diabetes and HDL were nominally significantly different between men and women. While diabetes is a SCA risk factor among both sexes, previous observational studies have consistently suggested a stronger, albeit not statistically different, association among women than men[58,59]. These findings may reflect different underlying SCA pathophysiology between men and women. While these differences may be due to chance as they do not remain significant after multiple test correction, it is also likely that our study is underpowered to detect these differences.

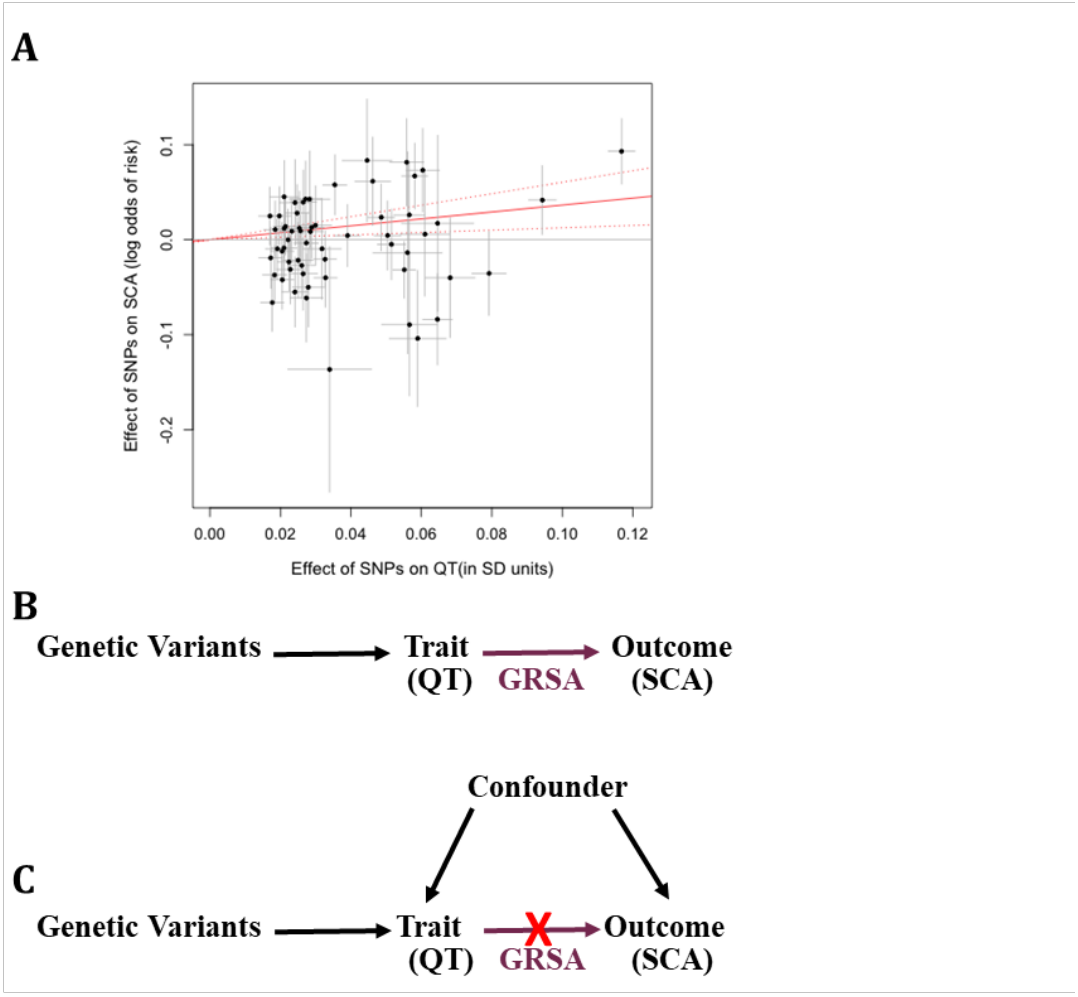
Several limitations deserve consideration. First, as somewhat addressed above, without detailed autopsy information, rhythm monitoring, and information on circumstances surrounding the cardiac arrest, the underlying etiology and mechanism of death may be heterogeneous and genetic associations are likely to be diluted. Nonetheless, clinical and autopsy studies have demonstrated a predominant, common pathophysiology of SCA in Western populations: VF in

the setting of CAD. Hence, it is reassuring that our genetic studies suggest an important role for both CAD and electrical instability in SCA. Second, the validity of the GRSA method as a Mendelian randomization instrument rests on the assumption that the variant causes differences in the outcome only by its effects on the risk factor of interest, and not directly or by influencing other risk factors. Although we did not explicitly exclude SNPs associated with multiple risk factors (genetic pleiotropy), we did utilize a goodness-of-fit approach to exclude putative "pleiotropic" effects from all GRSAs. Furthermore, we performed a sensitivity analysis using the MR-Egger method, which tests for the presence of pleiotropy. Only HDL was found to be significantly influenced by pleiotropy (P=0.02). Lastly, while genetic pleiotropy can bias our conclusions, important influence is less likely when using multiple SNPs aggregated in a genetic risk score.[60]

In conclusion, we have provided evidence for causal associations between some, but not all, observational risk factors for SCA. We show that differences in CAD status, BMI, and QT interval are causally associated with SCA risk. However, as evident in the second Mendelian randomization analysis, phenotypic heterogeneity is likely causing the lack of significantly causally associated risk factors. Further work into generating a more homogeneous phenotype is necessary to further explore causal SCA risk factors.

3.5 Figures

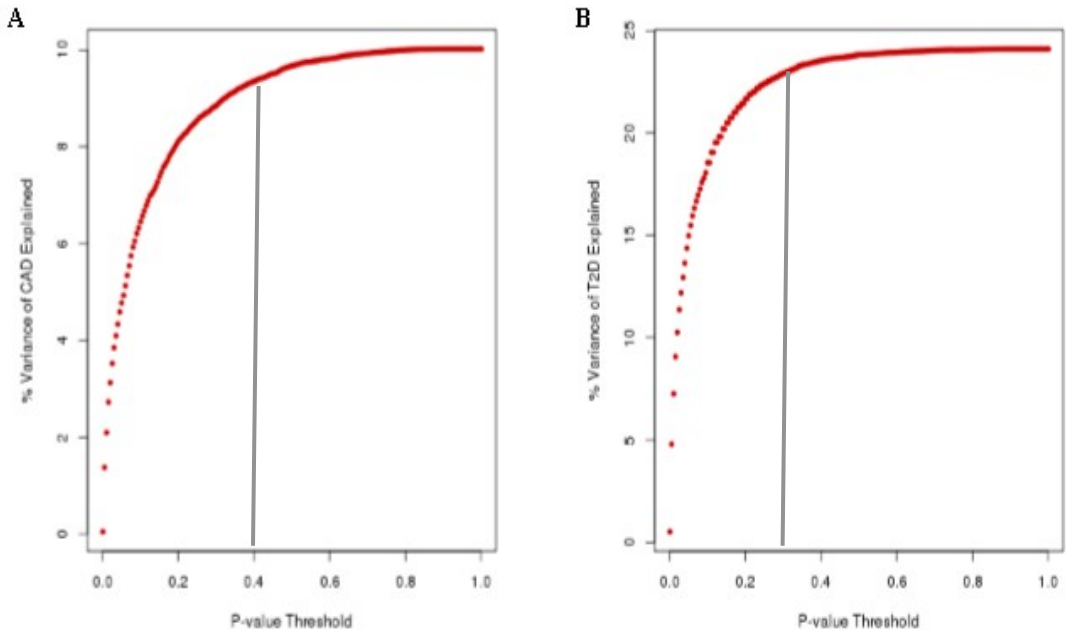
Figure 3.1: Genetic Risk Score Association (GRSA) Estimation



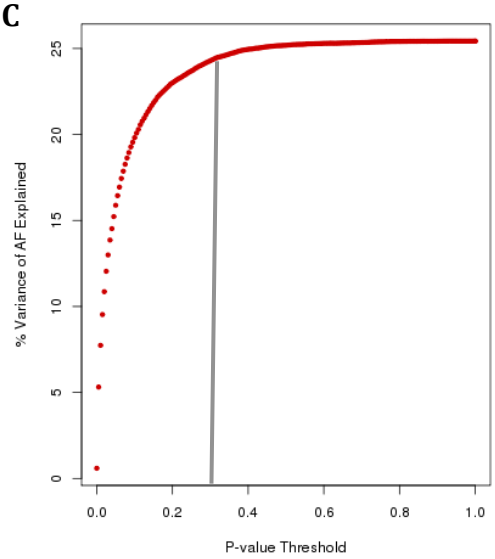
The plot (A) illustrates the process by which the QT-SCA GRSA is calculated using SNPs associated with QT at $P<5x10^{-8}$. The points represent the effect of each SNP on QT (in units of standard deviation of QT) on the x-axis, and the log odds effect on SCA risk (corresponding 95% confidence intervals in grey) on the y-axis. The estimate of the genetic risk score association is the slope of the zero-intercept weighted regression line (solid red line). For the GRSA used in our

analyses, the model contains a genome-wide LD-pruned SNP set (details in Methods). The top directed acyclic graph (B) represents a scenario in which the trait of interest has a causal effect on the outcome. If the GRSA, comprised of trait-associated variants (e.g., QT), has a significant effect on the outcome (e.g., SCA), it supports a causal role for the trait on the outcome. The bottom directed acyl graph (C) presents the case where an association is observed between the trait and outcome, but the GRSA comprised of trait-associated variants is not significantly associated with the outcome, suggesting that the observational association is likely being mediated by a confounding variable and the trait does not have a causal impact on the outcome.

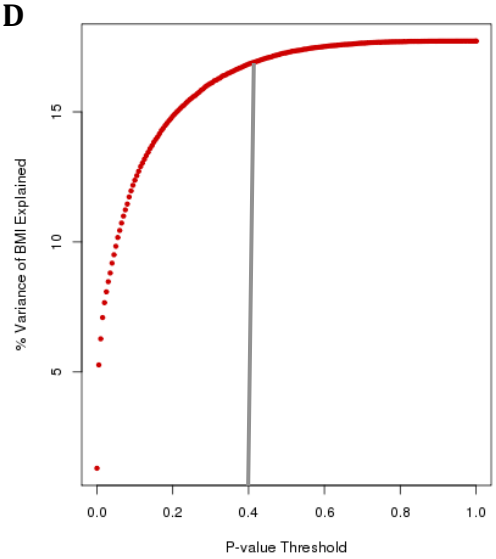
Figure 3.2: Variance of Trait Explained Plots



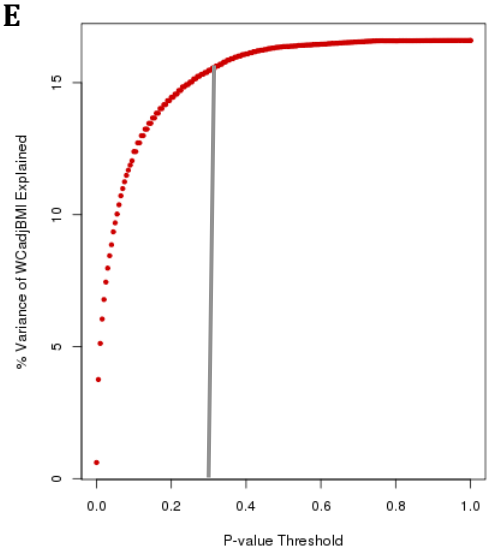
C


% Variance of AF Explained
0
5
10
15
20
25
0.0
0.2
0.4
0.6
0.8
1.0
P-value Threshold

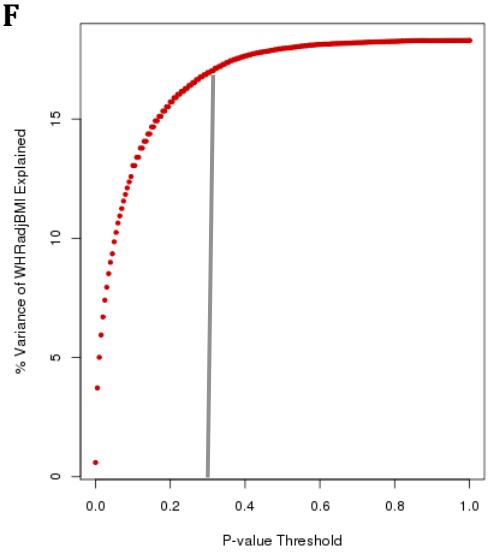
D


% Variance of BMI Explained
5
10
15
0.0
0.2
0.4
0.6
0.8
1.0
P-value Threshold

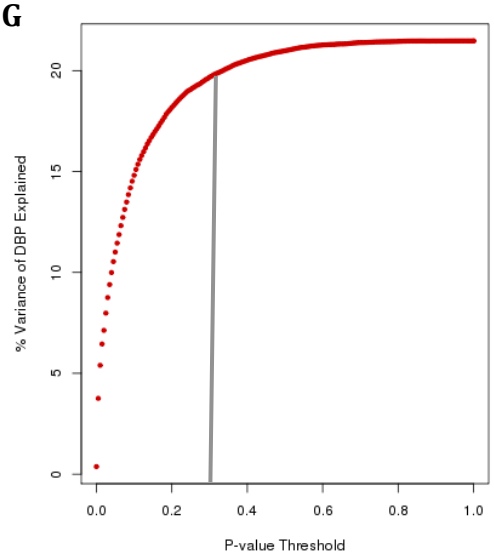
E


% Variance of WCadjBMI Explained
0
5
10
15
0.0
0.2
0.4
0.6
0.8
1.0
P-value Threshold

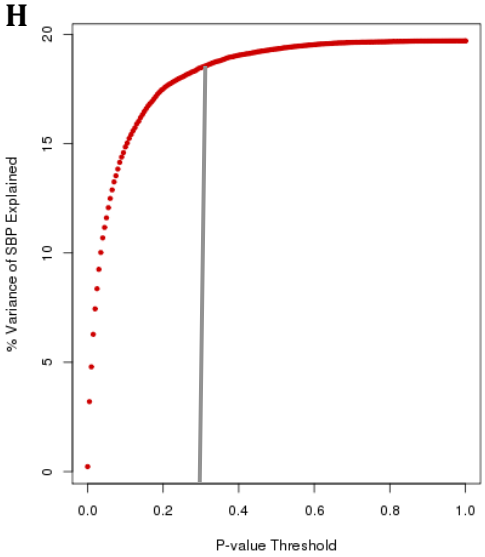
F


% Variance of WHRadjBMI Explained
0
5
10
15
0.0
0.2
0.4
0.6
0.8
1.0
P-value Threshold

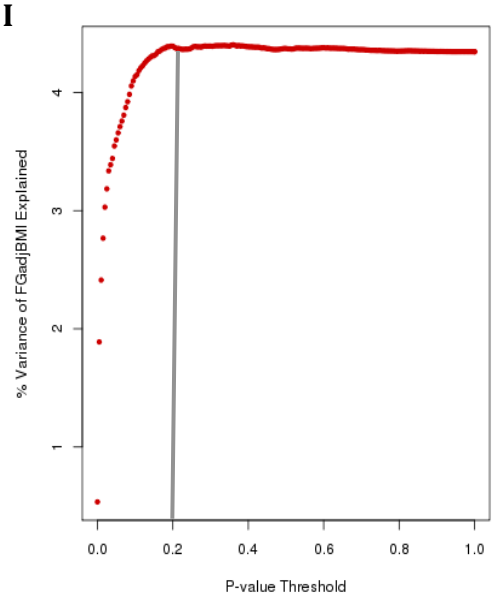
G


% Variance of DBP Explained
0
5
10
15
20
0.0
0.2
0.4
0.6
0.8
1.0
P-value Threshold

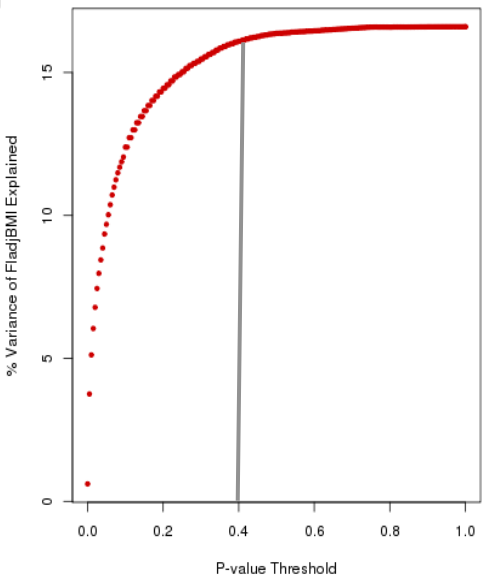
H


% Variance of SBP Explained
0
5
10
15
20
0.0
0.2
0.4
0.6
0.8
1.0
P-value Threshold

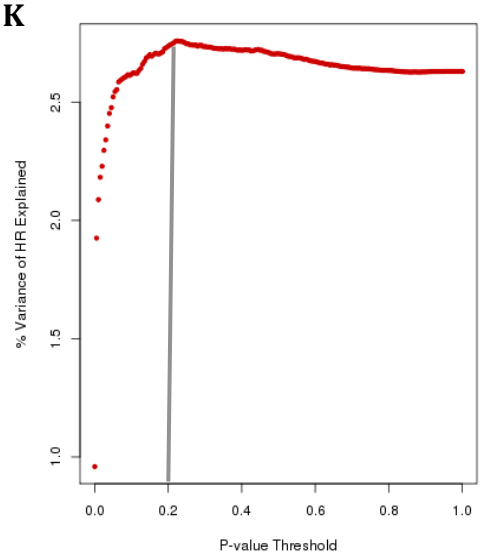
I


% Variance of FGadjBMI Explained
1
2
3
4
0.0
0.2
0.4
0.6
0.8
1.0
P-value Threshold

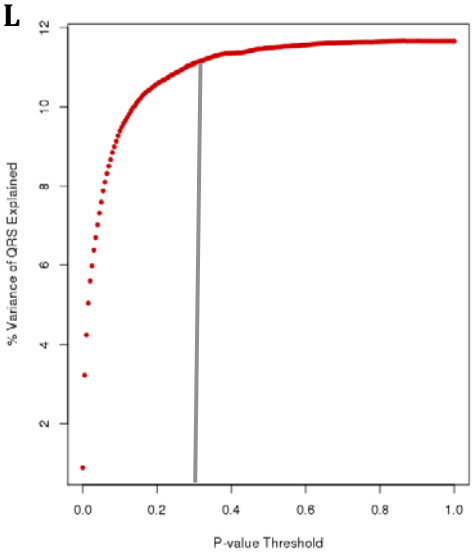
J


% Variance of FIadjBMI Explained
0
5
10
15
0.0
0.2
0.4
0.6
0.8
1.0
P-value Threshold

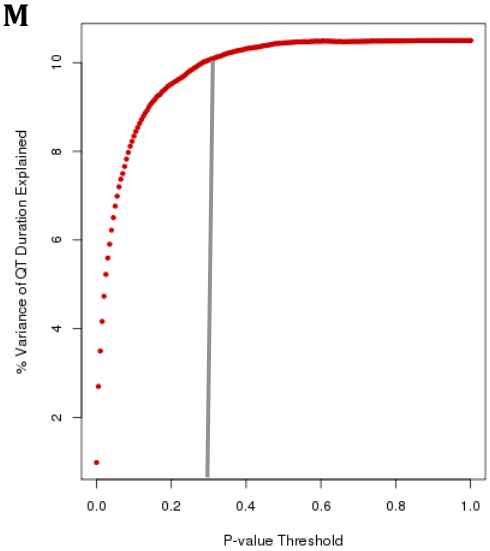
K


% Variance of HR Explained
1.0
1.5
2.0
2.5
0.0
0.2
0.4
0.6
0.8
1.0
P-value Threshold

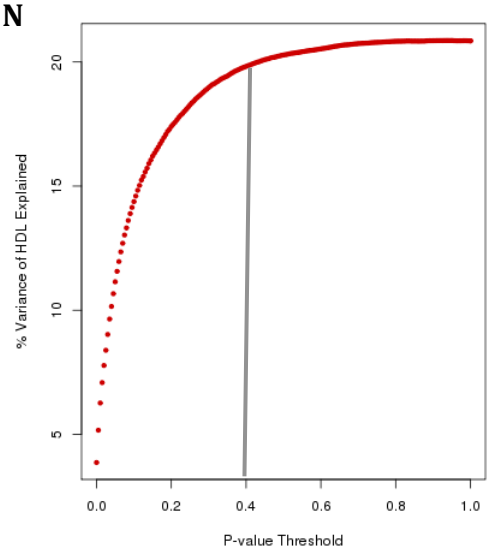
L


% Variance of QRS Explained
2
4
6
8
10
12
0.0
0.2
0.4
0.6
0.8
1.0
P-value Threshold

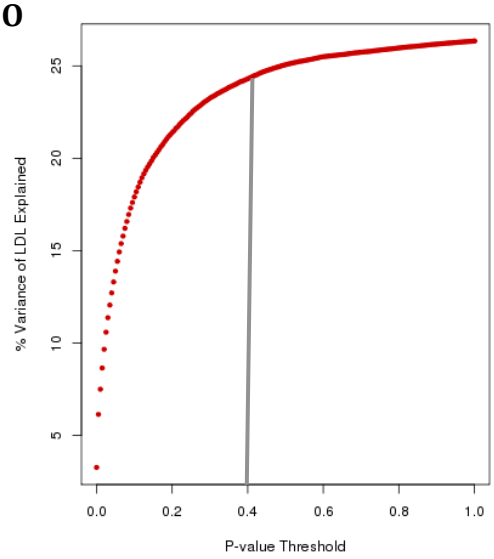
M


% Variance of QT Duration Explained
2
4
6
8
10
0.0
0.2
0.4
0.6
0.8
1.0
P-value Threshold

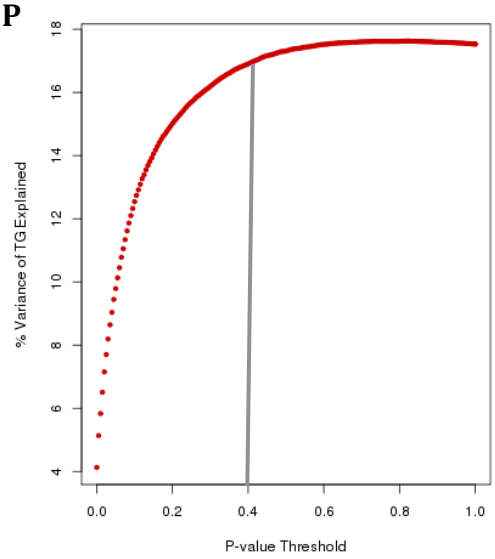
N


% Variance of HDL Explained
5
10
15
20
0.0
0.2
0.4
0.6
0.8
1.0
P-value Threshold

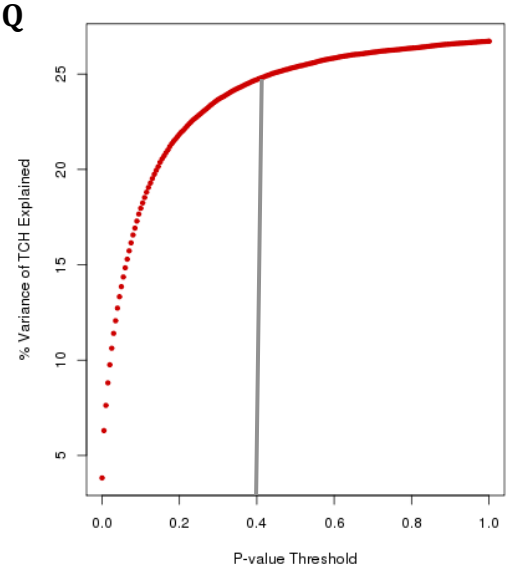
O


% Variance of LDL Explained
5
10
15
20
25
0.0
0.2
0.4
0.6
0.8
1.0
P-value Threshold

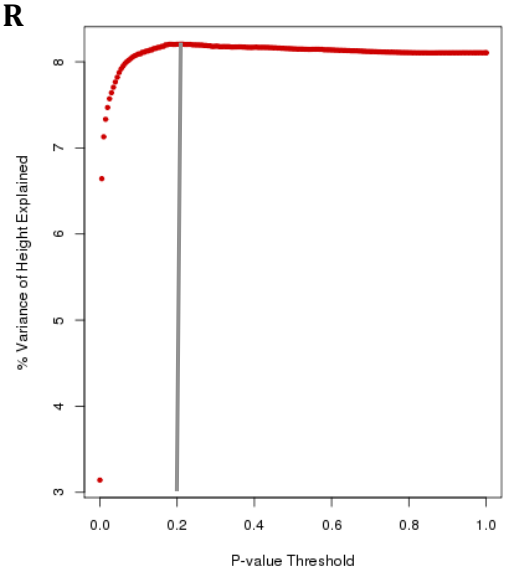
P


% Variance of TG Explained
4
6
8
10
12
14
16
18
0.0
0.2
0.4
0.6
0.8
1.0
P-value Threshold

Q

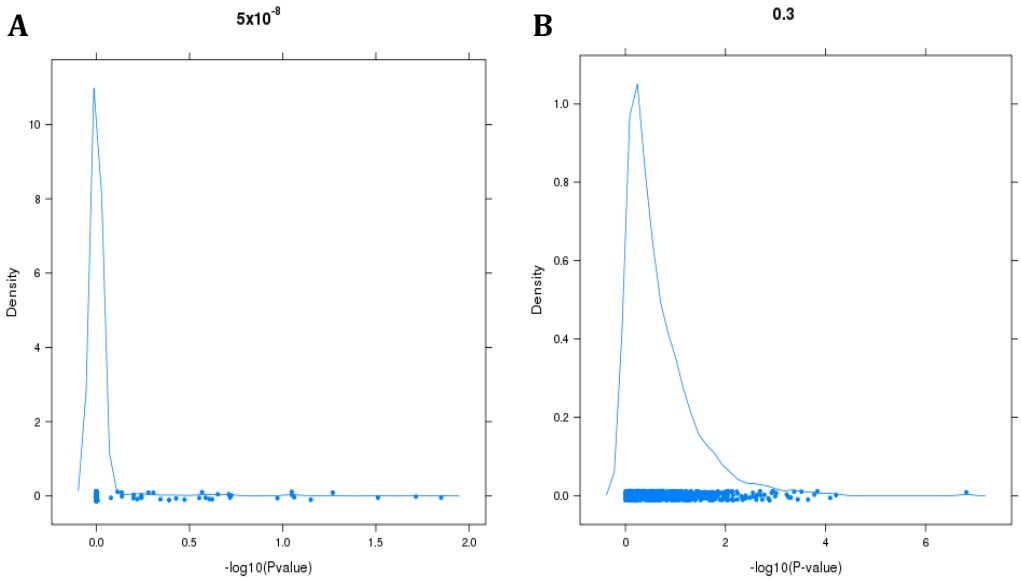

% Variance of TCH Explained
5
10
15
20
25
0.0
0.2
0.4
0.6
0.8
1.0
P-value Threshold

R


% Variance of Height Explained
3
4
5
6
7
8
0.0
0.2
0.4
0.6
0.8
1.0
P-value Threshold

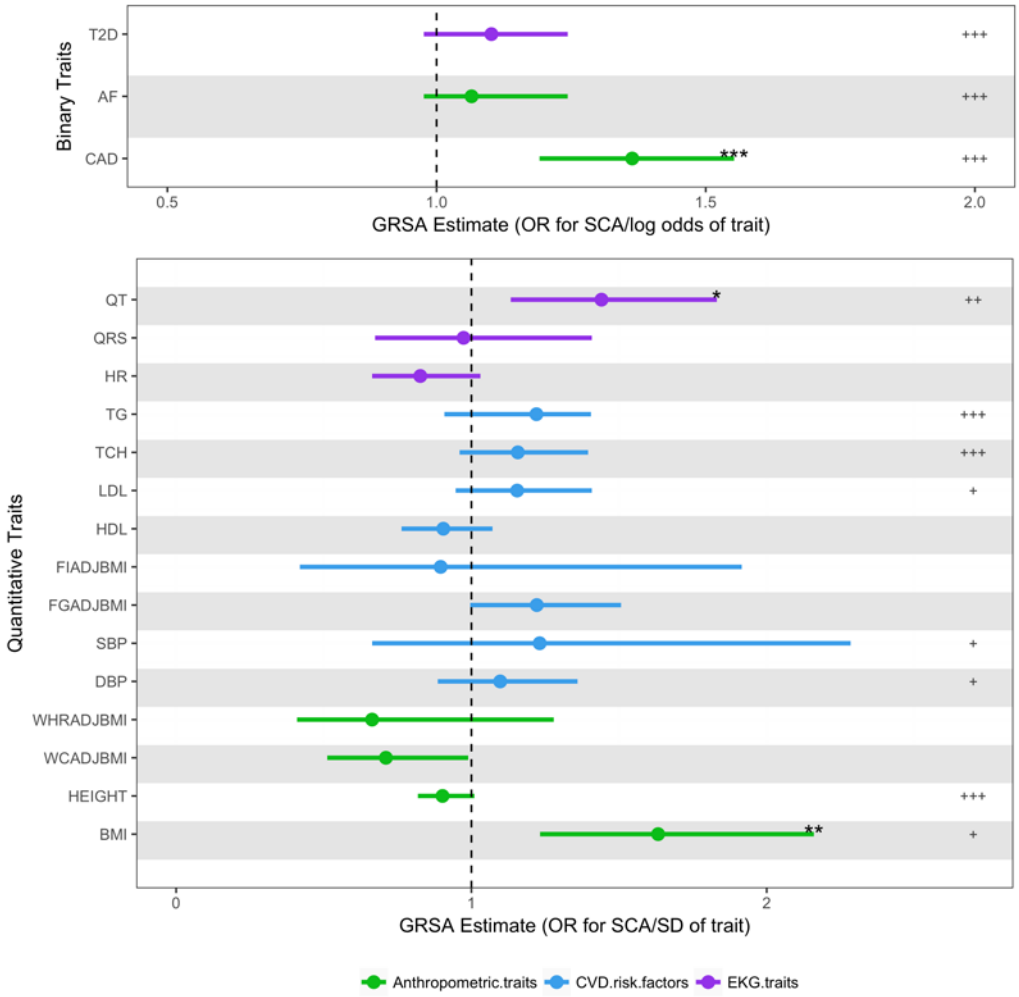
Each data point represents a different group of SNPs set by an increasing *P*-value threshold (x-axis) and the percent of variance of a trait explained by those SNPs (y-axis). These plots are used to determine the fewest number of SNPs required to maximize the variance explained, as indicated by the dotted gray line, and are reported in **Table 3.4** for each trait. (A) CAD = coronary artery disease; (B) T2D = type 2 diabetes; (C) AF = atrial fibrillation; (D) BMI = body mass index; (E) WCadjBMI = waist circumference adjusted for BMI; (F) WHRadjBMI = waist to hip ratio adjusted for BMI; (G) DBP = diastolic blood pressure; (H) SBP = systolic blood pressure; (I) FGadjBMI = fasting glucose adjusted for BMI; (J) FIadjBMI = fasting insulin adjusted for BMI; (K) HR = heart rate; (L) QRS = QRS interval; (M) QT = QT interval; (N) HDL = high-density lipoproteins; (O) LDL = low-density lipoproteins; (P) TG = triglycerides; (Q) TCH = total cholesterol; (R) Height.

Figure 3.3: *P*-value Distributions from 1000 Null Datasets



1000 dummy GWAS datasets were created using genotypes of 9,533 European participants from the ARIC cohort and 1000 randomly generated quantitative phenotypes (mean=0, sd=1). These datasets were subsequently used to compute a GRSA estimate for SCA with (A) $P<5x10^{-8}$ and (B) $P<0.3$. Each panel plots the $-\log_{10}(P\text{-value})$ of GRSs constructed from these datasets at the different alphas, and represents the null distribution of GRSA *P*-values. These null distributions were used to determine the permutated *P*-value in **Table 3.4.**

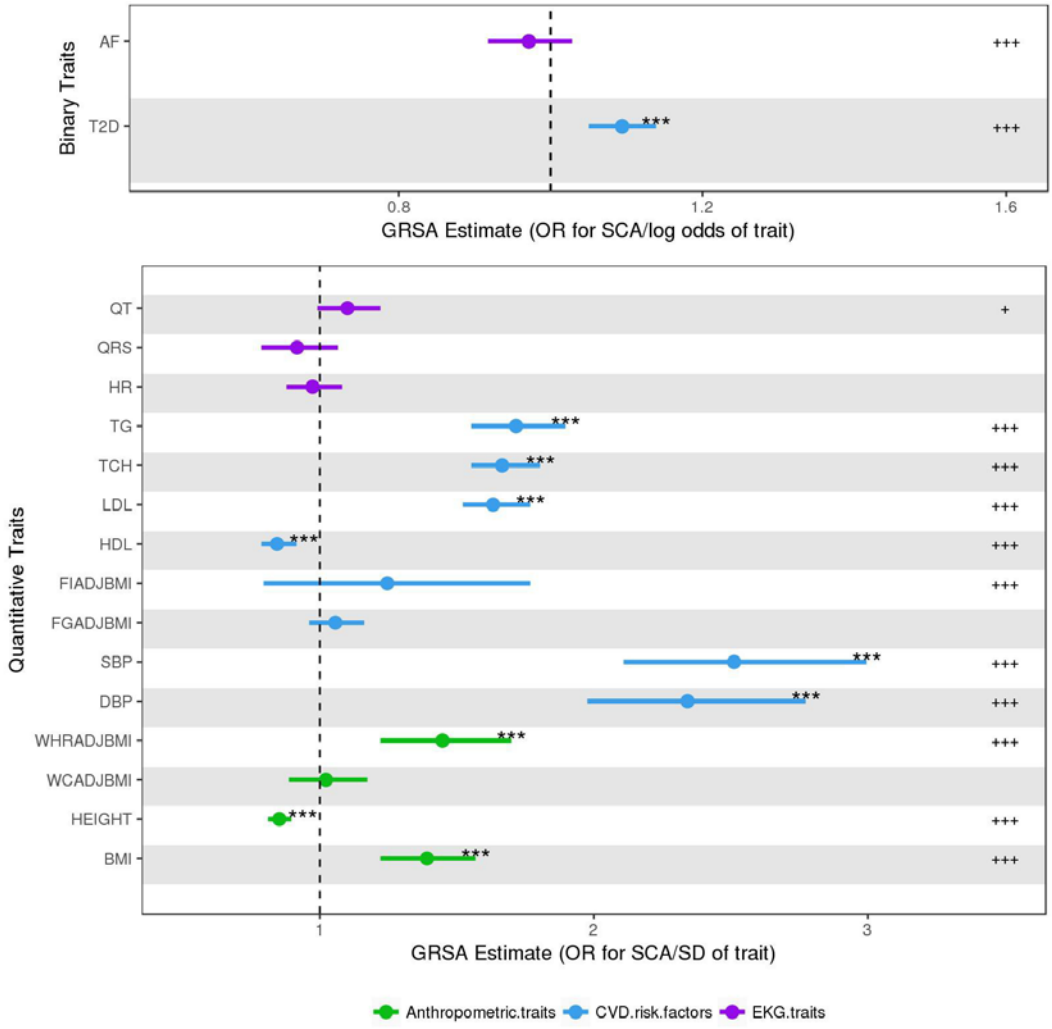
Figure 3.4: Genetic Risk Scores Association (GRSA) Estimates for SCA using SCA GWAS with 3,939 cases and 35,989 controls



These data points represent the exponentiated GRSA estimates of 18 traits on sudden cardiac arrest (SCA) and corresponding 95% confidence interval values using the previously published SCA GWAS with 3,989 cases and 25,989 controls. The GRSA estimates in the top panel for the binary traits are in log odds units. Values in bottom panel are in SD units of the quantitative traits. GRSA estimates and significance are derived from SNPs associated with each trait at $P<5x10^{-8}$.

The significance of the $GRSA_{GWS}$ estimates (FDR adjusted P_{GWS}) are represented as "*" for $P<0.05$, "**" for $P<0.01$, and "***" for $P<0.001$. The significance of the secondary analysis using $GRSA_{max}$ estimates (FDR adjusted permuted P_{max}) are represented as "+" for $P<0.05$, "++" for $P<0.01$ and "+++" for $P<0.001$. For details on values of GRSA estimates and *P*-values, see **Table 3.3.** CAD = coronary artery disease; T2D = type 2 diabetes; AF = atrial fibrillation; BMI = body mass index; WCadjBMI = waist circumference adjusted for BMI; WHRadBMI = waist to hip ratio adjusted for BMI; DBP = diastolic blood pressure; SBP = systolic blood pressure; FGadjBMI = fasting glucose adjusted for BMI; FIadjBMI = fasting insulin adjusted for BMI; HR = heart rate; QRS = QRS interval; QT = QT interval; HDL = high-density lipoproteins; LDL = low-density lipoproteins; TCH = total cholesterol; TG = triglycerides.

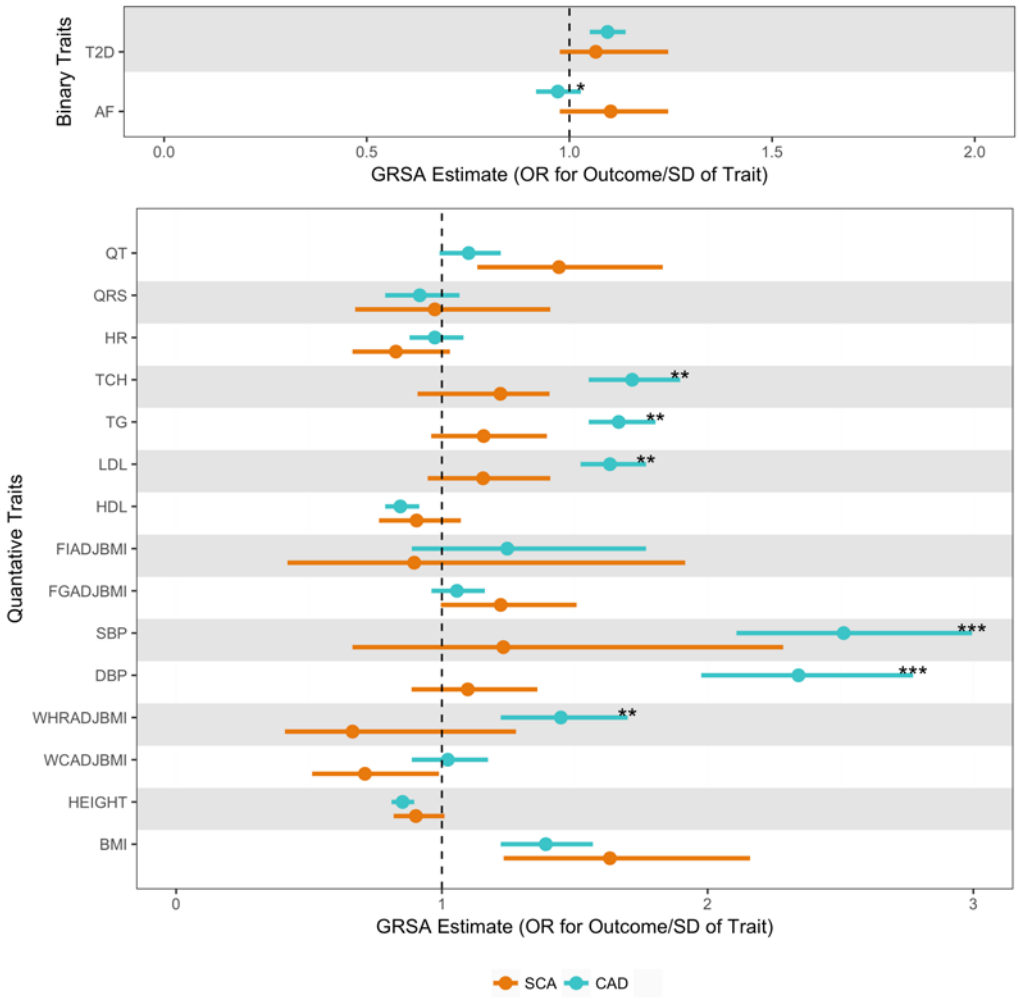
Figure 3.5 Genetic Risk Scores Association (GRSA) Estimates for CAD



These data points represent the exponentiated GRSA estimates of 17 traits on coronary artery disease (CAD) and corresponding 95% confidence interval values. The GRSA estimates in the top panel for the binary traits are in log odds units. Values in bottom panel are in SD units of the quantitative traits. GRSA estimates and significance are derived from SNPs at $P<5x10^{-8}$. The significance of the $GRSA_{GWS}$ estimates (P_{GWS}) are represented as "*" for $P<0.05$, "**" for $P<0.01$, and "***" for $P<0.001$. The significance of the secondary analysis using $GRSA_{max}$ estimates (permuted P_{max}) are represented as "+" for $P<0.05$, "++" for

P<0.01 and "+++" for *P*<0.001. For details on values of GRSA estimates and *P*-values, see **Table 3.3.** T2D = type 2 diabetes; AF = atrial fibrillation; BMI = body mass index; WCadjBMI = waist circumference adjusted for BMI; WHRadBMI = waist to hip ratio adjusted for BMI; DBP = diastolic blood pressure; SBP = systolic blood pressure; FGadjBMI = fasting glucose adjusted for BMI; FIadjBMI = fasting insulin adjusted for BMI; HR = heart rate; QRS = QRS interval; QT = QT interval; HDL = high-density lipoproteins; LDL = low-density lipoproteins; TCH = total cholesterol; TG = triglycerides.

Figure 3.6: Comparison of GRSA for SCA and CAD.



These data represent exponentiated GRSAs of all 17 traits. GRSA estimates for SCA and CAD, are plotted in orange and teal respectively. Bars around the estimates represent the 95% confidence interval. The GRSA estimates in the top panel for the binary traits are in log odds units. Values in bottom panel are in SD units of the quantitative traits. The level of significance for 1 degree of freedom Wald test of difference in $GRSA_{GWS}$ estimates between SCA and CAD is represented "*" for $P<0.05$, "**" for $P<0.01$, and "***" for $P<0.001$. T2D = type 2 diabetes; AF = atrial fibrillation; BMI = body mass index; WCadjBMI = waist

circumference adjusted for BMI; WHRadBMI = waist to hip ratio adjusted for BMI; DBP = diastolic blood pressure; SBP = systolic blood pressure; FGadjBMI = fasting glucose adjusted for BMI; FIadjBMI = fasting insulin adjusted for BMI; HR = heart rate; QRS = QRS interval; QT = QT interval; HDL = high-density lipoproteins; LDL = low-density lipoproteins; TCH = total cholesterol; TG = triglycerides.

3.6 Tables

Table 3.1 Study Details

Trait	Abbreviation	Consortium	Title	Total Sample Size
Coronary Artery Disease	CAD	CARDIOGRAM+C4D	Large-scale association analysis identifies new risk loci for coronary artery disease (Deloukas et al., 2012)	194,427
Type 2 Diabetes	T2D	DIAGRAM	Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes (Morris et al, 2012)	149,821
Atrial Fibrillaiton	AF	Atrial Fibrillation Genetics (AFGen)	Large-scale analyses of common and rare variants identify 12 new loci associated with atrial fibrillation (Christophersen et al., 2017)	110,099
Body-mass Index	BMI	GIANT	Genetic studies of body mass index yield new insights for obesity biology (Locke et al., 2015)	339,224
Height	HEIGHT	GIANT	Defining the role of common variation in the genomic and biological architecture of adult human height (Wood et al., 2014)	253,288
Waist Circumference adjusted for BMI	WCADJBMI	GIANT	New genetic loci link adipose and insulin biology to body fat distribution (Shungin et al., 2015)	224,459
Waist-to-Hip Ratio adjusted for BMI	WHRADJBMI	GIANT	New genetic loci link adipose and insulin biology to body fat distribution (Shungin et al., 2015)	142,762
Diastolic Blood Pressure	DBP	ICBP	Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk (Ehret et al., 2011) & The genetics of blood pressure regulation and its target organs from association studies in 342,415 individuals (Ehret et al., 2016)	69,395 / 201,529

Trait	Abbreviation	Consortium	Title	Total Sample Size
Systolic Blood Pressure	SBP	ICBP	Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk (Ehret et al., 2011) & The genetics of blood pressure regulation and its target organs from association	69,395 / 201,529
Fasting Glucose adjusted for BMI	FGADJBMI	DIAGRAM	A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance (Manning et al., 2012)	58,074
Fasting Insulin adjusted for BMI	FIADJBMI	DIAGRAM	A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance (Manning et al., 2012)	51,750
Heart Rate	HR	UK Biobank	Identification of genomic loci associated with resting heart rate and shared genetic predictors with all-cause mortality (Eppinga et al., 2016)	134,251
QRS Interval	QRS	CHARGE QRS	52 Genetic Loci Influencing Myocardial Mass (van der Harst et al., 2016)	60,255
QT Interval	QT	QT-IGC	Genetic association study of QT interval highlights role for calcium signaling pathways in myocardial repolarization (Arking et al., 2014)	76,061
High-Density Lipoprotein	HDL	Global Lipids Genetics Consortium	Discovery and refinement of loci associated with lipid levels (Willer et al., 2013)	188,577
Low-Density Lipoprotein	LDL	Global Lipids Genetics Consortium	Discovery and refinement of loci associated with lipid levels (Willer et al., 2013)	188,577
Total Cholesterol	TCH	Global Lipids Genetics Consortium	Discovery and refinement of loci associated with lipid levels (Willer et al., 2013)	188,577
Triglycerides	TG	Global Lipids Genetics Consortium	Discovery and refinement of loci associated with lipid levels (Willer et al., 2013)	188,577

Table 3.2 Comparison of Mendelian Randomization Methods

Trait	Number of SNPs	IVW Estimate [95% CI]	IVW *P*-value	MR-Egger Estimate [95% CI]	MR-Egger *P*-value	MR-Egger Intercept	MR-Egger Intercept *P* value
CAD	23	0.307 [0.175,0.440]	$5.16x10^{-6}$	0.365 [0.032,0.698]	0.032	-0.007	0.71
T2D	38	0.063 [-0.052,0.177]	0.29	0.014 [-0.265,0.293]	0.92	0.006	0.71
AF	19	0.097 [-0.024,0.218]	0.12	0.009 [-0.229,0.247]	0.94	0.013	0.40
BMI	72	0.488 [0.209,0.767]	$6.02x10^{-4}$	0.864 [0.183,1.55]	0.013	-0.012	0.24
Height	447	-0.100 [-0.212,0.013]	0.08	-0.06 [-0.368,0.248]	0.70	-0.001	0.79
WCadjBMI	67	-0.358 [-0.715,-0.001]	0.05	-0.429 [01.95,1.09]	0.58	0.002	0.92
WHRadjBMI	39	-0.404 [-0.864,0.056]	0.09	-1.92 [-4.26,0.419]	0.11	0.043	0.20
DBP	54	0.093 [-0.121,0.307]	0.39	0.363 [-0.224,0.951]	0.23	-0.013	0.33
SBP	54	0.208 [-0.410,0.826]	0.51	0.593 [-1.11,2.30]	0.49	-0.006	0.63
FGadjBMI	22	0.203 [-0.051,0.456]	0.12	0.432 [-0.103,0.966]	0.11	-0.019	0.34
FIadjBMI	9	-0.113 [-0.874,0.648]	0.77	-1.247 [-5.63,3.132]	0.58	0.036	0.61
HR	67	-0.188 [-0.414,0.038]	0.10	-0.760 [-1.38,-0.137]	0.017	0.023	0.05
QRS	22	-0.027 [-0.395,0.342]	0.89	-0.834 [-2.15,0.481]	0.21	0.036	0.21
QT	60	0.365 [0.125,0.605]	0.003	0.571 [0.101,1.041]	0.017	-0.009	0.32
HDL	89	-0.099 [-0.279,0.081]	0.28	0.230 [-0.096,0.556]	0.17	-0.017	0.019
LDL	68	0.142 [-0.070,0.355]	0.19	0.198 [-0.190,0.586]	0.32	-0.003	0.73
TCH	85	0.158 [-0.032,0.349]	0.10	0.021 [-0.340,0.381]	0.91	0.008	0.38
TG	58	0.15 [-0.075,0.376]	0.19	-0.050 [-0.415,0.315]	0.79	0.012	0.17

Trait	Number of SNPs	Simple Median Estimate [95% CI]	SimpleMedian *P*-value	Median Weighted Estimate [95% CI]	Median value
CAD	23	0.304 [0.100,0.507]	0.003	0.329 [0.128,0.531]	0.001
T2D	38	0.018 [-0.139,0.175]	0.82	0.056 [-0.106,0.218]	0.50
AF	19	0.158 [-0.035,0.351]	0.11	0.053 [-0.112,0.219]	0.53
BMI	72	0.426 [-0.025,0.877]	0.06	0.729 [0.250,1.21]	0.003
Height	447	-0.165 [-0.340,0.011]	0.07	-0.084 [-0.264,0.095]	0.36
WCadjBMI	67	-0.527 [-0.985,-0.069]	0.024	-0.519 [-0.977,-0.061]	0.026
WHRadjBMI	39	-0.45 [-1.02,0.122]	0.12	-0.547 [-1.12,0.022]	0.06
DBP	54	0.090 [-0.228,0.407]	0.58	0.070 [-0.239,0.380]	0.66
SBP	54	0.218 [-0.694,1.13]	0.64	0.180 [-0.707,1.07]	0.69
FGadjBMI	22	0.160 [-0.236,0.556]	0.43	0.472 [0.147,0.798]	0.004
FIadjBMI	9	-0.225 [-1.27,0.824]	0.67	-0.271 [-1.27,0.729]	0.60
HR	67	-0.184 [-0.531,0.162]	0.30	-0.260 [-0.601,0.082]	0.14
QRS	22	-0.061 [-0.535,0.413]	0.80	-0.074 [-0.543,0.395]	0.76
QT	60	0.428 [0.027,0.829]	0.037	0.470 [0.106,0.835]	0.011
HDL	89	-0.053 [-0.347,0.240]	0.72	0.029 [-0.254,0.312]	0.84
LDL	68	0.088 [-0.225,0.400]	0.58	0.196 [-0.100,0.491]	0.19
TCH	85	0.100 [-0.183,0.383]	0.49	0.025 [-0.230,0.280]	0.85
TG	58	0.066 [-0.308,0.439]	0.73	-0.129 [-0.449,0.190]	0.43

Table 3.3A: SCA GRSA with 18 traits using SNPs with P<5x10^{-8}

TRAIT	Number of SNPs	GRSA estimate [95% CI]	*P*-value for GRSA	FDR adjusted *P*-value
CAD	23	0.307 [0.175,0.440]	5.16x10^{-6}	9.29x10^{-5}
T2D	38	0.063 [-0.052,0.177]	0.29	0.38
AF	19	0.097 [-0.024,0.218]	0.12	0.24
BMI	72	0.488 [0.209,0.767]	6.02x10^{-4}	5.42x10^{-3}
HEIGHT	449	-0.100 [-0.212,0.013]	0.08	0.29
WCADJBMI	65	-0.358 [-0.715,-0.001]	0.05	0.23
WHRADJBM	39	-0.404 [-0.864,0.056]	0.09	0.27
DBP	54	0.093 [-0.121,0.307]	0.39	0.47
SBP	54	0.208 [-0.410,0.826]	0.51	0.58
FGADJBMI	22	0.203 [-0.051,0.456]	0.12	0.22
FIADJBMI	9	-0.113 [-0.874,0.648]	0.77	0.82
HR	67	-0.188 [-0.414,0.038]	0.10	0.26
QRS	22	-0.027 [-0.395,0.342]	0.89	0.89
QT	60	0.365 [0.125,0.605]	0.003	0.018
HDL	90	-0.099 [-0.279,0.081]	0.28	0.39
LDL	66	0.142 [-0.070,0.355]	0.19	0.32
TCH	84	0.158 [-0.032,0.349]	0.10	0.23
TG	55	0.150 [-0.075,0.376]	0.19	0.29

Table 3.3B: CAD GRSA with 17 traits using SNPs with P<5x10^{-8}

TRAIT	Number of SNPs	GRSA estimate [95% CI]	*P*-value for GRSA	FDR adjusted *P*-value	*P*-value for difference in GRSA between SCA and CAD
T2D	38	0.090 [0.049, 0.131]	1.66x10^{-5}	3.53x10^{-5}	0.817
AF	19	-0.029 [-0.086, 0.028]	0.317	0.337	0.017
BMI	71	0.327 [0.201, 0.452]	3.30x10^{-7}	8.01x10^{-7}	0.303
HEIGHT	455	-0.156 [-0.207, -0.106]	1.07x10^{-9}	3.03x10^{-9}	0.269
WCADJBMI	67	0.022 [-0.116, 0.160]	1.66x10^{-5}	3.14x10^{-5}	0.066
WHRADJBM	40	0.365 [0.195, 0.535]	2.51x10^{-5}	3.88x10^{-5}	0.005
DBP	54	0.489 [0.396, 0.583]	1.41x10^{-24}	5.99x10^{-24}	0.023
SBP	54	1.37 [1.10, 1.64]	2.62x10^{-23}	8.90x10^{-23}	2.00x10^{-4}
FGADJBMI	22	0.0547 [-0.040, 0.150]	0.258	0.313	0.277
FIADJBMI	9	0.223 [-0.121, 0.566]	0.204	0.267	0.426
HR	63	-0.026 [-0.128, 0.075]	0.610	0.610	0.208
QRS	21	-0.087 [-0.237, 0.064]	0.258	0.292	0.776
QT	61	0.096 [-0.001, 0.201]	0.073	0.103	0.063
HDL	82	-0.166 [-0.243, -0.089]	2.16x10^{-5}	3.67x10^{-5}	0.533
LDL	66	0.493 [0.419, 0.566]	1.41x10^{-39}	1.20x10^{-38}	0.001
TCH	80	0.515 [0.444, 0.586]	9.65x10^{-46}	1.64x10^{-44}	0.004
TG	54	0.538 [0.440, 0.637]	6.31x10^{-27}	3.58x10^{-26}	0.003

Table 3.4A: SCA GRSA for 18 traits using SNPs from custom P-value cutoff

TRAIT	Pvalue Threshold	Number of SNPs	GRSA estimate (95% CI)	*P*-value for GRSA	Empirical *P* value from 1000 permutations	FDR adjusted *P*-
CAD	0.4	32,252	0.050 [0.039, 0.061]	4.81×10^{-19}	<0.001	<0.001
T2D	0.3	29,276	0.041 [0.031, 0.050]	1.72×10^{-17}	<0.001	<0.001
AF	0.3	31,196	0.045 [0.032, 0.058]	2.73×10^{-11}	<0.001	<0.001
BMI	0.4	29,478	0.095 [0.044, 0.146]	2.57×10^{-4}	0.006	0.012
HEIGHT	0.2	19,651	-0.162 [-0.215, -0.110]	1.29×10^{-9}	<0.001	<0.001
WCADJBMI	0.3	27,081	-0.044 [-0.089, -0.001]	0.047	0.13	0.18
WHRADJBMI	0.3	28,111	-0.004 [-0.047, 0.039]	0.85	0.89	0.89
DBP	0.3	27,871	0.051 [0.020, 0.081)	0.001	0.017	0.031
SBP	0.3	27,820	0.052 [0.018, 0.085]	0.002	0.023	0.038
FGADJBMI	0.2	24,415	0.003 [-0.023, 0.029]	0.82	0.88	0.93
FIADJBMI	0.4	34,051	-0.014 [-0.051, 0.023]	0.45	0.57	0.69
HR	0.2	24,471	-0.013 [-0.056, 0.030]	0.56	0.66	0.74
QRS	0.3	29,636	-0.030 [-0.067, 0.007]	0.11	0.24	0.30
QT	0.3	30,129	0.108 [0.056, 0.159]	4.06×10^{-5}	0.002	0.005
HDL	0.4	29,737	-0.050 [-0.088, -0.011]	0.011	0.05	0.08
LDL	0.4	30,177	0.066 [0.032, 0.101]	1.86×10^{-4}	0.005	0.011
TCH	0.4	30,036	0.097 [0.062, 0.133]	5.90×10^{-8}	<0.001	<0.001
TG	0.4	30,505	0.095 [0.057, 0.134]	9.89×10^{-7}	<0.001	<0.001

Table 3.4B: CAD GRSA for 17 traits using SNPs from custom P-value cutoff

TRAIT	Pvalue Threshold	Number of SNPs	GRSA estimate (95% CI)	*P*-value for GRSA	Empirical *P*-value from 1000 permutations	FDR adjusted *P*-value
T2D	0.3	28,184	0.051 [0.047, 0.056]	2.30×10^{-109}	<0.001	<0.001
AF	0.3	29,300	0.033 [0.026, 0.039]	2.46×10^{-24}	<0.001	<0.001
BMI	0.4	27,872	0.089 [0.065, 0.113]	3.93×10^{-13}	<0.001	<0.001
HEIGHT	0.2	18,695	-0.117 [-0.142, -0.092]	1.65×10^{-20}	<0.001	<0.001
WCADJBMI	0.3	25,780	-0.003 [-0.023, 0.018]	0.80	0.85	0.85
WHRADJBMI	0.3	26,709	0.062 [0.042, 0.082]	2.33×10^{-9}	<0.001	<0.001
DBP	0.3	26,029	0.054 [0.038, 0.067]	1.55×10^{-11}	<0.001	<0.001
SBP	0.3	25,859	0.078 [0.061, 0.095]	2.43×10^{-19}	<0.001	<0.001
FGADJBMI	0.2	22,814	0.010 [-0.003, 0.023]	0.14	0.26	0.29
FIADJBMI	0.4	31,601	0.043 [0.024, 0.063]	1.47×10^{-5}	<0.001	<0.001
HR	0.2	22,991	0.014 [-0.006, 0.034]	0.17	0.29	0.31
QRS	0.3	27,648	0.018 [0.001, 0.049]	0.046	0.13	0.15
QT	0.3	27,905	0.044 [0.012, 0.068]	6.95×10^{-4}	0.011	0.014
HDL	0.4	28,799	-0.062 [-0.075, -0.049]	1.24×10^{-11}	<0.001	<0.001
LDL	0.4	29,412	0.078 [0.062, 0.094]	3.83×10^{-21}	<0.001	<0.001
TCH	0.4	29,101	0.050 [0.034, 0.067]	4.07×10^{-9}	<0.001	<0.001
TG	0.4	29,650	0.075 [0.057, 0.093]	2.27×10^{-16}	<0.001	<0.001

Table 3.5A: Sex-stratified SCA GRSA for 18 traits using SNPs with P<5x10^{-8} -Men

Trait	Number of SNPs	GRSA estimate (95% CI)	*P*-value for GRSA
CAD	23	0.329 [0.168, 0.490]	5.98x10^{-5}
T2D	39	0.021 [-0.100, 0.141]	0.74
AF	19	0.063 [-0.083, 0.209]	0.40
BMI	72	0.603 [0.269, 0.937]	3.97x10^{-4}
HEIGHT	462	-0.023 [-0.157, 0.110]	0.73
WCADJBMI	68	-0.268 [-0.641, 0.105]	0.16
WHRADJBMI	40	-0.494 [-0.963, -0.026]	0.039
DBP	47	0.156 [-0.323, 0.646]	0.52
SBP	45	0.149 [-0.362, 0.660]	0.57
FGADJBMI	22	0.072 [-0.177, 0.321]	0.57
FIADJBMI	9	-0.687 [-1.61, 0.232]	0.14
HR	68	-0.205 [-0.477, 0.066]	0.14
QRS	22	0.063 [-0.351, 0.477]	0.77
QT	64	0.245 [-0.037, 0.528]	0.09
HDL	90	0.026 [-0.181, 0.232]	0.81
LDL	70	0.253 [0.051, 0.455]	0.014
TCH	86	0.189 [-0.001, 0.380]	0.05
TG	59	0.077 [-0.160, 0.313]	0.53

Table 3.5B: Sex-stratified SCA GRSA for 18 traits using SNPs with P<5x10^{-8} -Women

Trait	Number of SNPs	GRSA estimate (95% CI)	*P*-value for GRSA	*P*-value for difference in GRSA by sex
CAD	23	0.267 [0.014, 0.519]	0.039	0.64
T2D	39	0.240 [0.053, 0.427]	0.012	0.048
AF	19	0.107 [-0.124, 0.338]	0.37	0.65
BMI	72	0.469 [-0.060, 0.997]	0.08	0.68
HEIGHT	464	-0.220 [-0.429, -0.011]	0.039	0.15
WCADJBMI	68	-0.492 [-1.08, 0.093]	0.09	0.54
WHRADJBMI	40	-0.005 [-0.740, 0.730]	0.99	0.3
DBP	47	0.334 [-0.413, 1.08]	0.38	0.71
SBP	45	0.268 [-0.531, 1.07]	0.51	0.81
FGADJBMI	22	0.546 [0.155, 0.937]	0.006	0.044
FIADJBMI	9	0.778 [-0.693, 2.25]	0.30	0.11
HR	68	-0.235 [-0.663, 0.194]	0.28	0.11
QRS	22	-0.215 [-0.857, 0.427]	0.51	0.37
QT	63	0.615 [0.167, 1.06]	0.007	0.32
HDL	90	-0.417 [-0.743, -0.091]	0.012	0.043
LDL	71	-0.039 [-0.343, 0.264]	0.80	0.15
TCH	85	0.034 [-0.263, 0.331]	0.82	0.23
TG	60	0.283 [-0.094, 0.660]	0.14	0.47

Chapter 4: The effect of sex and underlying disease on the genetic association of QT interval and sudden cardiac arrest

4.1 Introduction

SCA is often the first manifestation of heart disease, particularly for women; several studies have found that women are less likely than men to have a prior history of known cardiac disease.[15,61] It has been hypothesized that SCA is a much more heterogeneous condition in women, potentially due to the different underlying diseases, leading to differences in the associated risk factors.

Prolonged QT interval, a measure of ventricular repolarization, has been previously established as a risk factor for SCA,[62,63] and recent studies using Mendelian randomization have demonstrated that this risk factor is causal.[38] Women, on average, exhibit longer QT intervals than men in the general population once puberty is reached.[64,65] In addition, a previous study found that the increase in risk for overall cardiac death associated with longer QT interval was more pronounced in women.[66] Women also have higher risk of arrhythmic events than men in the setting of inherited or acquired (drug-induced) QT prolongation.[67] Based on the sex differences in QT interval in the general population and its association with overall cardiac mortality, we hypothesize that the risk of SCA associated with longer QT interval could differ by sex. Likewise, we also hypothesize that QT interval could differentially affect SCA risk depending on the underlying pathology (e.g. ischemic vs. non-ischemic disease).

Previous studies have shown that ~34% of QT interval variation is heritable[68,69]. In addition, recent research indicates that ~21% of variation can be

explained by common autosomal SNPs found genome-wide, including SNPs in genes such as *KCNQ1*, *KCNH2*, *SCN5A* and *NOS1AP*.[70] The top SNP from the most recent QT interval genome-wide association study (GWAS) was the *NOS1AP* locus SNP rs12143842, which increased QT interval by 3.50 ms per T-allele (p-value=$1x10^{-213}$)[33] and accounts for ~1% of the variation in QT interval.[71] This SNP has been previously associated with increased SCA risk[72,73], and has also been found to have stronger effect on QT interval in women than men.[71]

In this study we examined a large Finnish study of post-mortem autopsy-confirmed SCA subjects to study the genetic association between QT interval and SCA risk. More specifically, we compared the association of the *NOS1AP* locus variant rs12143842 with SCA risk between subjects with underlying ischemic vs. non-ischemic disease. We also performed sex-stratified analyses within these groups to investigate any sex-specific association of the *NOS1AP* locus SNP with SCA risk. Finally, we performed Mendelian randomization to test for differences in the causal association between a previously identified causal risk factor, longer QT interval, and SCA in the setting of different underlying disease and/or between sexes.

4.2 Methods

4.2.1 Genotyping

Samples from the Fingesture and NFBC1966 studies, as described in Chapter 3, were genotyped for rs12143842 using five different platforms: Illumina Infinium Global Screening Array (GSA); Affymetrix Genome-wide Human SNP Array 6.0; Agena Biosciences MassARRAY; Applied Biosystems

Taqman real-time PCR; and Illumina TruSeq sequencing. All genotyping and sequencing were performed according to the manufacturer's instructions. Quality control was performed separately on each dataset before merging. Dataset and QC information is summarized in **Table 4.1**. Overlapping samples between platforms were used to evaluate the accuracy of the genotyping (reported in Supplementary Table 1). After exclusions, the study population included 2,282 SCA victims and 3,561 Finnish controls.

4.2.2 Statistical Analysis

P-values for differences in the Fingesture study characteristics were calculated using a two sample t-test for continuous variables and Pearson chi-square test for categorical variables. The genotypes for rs12143842 for all samples were merged and logistic regression was performed using R (version 3.3.3), with sex as the only covariate. The SCD cases were stratified by sex and underlying disease (ischemic, non-ischemic and other disease) to examine the SNP effects in each group. Differences between sexes were determined by incorporating an interaction term into the regression model. P-values for differences in effect sizes between the underlying disease groups were obtained from a 1-degree of freedom Wald test. Multi-dimensional scaling (MDS) using PLINKv1.9 was used for samples run on the GSA microarray (1,168 cases/761 controls) to assess potential population substructure between the Fingesture and NFBC1966 studies. MDS is a method that reduces the high number of dimensions (i.e. the number of SNPs) to a smaller number of dimensions based on similarities in the data and orders these MDS dimensions (called components) based on the

amount of variation explained in the data.[74] Most often, population substructure accounts for the most variation within the data and is captured in the first several MDS components.

4.2.3 Mendelian Randomization

While association tests establish observational relationships between a trait (i.e. QT interval) and an outcome (i.e. SCD), they cannot establish causality. Confounding variables, variables affecting both the trait and the outcome, can result in false positive associations. Mendelian randomization circumvents these potential confounders to establish causality by exploiting certain characteristics of SNPs: that they are (1) assigned at conception and (2) randomly distributed in the large population.[60,75] Mendelian randomization has other assumptions that must be met as well, including the absence of pleiotropy.[76] This assumption is often hard to fully meet, leading to potential bias of the results. However, recent methods have been developed to remove potentially pleiotropic SNPs in order to meet this assumption.

Mendelian randomization uses genetic variants as instrumental variables to test for causal relationships between a trait and an outcome. We used a multi-SNP genetic risk score association (GRSA) model to test for causality between QT interval and SCD in our stratified datasets. The SNPs used in the model are known to be associated with the trait of interest. In this study, we used genome-wide significant SNPs from the most recent QT interval GWAS.[33] The SNPs were pruned for linkage disequilibrium (LD) using the 'clump' method in PLINKv1.9, which removes any SNP within a 1Mb window of the SNP with the lowest P-

value. This step is performed in order to remove any correlated SNPs and reduce any potential bias. The GRSA model uses 57 LD-pruned SNPs to compare the association of these SNPs with the trait of interest (β_{trait}) to the association of the SNPs with SCD ($\beta_{outcome}$) using the R package 'MendelianRandomization'. [46] Zero-intercept inverse-weighted (IVW) linear regression is used to calculate the GRSA estimate, which is the slope of the resultant regression line, and estimates the difference in log odds of SCD risk per SD increase in QT interval. We used the HEIDI-outlier method from the 'gsmr' R package to detect and remove potentially pleiotropic SNPs.[48] P-values for difference in GRSA estimates were obtained from a 1-degree of freedom Wald test.

Genome-wide SNP data is required for Mendelian randomization analyses and therefore only the Fingesture and NFBC1966 samples genotyped using the Infinium Global Screening Array (GSA) and imputed to the NHLBI Trans-Omics for Precision Medicine (TOPMed) imputation panel using the University of Michigan imputation server[77] were used in this analysis (1,168 SCD victims and 761 Finnish controls). Logistic regression for single SNP association tests were run using FASTv2.4.[37] We performed several stratified analyses, including by sex and underlying disease (ischemic and non-ischemic disease). There were a small number of SCD cases with other underlying disease genotyped on this array and therefore were only included in the overall analysis and sex-stratified analyses but were excluded from the underlying disease-stratified analysis and subsequent sex-stratified analyses.

4.3 Results

4.3.1 Sample population

The SCA population is comprised of a subset of the Fingesture study of Finnish SCA subjects with autopsy-confirmed assessment of underlying heart disease in whom DNA was available at the time of this study (n=2,282). Controls were drawn from the Northern Finland Birth Cohort of 1966 (NFBC1966) and are comprised of 3,561 Finnish individuals born in 1966. Characteristics of the Fingesture study are detailed in **Table 4.2**. Additional information about the different sample subgroups is provided in **Table 4.3**. To assess for potential population stratification, we ran multi-dimensional scaling (MDS) on a subset of the samples with genome-wide SNP data (1,168 cases/761 controls). We assessed the top 10 MDS components, which can be used to visualize potential population substructure, for association with SCD status to test for possible confounding of our SNP association results. We ran logistic regression for SCD status, including sex and the top 10 MDS components as covariates. Results are in **Table 4.4**. Plots for the top 10 MDS components, colored by SCA status, are found in **Figure 4.1**. MDS component 7 was associated with SCA status after multi-test correction (P<0.002) (**Table 4.4**), indicating the potential for confounding due to population substructure. However, combined, the top 10 components explained only 0.9% of the variance in SCA status, suggesting likely minimal impact. This minimal impact was confirmed by sensitivity analyses (described below).

4.3.2 *NOS1AP* locus SNP analysis

Given the previously established relationship between QT interval and SCA risk, and with *NOS1AP* locus SNPs and SCA in other cohorts[63,78], we first sought to assess the association between SCA and the *NOS1AP* locus SNP rs12143842. When analyzing all 2,282 SCA cases and 3,561 controls, the T allele of rs12143842 was significantly associated with increased SCA risk with an OR of 1.14 for each copy of the QT lengthening allele (95% CI, 1.04-1.25; P = 0.005). In sensitivity analyses, including the 10 top components from the MDS analysis in the model minimally increased the effect estimate (**Table 4.5**). All SNP association results are summarized in **Figure 4.2** and **Table 4.6**.

4.3.2a Ischemic vs. Non-ischemic

To explore whether the association of rs12143842 differs by underlying disease pathology, we stratified the SCA cases into those with (1) underlying ischemic heart disease (n=1,478), (2) non-ischemic heart disease (n=750), and (3) other pathologies (myocarditis, cardiac anomaly, and normal autopsy, n=54). The rs12143842 T allele had the strongest association in non-ischemic SCA individuals with an OR of 1.23 (95% CI, 1.07-1.39; P=0.003). A weaker non-significant association was observed in both ischemic SCA individuals (OR =1.09; 95% CI, 0.98-1.21; P=0.12), and those with other underlying conditions (OR = 1.11; 95% CI, 0.71-1.73; P=0.64).

4.3.2b Men vs. Women

Given that QT interval is a stronger SCA risk factor in men than women, and rs12143842 has a larger effect on QT interval in women than in men,[71] we

next investigated whether the effect of rs12143842 on SCA risk differed between men and women. We limited sex-stratified analyses to SCA cases with underlying ischemic and non-ischemic pathology and excluded those with other underlying conditions due to the small sample size of those with other conditions.

Among 1,862 SCA male victims and 1,641 male controls, the rs12143842 QT lengthening allele was marginally associated with an increased risk of SCA (OR of 1.11; 95% CI, 0.99-1.23; *P*=0.07). When stratified by underlying disease pathology, the association was significant among non-ischemic SCA males (579 cases/1,641 controls) with an OR of 1.17 (95% CI, 1.00-1.37; *P*=0.045), while there was no statistically significant association in ischemic SCA males (1,245 cases/1,641 controls) for SCA risk (OR =1.09; 95% CI, 0.96-1.23; *P*=0.18; *P* for difference between ischemic/non-ischemic males=0.48).

Overall, among 420 female SCA cases and 1,920 female controls, the rs12143842 QT lengthening allele was associated with increased SCA risk (OR of 1.24; 95% CI, 1.04-1.46; *P*=0.015). Similar to findings among men, a stronger association was observed in the non-ischemic SCA women (171 cases/1,920 controls), with the rs12143842 T allele association with a 1.37-fold increased SCA risk (95% CI, 1.07-1.75; *P*=0.013) than among ischemic SCA women (233 cases/1,920 controls) (OR = 1.11 for each copy of the variant allele; 95% CI, 0.88-1.38; *P*=0.39; *P* for difference between ischemic/non-ischemic women=0.08).

4.3.3 Mendelian Randomization of QT Interval

Using Mendelian randomization approaches, we have previously established that QT interval is causally associated with SCA.[38] To investigate whether these causal associations differ based on sex and underlying disease, we calculated genetic risk score association (GRSA) estimates using the genome-wide significant SNPs from the most recent QT interval GWAS.[33] Inverse-weighted (IVW) linear regression was performed to compare the effect of the SNP on QT interval to the effect of the SNP on SCA risk in the sex-stratified and underlying disease-stratified datasets. Results are summarized in **Figure 4.3** and **Table 4.7**.

Among all SCA victims (n=1,168 cases/761 controls), a one standard deviation (SD) increase in QT interval was associated with a 1.42-fold increased risk of SCA (95% CI, 0.83-2.45; *P*=0.20), which translates in our sample population to a 1.10-fold increased risk of SCD per 10 ms increase in QT interval (95% CI, 0.90-1.34, *P*=0.20). While not statistically significant, these findings are consistent with our previous work (previous findings: odds ratio in cardiac arrest risk per SD increase in QT, 1.44; 95% CI, 1.13-1.83; *P*=0.018)[38]. Similar to our findings with *NOS1AP* locus SNP rs12143842, we found that the causal relationship of QT interval and SCA differs between individuals with ischemic heart disease and individuals with non-ischemic disease. Among non-ischemic SCA victims (507 cases/761 controls), there was a 1.96-fold increase in SCA risk per SD increase in QT (95% CI, 1.00-3.82; *P*=0.05). By contrast, there was no

evidence of a causal association of QT interval with SCA among SCA cases with ischemic disease (611 cases/761 controls; OR= 0.88; 95% CI, 0.47-1.67; *P*=0.70).

Non-ischemic female SCA cases had the strongest causal association of QT interval with SCA (odds ratio in SCA risk per SD increase in QT, 3.60; 95% CI, 1.22-10.59; *P*=0.02). Non-ischemic males had a large but non-significant causal association estimate between QT interval and SCA (odds ratio in SCA risk per SD increase in QT, 1.47; 95% CI, 0.64-3.39; *P*=0.36). Among those with underlying ischemic disease, there was no evidence for a causal relationship of QT interval with SCA for men or women (odds ratio in SCA risk per SD increase in QT, 0.92; 95% CI, 0.41-2.05; *P*=0.84 and odds ratio in SCA risk per SD increase in QT, 0.80; 95% CI 0.22-2.94; *P*=0.74, respectively).

4.4 Discussion

In the general population, women have longer QT intervals than men; women experience a higher rate of arrhythmias in the setting of prolonged QT interval; and prolonged QT interval is causally associated with SCD. We therefore hypothesized that women would show a greater association between genetically determined longer QT interval and SCD. Given the different etiologies between ischemic and non-ischemic cardiac disease, we further hypothesized that the genetic association with longer QT interval would also differ between the different underlying diseases. Our results, while not conclusive, support both of these hypotheses. We found that rs12143842, the top QT interval-associated SNP from previous GWAS[33], was associated with SCD risk in our overall dataset. We observed a larger, yet not statistically significant, genetic effect on SCD risk in

non-ischemic individuals compared to ischemic individuals. Furthermore, the women with SCD in the setting of non-ischemic cardiac disease had the largest genetic effect of rs12143842 on SCD risk. Our Mendelian randomization analyses had similar findings; non-ischemic individuals showed a potential causal association between longer QT interval and SCD, and female non-ischemic individuals had the strongest causal association. By contrast, both the SNP association and Mendelian randomization analyses did not show evidence for a genetic (causal) association between QT interval and SCD due to underlying ischemic disease in men or women. These results suggest that SCD in the setting of ischemic disease may not be strongly influenced by myocardial repolarization (QT interval), or that the effect of longer QT interval on ischemic SCD risk is masked by other risk factors exerting a larger effect. While the differences in sex- and underlying disease-stratified associations were not statistically significant, the directionality of our findings is nevertheless consistent with our underlying hypotheses; together these results provide evidence that SCD risk in non-ischemic individuals, particularly women with non-ischemic disease, may be influenced by genetically determined QT interval.

The underlying cause(s) of the sex differences in the association between longer QT interval and SCD remains unknown, however, sex hormones may play a role. Studies have previously established that testosterone and progesterone shorten the QT interval, while estrogen lengthens the QT interval.[79,80] While the underlying mechanism is unknown, our findings support the hypothesis that non-ischemic individuals are more susceptible to the effects of longer QT interval on

developing SCD. Given that women already have underlying lengthened QT due to sex hormones, the addition of QT lengthening genetic susceptibility (i.e. the T allele of the *NOS1AP* SNP rs12143842) may result in the higher observed risk of SCD in women with non-ischemic disease.

While our study provides evidence for differences in SCD risk factors between both underlying disease and sex, several limitations should be noted. First, many of our analyses did not meet traditional statistical significance cut-offs, though we note that the directionality of the results is entirely consistent with our original hypotheses. The study is underpowered to detect interactions and thus, additional samples are necessary to confirm our results. Our findings in the subgroup analyses also require additional replication. Second, there is likely additional phenotypic heterogeneity within the underlying disease subgroups. The non-ischemic group, as noted in the supplementary methods, consists of eight different cardiac conditions. It is possible these different conditions, while similar in nature, may differ in their relationship between QT interval and SCD risk. Additional samples are needed to further stratify the non-ischemic group to investigate whether a particular condition is driving the association. Third, while our MDS components indicated potential population substructure within a subset of samples, when we included the components as covariates in our analysis, the effect was actually stronger. Therefore, not adjusting our main analysis for population substructure is likely resulting in a downward bias of the true association. Fourth, the NFBC1966 cohort used for our controls consisted of relatively young individuals (31 years old). Given the mean age of our SCD

cohort was 60 years old, it is likely some of our "controls" will go on to have an SCD event later in life, and by not excluding these individuals, we bias our estimates towards the null. Fifth, the Finnish population is quite homogenous and therefore our findings may not be able to be applicable to other populations, including other Europeans. Lastly, the strongest associations were seen in women and since women on average have lower rates of SCD, we have the least power to detect differences within this group. Nevertheless, our findings that female SCD victims with non-ischemic disease had the greatest association between longer QT interval and SCD risk were consistent between the various analyses performed, including both SNP association tests and Mendelian randomization. The directionality of our findings is consistent with our original hypothesis, which stated that the effect of longer QT interval will differ by underlying disease pathology and would be stronger in females than males.

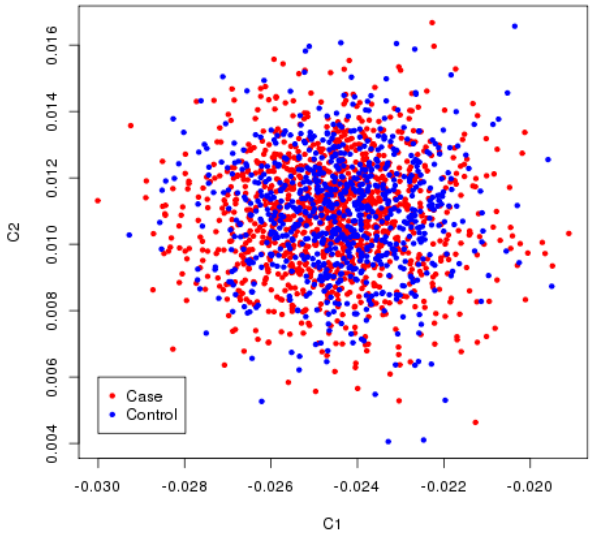
In conclusion, our study of autopsy-confirmed SCD victims provides consistent evidence to support the hypotheses that SCD risk factors, specifically lengthening of QT interval, may differ by both the underlying disease and sex. We found evidence of a genetic association in non-ischemic SCD victims, as well as a potentially causal association, between longer QT interval and SCD risk, with the largest genetic effect observed in female non-ischemic SCD individuals. SCD victims with underlying ischemic disease did not provide evidence for a genetic association, nor a causal association, between longer QT interval and SCD, regardless of sex. Our findings provide evidence that SCD risk factors,

particularly longer QT interval, may differ between sex and underlying disease etiology.

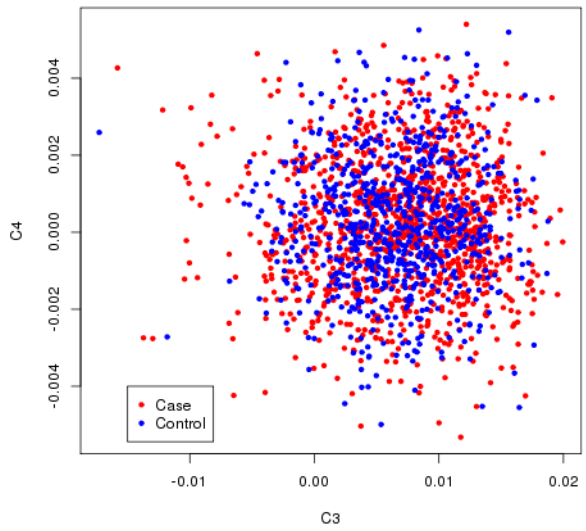
4.5 Figures

Figure 4.1 Multi-dimensional scaling (MDS) plot of Fingesture and NFBC1966 cohort samples

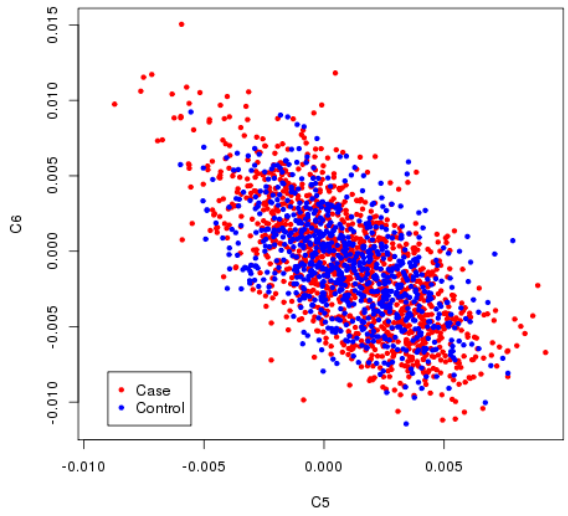
A. Component 1 vs. Component 2



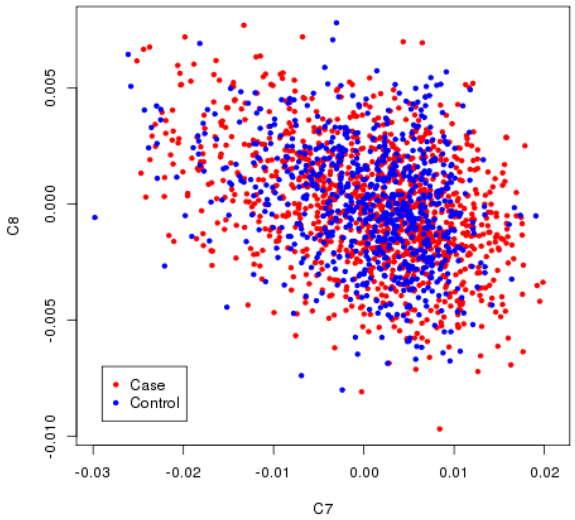
B. Component 3 vs. Component 4



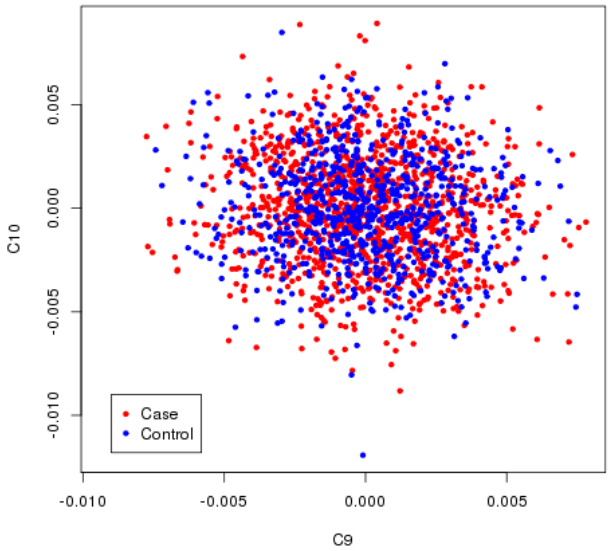
C. Component 5 vs. Component 6



D. Component 7 vs. Component 8

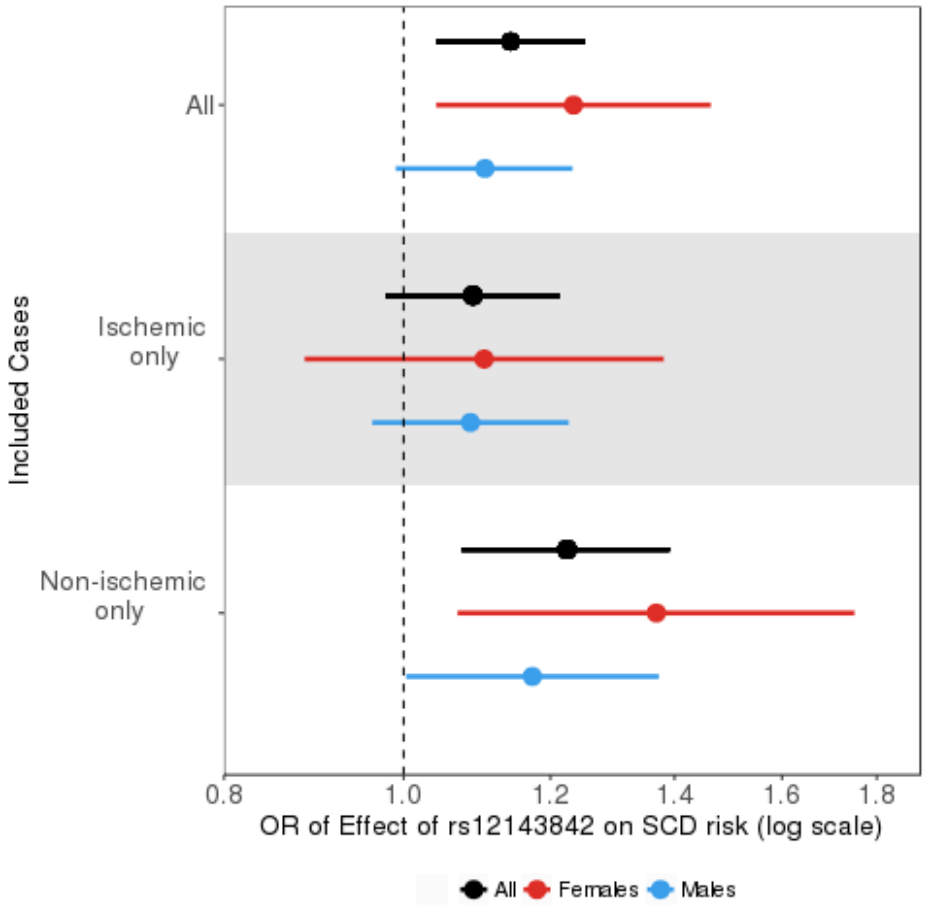


E. Component 9 vs. Component 10



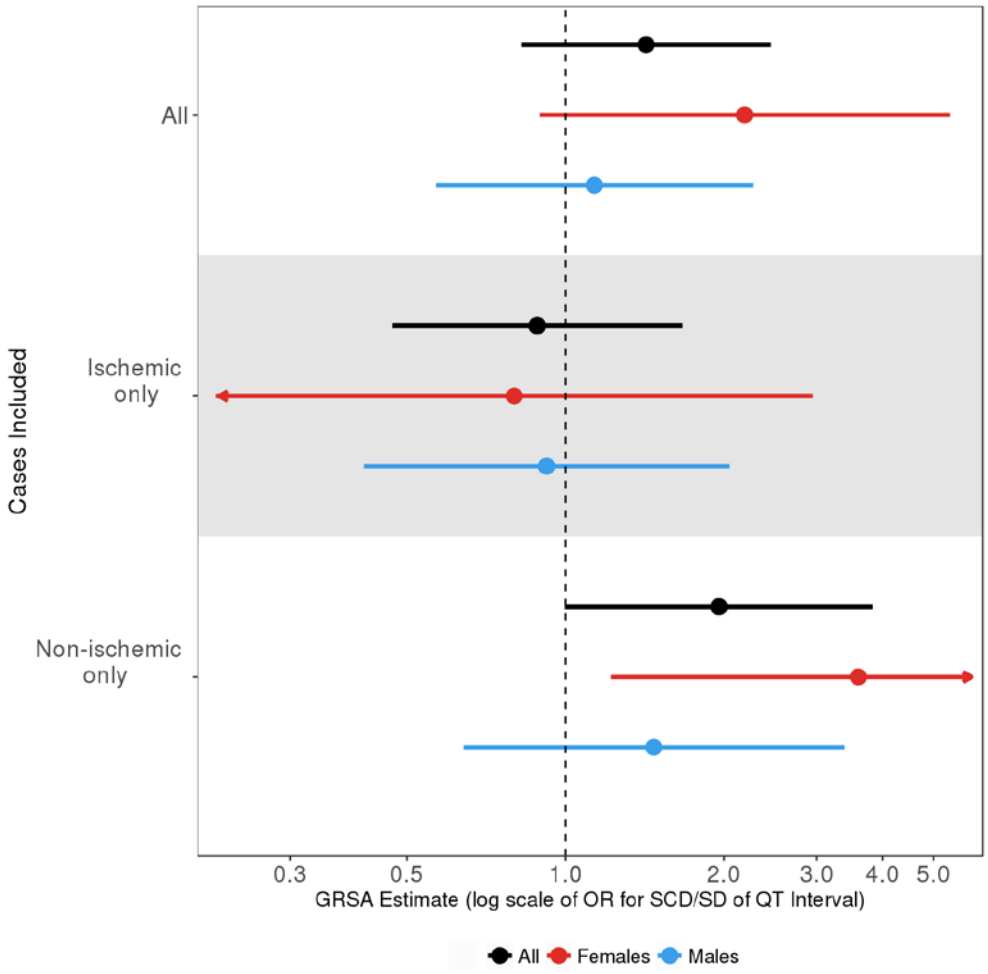
The plots A-E demonstrates strong genetic overlap between the Fingesture cohort (red) and the NFBC1966 cohort (blue).

Figure 4.2 Forest plot of the association of rs12143842 with SCA risk



The top white panel represents the analysis including all SCA victims (2,282 cases); the middle gray panel includes ischemic-only SCA victims (1,478 cases); and the bottom white panel includes only non-ischemic SCA victims (750 cases). The dots represent the odds ratio of the rs12143842 QT prolonging allele on SCA risk and the lines represent the 95% confidence intervals. Both sexes (black), females only (red), and males only (blue). Additional information found in **Table 4.6**.

Figure 4.3 Genetic risk score association (GRSA) estimates for QT interval with SCA



The data points in the top plot represent the exponentiated GRSA estimates of QT interval on SCA (in log odds of SCA/SD of QT interval) and corresponding 95% confidence intervals. The top white panel represents the analysis including all SCA cases used in the MR analysis (1,168 cases); the middle gray panel includes ischemic-only SCA cases (611 cases); the bottom white panel includes only non-ischemic SCA cases (507 cases). Each panel includes analyses using: both sexes

(black), females only (red), and males only (blue). Additional information found in **Table 4.7**.

4.6 Tables

Table 4.1 Genotyping Platform Sample Characteristics

Genotyping platform	Illumina Infinium Global Screening Array (GSA)	Affymetrix Genome-wide Human SNP Array 6.0	Agena Biosciences MassARRAY	AB Taqman	Illumina Sequencing
N, number of total cases	1168	358	574	572	825
N, number of total controls	761	NA	422	2175	563
N, number of independent cases	1168	315	122	496	181
N, number of independent controls	761	NA	251	2140	408
QC critieria	Sample and SNP call rate (<95%); sex check; duplicate removal; cryptic relatedeness; genetic outlier removal using PCA	Sample and SNP call rate (<95%); sex check; duplicate removal; cryptic relatedeness; genetic outlier removal using PCA	Sample and SNP call rate (<95%)	NA	Minimum SNP read depth (10x); Sample and SNP call rate (<95%); sex check; duplicate removal; cryptic relatedeness; genetic outlier removal using PCA
Sex, number of women among independent cases	218	50	30	91	31
Sex, number of women among independent controls	407	NA	145	1140	228
Age, mean age at SCD event	60.1	62.8	59.8	64.3	58
N, number of ischemic SCD cases	610	310	44	427	87
N, number of non-ischemic SCD cases	557	5	78	69	94
Number of non-matching alleles between overlap samples	0	0	1*	0	1*

*Same sample; removed from both analyses

Table 4.2 Fingesture study characteristics.

Variable	All (N=2,282)	Men (N=1,862)	Women (N=420)	*P**
Mean age, year (SD)	61.23 (10.71)	60.65 (10.43)	63.84 (11.56)	<0.001
N, ischemic disease (%)	1,478 (64.8%)	1,245 (66.9%)	233 (55.5%)	<0.001
N, non-ischemic disease (%)	750 (32.8%)	579 (31.1%)	171 (40.7%)	<0.001
N, other (%)	54 (2.4%)	38 (2.0%)	16 (3.8%)	0.03
BMI, kg/m^2 (SD)	28.36 (6.61)	28.16 (6.23)	29.26 (8.10)	0.06
Heart weight, g (SD)	493.60 (129.23)	509.60 (127.83)	421.40 (109.47)	<0.001

**P* calculated for difference between men and women

Table 4.3 Sample subgroup characteristics

Subgroup	N	Mean Age (SD)	N, Female	*NOS1AP* SNP T Allele Frequency
All Fingesture cases	2,282	61.23 (10.71)	420	0.264
Female Fingesture cases	420	63.84 (11.56)	420	0.285
Male Fingesture cases	1,862	60.65 (10.43)	0	0.259
Ischemic Fingesture cases	1,478	64.10 (9.70)	233	0.258
Non-ischemic Fingesture cases	750	56.22 (10.48)	171	0.276
Fingesture cases, age 30-55	658	48.12 (5.99)	93	0.270
Fingesture cases, age 56-85	1,604	66.70 (6.84)	322	0.262
All NFBC1966 controls	3,561	31 (0)	1,920	0.242
Female NFBC1966 controls	1,920	31 (0)	1,920	0.244
Male NFBC1966 controls	1,641	31 (0)	0	0.240

Table 4.4 Multi-dimensional scaling (MDS) regression results

Covariate	Beta	SE	*P*
Sex	-1.61	0.105	<0.001
MDS Component 1	-0.296	0.328	0.36
MDS Component 2	-0.231	0.288	0.42
MDS Component 3	0.477	0.189	0.011
MDS Component 4	-0.265	0.292	0.37
MDS Component 5	0.099	0.269	0.71
MDS Component 6	0.147	0.245	0.55
MDS Component 7	0.316	0.102	0.002
MDS Component 8	0.009	0.210	0.97
MDS Component 9	-0.002	0.197	0.99
MDS Component 10	0.001	0.194	0.99

*Components were re-scaled by mulitiplying by 100 before regression to avoid numerical errors in R

Table 4.5 Multi-dimensional scaling (MDS) regression results for rs12143842

Covariates used in model	Beta	SE	*P*	Variance Explained
Sex	0.211	0.083	0.011	0.101
Sex + MDS Components 1-10	0.227	0.084	0.007	0.108

Table 4.6 rs12143842 SNP association results

	All				
Dataset	**cases/controls**	**Beta**	**SE**	***P***	***P* for ischemic/ non-ischemic difference**
All cases/population controls	2282/3561	0.133	0.047	0.005	
ischemic cases/population controls	1478/3561	0.086	0.055	0.11	0.15
non-ischemic cases/population controls	750/3561	0.203	0.067	0.003	
Other cases/population controls	54/3561	0.106	0.226	0.64	

	Males only			
Dataset	**cases/controls**	**Beta**	**SE**	***P***
All cases/population controls	1862/1641	0.101	0.056	0.07
ischemic cases/population controls	1245/1641	0.083	0.062	0.18
non-ischemic cases/population controls	579/1641	0.160	0.080	0.045
Other cases/population controls	38/1641	-0.172	0.290	0.55

	Females only				
Dataset	**cases/controls**	**Beta**	**SE**	***P***	**P-value for interaction term (Sex*SNP)**
All cases/population controls	420/1920	0.211	0.087	0.015	0.14
ischemic cases/population controls	233/1920	0.100	0.114	0.39	0.86
non-ischemic cases/population controls	171/1920	0.314	0.126	0.013	0.14
Other cases/population controls	16/1920	0.649	0.377	0.09	0.015

Table 4.7 Mendelian randomization of QT interval results

	All				
Dataset	**cases/controls**	**SNPs included**	**GRSA Estimate [95% CI]**	***P***	***P* for ishemic/non-ischemic difference**
All cases/population controls	1168/761	57	0.352 [-0.191, 0.895]	0.20	
ischemic cases/population controls	611/761	57	-0.124 [-0.757, 0.510]	0.70	0.09
non-ischemic cases/population controls	507/761	57	0.671 [-0.003, 1.340]	0.05	

	Males only			
Dataset	**cases/controls**	**SNPs included**	**GRSA Estimate [95% CI]**	***P***
All cases/population controls	950/354	57	0.126 [-0.567, 0.820]	0.72
ischemic cases/population controls	528/354	57	-0.083 [-0.881,0.716]	0.84
non-ischemic cases/population controls	387/354	57	0.386 [-0.445,1.220]	0.36

	Females only				
Dataset	**cases/controls**	**SNPs included**	**GRSA Estimate [95% CI]**	***P***	***P* for male/female difference**
All cases/population controls	218/407	57	0.783 [-0.112, 1.680]	0.09	0.26
ischemic cases/population controls	83/407	57	-0.224 [-1.530, 1.080]	0.74	0.86
non-ischemic cases/population controls	120/407	57	1.28 [0.202, 2.360]	0.020	0.20

References

1. Benjamin, E. J. *et al.* Heart Disease and Stroke Statistics—2018 Update: A Report From the American Heart Association. *Circulation* **137,** (2018).

2. Marsman, R. F., Tan, H. L. & Bezzina, C. R. Genetics of sudden cardiac death caused by ventricular arrhythmias. *Nat. Rev. Cardiol.* **11,** 96–111 (2014).

3. Gorgels, A. P. M., Gijsbers, C., de Vreede-Swagemakers, J., Lousberg, A. & Wellens, H. J. J. Out-of-hospital cardiac arrest--the relevance of heart failure. The Maastricht Circulatory Arrest Registry. *Eur. Heart J.* **24,** 1204–9 (2003).

4. Goldberger, J. J. *et al.* American Heart Association/American College of Cardiology Foundation/Heart Rhythm Society Scientific Statement on Noninvasive Risk Stratification Techniques for Identifying Patients at Risk for Sudden Cardiac Death: A Scientific Statement From the American Heart Association Council on Clinical Cardiology Committee on Electrocardiography and Arrhythmias and Council on Epidemiology and Prevention. *J. Am. Coll. Cardiol.* **52,** 1179–1199 (2008).

5. Bailey, J. J., Berson, A. S., Handelsman, H. & Hodges, M. Utility of current risk stratification tests for predicting major arrhythmic events after myocardial infarction. *J. Am. Coll. Cardiol.* **38,** 1902–1911 (2001).

6. Friedlander, Y. *et al.* Sudden death and myocardial infarction in first degree relatives as predictors of primary cardiac arrest. *Atherosclerosis* **162,** 211–216 (2002).

7. Jouven, X., Desnos, M., Guerot, C. & Ducimetière, P. Predicting sudden death in the population: the Paris Prospective Study I. *Circulation* **99,** 1978–83

(1999).

8. Arking, D. E. *et al.* Identification of a sudden cardiac death susceptibility locus at 2q24.2 through genome-wide association in European ancestry individuals. *PLoS Genet.* **7,** e1002158 (2011).

9. Noseworthy, P. A. *et al.* Common Genetic Variants, QT Interval, and Sudden Cardiac Death in a Finnish Population-Based Study. *Circ. Cardiovasc. Genet.* **4,** 305–311 (2011).

10. Albert, C. M. *et al.* Common variants in cardiac ion channel genes are associated with sudden cardiac death. *Circ. Arrhythm. Electrophysiol.* **3,** 222–9 (2010).

11. Milano, A. *et al.* Sudden Cardiac Arrest and Rare Genetic Variants in the Community. *Circ. Cardiovasc. Genet.* **9,** 147–53 (2016).

12. Ashar, F. N. *et al.* A comprehensive evaluation of the genetic architecture of sudden cardiac arrest. *Eur. Heart J.* **39,** 3961–3969 (2018).

13. Yousuf, O., Chrispin, J., Tomaselli, G. F. & Berger, R. D. Clinical Management and Prevention of Sudden Cardiac Death. *Circ. Res.* **116,** (2015).

14. Albert, C. M. *et al.* Prospective Study of Sudden Cardiac Death Among Women in the United States. *Circulation* **107,** 2096–2101 (2003).

15. Kannel, W. B., Wilson, P. W. F., D'Agostino, R. B. & Cobb, J. Sudden coronary death in women. *Am. Heart J.* **136,** 205–212 (1998).

16. Haukilahti, M. A. E. *et al.* Sudden Cardiac Death in Women. *Circulation* **139,** 1012–1021 (2019).

17. Albert, C. M., McGovern, B. A., Newell, J. B. & Ruskin, J. N. Sex

differences in cardiac arrest survivors. *Circulation* **93,** 1170–6 (1996).

18. Chugh, S. S. *et al.* Women Have a Lower Prevalence of Structural Heart Disease as a Precursor to Sudden Cardiac Arrest: The Ore-SUDS (Oregon Sudden Unexpected Death Study). *J. Am. Coll. Cardiol.* **54,** 2006–2011 (2009).

19. Traylor, M. *et al.* Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE collaboration): a meta-analysis of genome-wide association studies. *Lancet. Neurol.* **11,** 951–62 (2012).

20. Anttila, V. *et al.* Genome-wide meta-analysis identifies new susceptibility loci for migraine. *Nat. Genet.* **45,** 912–917 (2013).

21. Perry, J. R. B. *et al.* Stratifying Type 2 Diabetes Cases by BMI Identifies Genetic Risk Variants in LAMA1 and Enrichment for Risk Variants in Lean Compared to Obese Cases. *PLoS Genet.* **8,** e1002741 (2012).

22. Padyukov, L. *et al.* A genome-wide association study suggests contrasting associations in ACPA-positive versus ACPA-negative rheumatoid arthritis. *Ann. Rheum. Dis.* **70,** 259–65 (2011).

23. Becker, L. B. *et al.* Racial Differences in the Incidence of Cardiac Arrest and Subsequent Survival. *N. Engl. J. Med.* **329,** 600–606 (1993).

24. Galea, S. *et al.* Explaining Racial Disparities in Incidence of and Survival from Out-of-Hospital Cardiac Arrest. *Am. J. Epidemiol.* **166,** 534–543 (2007).

25. Reinier, K. *et al.* Distinctive Clinical Profile of Blacks Versus Whites Presenting With Sudden Cardiac Arrest. *Circulation* **132,** 380–387 (2015).

26. Hertz, R. P., Unger, A. N., Cornell, J. A. & Saunders, E. Racial Disparities in Hypertension Prevalence, Awareness, and Management. *Arch. Intern. Med.*

165, 2098 (2005).

27. Drazner, M. H. *et al.* Left Ventricular Hypertrophy Is More Prevalent in Blacks Than Whites in the General Population. *Hypertension* **46,** 124–129 (2005).

28. Fender, E. A., Henrikson, C. A. & Tereshchenko, L. Racial differences in sudden cardiac death. *J. Electrocardiol.* **47,** 815–8 (2014).

29. Chan, P. S. *et al.* Racial Differences in Survival After In-Hospital Cardiac Arrest. *JAMA* **302,** 1195 (2009).

30. Mezu, U., Ch, I., Halder, I., London, B. & Saba, S. Women and minorities are less likely to receive an implantable cardioverter defibrillator for primary prevention of sudden cardiac death. *Europace* **14,** 341–344 (2012).

31. Zhao, D. *et al.* Racial Differences in Sudden Cardiac Death. *Circulation* **139,** 1688–1697 (2019).

32. Smith, J. G. *et al.* Impact of Ancestry and Common Genetic Variants on QT Interval in African Americans. *Circ. Cardiovasc. Genet.* **5,** 647–655 (2012).

33. Arking, D. E. *et al.* Genetic association study of QT interval highlights role for calcium signaling pathways in myocardial repolarization. *Nat. Genet.* **46,** 826–836 (2014).

34. Junttila, M. J. *et al.* Type 2 diabetes and coronary artery disease: Preserved ejection fraction and sudden cardiac death. *Hear. Rhythm* **15,** 1450–1456 (2018).

35. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinforma. Appl. NOTE* **26,** 2190–2191 (2010).

36. Han, B. & Eskin, E. ARTICLE Random-Effects Model Aimed at

Discovering Associations in Meta-Analysis of Genome-wide Association Studies. (2011). doi:10.1016/j.ajhg.2011.04.014

37. Chanda, P., Huang, H., Arking, D. E. & Bader, J. S. Fast Association Tests for Genes with FAST. *PLoS One* **8,** e68585 (2013).

38. Ashar, F. N. *et al.* A comprehensive evaluation of the genetic architecture of sudden cardiac arrest. *Eur. Heart J.* (2018). doi:10.1093/eurheartj/ehy474

39. Mozaffarian, D. *et al.* Heart Disease and Stroke Statistics—2015 Update. *Circulation* **131,** (2015).

40. Blom, M. T. *et al.* Improved survival after out-of-hospital cardiac arrest and use of automated external defibrillators. *Circulation* **130,** 1868–1875 (2014).

41. Jansen, H., Samani, N. J. & Schunkert, H. Mendelian randomization studies in coronary artery disease. *Eur. Heart J.* **35,** 1917–1924 (2014).

42. Smith, G. D. & Ebrahim, S. Mendelian randomization: prospects, potentials, and limitations. *Int. J. Epidemiol.* **33,** 30–42 (2004).

43. Voight, B. F. *et al.* Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. *Lancet* **380,** 572–580 (2012).

44. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **81,** 559–575 (2007).

45. Purcell, S. & Chang, C. *PLINK 1.9*.

46. Olena, A., Ie> Maintainer, Y. <yavorsko@tcd & Burgess, S. Package 'MendelianRandomization'. (2018).

47. Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian Randomization Analysis With Multiple Genetic Variants Using Summarized

Data. *Genet. Epidemiol.* **37,** 658–665 (2013).

48. Zhu, Z. *et al.* Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat. Commun.* **9,** 224 (2018).

49. Albert, C. M., McGovern, B. A., Newell, J. B. & Ruskin, J. N. Sex Differences in Cardiac Arrest Survivors. *Circulation* **93,** 1170–1176 (1996).

50. Straus, S. M. J. M. *et al.* Prolonged QTc Interval and Risk of Sudden Cardiac Death in a Population of Older Adults. *J. Am. Coll. Cardiol.* **47,** 362–367 (2006).

51. Nelson, C. P. *et al.* Genetically Determined Height and Coronary Artery Disease. *N. Engl. J. Med.* **372,** 1608–1618 (2015).

52. Cupples, L. A., Gagnon, D. R. & Kannel, W. B. Long- and short-term risk of sudden coronary death. *Circulation* **85,** I11-8 (1992).

53. Jouven, X., Zureik, M., Desnos, M., Guérot, C. & Ducimetière, P. Resting heart rate as a predictive risk factor for sudden death in middle-aged men. *Cardiovasc. Res.* **50,** 373–8 (2001).

54. Laukkanen, J. A. *et al.* T-wave inversion, QRS duration, and QRS/T angle as electrocardiographic predictors of the risk for sudden cardiac death. *Am. J. Cardiol.* **113,** 1178–1183 (2014).

55. Adabag, S. *et al.* Obesity related risk of sudden cardiac death in the atherosclerosis risk in communities study. *Heart* **101,** 215–221 (2015).

56. Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518,** 197–206 (2015).

57. Shungin, D. *et al.* New genetic loci link adipose and insulin biology to

body fat distribution. *Nature* **518,** 187–196 (2015).

58. Atrial fibrillation and the risk of sudden cardiac death: The atherosclerosis risk in communities study and cardiovascular health study. *JAMA Intern. Med.* **173,** 29–35 (2013).

59. Jouven, X. *et al.* Diabetes, glucose level, and risk of sudden cardiac death. *Eur. Heart J.* **26,** 2142–2147 (2005).

60. Brion, M.-J. A., Benyamin, B., Visscher, P. M. & Smith, G. D. Beyond the Single SNP: Emerging Developments in Mendelian Randomization in the 'Omics' Era. *Curr. Epidemiol. Reports* **1,** 228–236 (2014).

61. de Vreede-Swagemakers, J. J. . *et al.* Out-of-Hospital Cardiac Arrest in the 1990s: A Population-Based Study in the Maastricht Area on Incidence, Characteristics and Survival. *J. Am. Coll. Cardiol.* **30,** 1500–1505 (1997).

62. Algra, A., Tijssen, J. G., Roelandt, J. R., Pool, J. & Lubsen, J. QTc prolongation measured by standard 12-lead electrocardiography is an independent risk factor for sudden death due to cardiac arrest. *Circulation* **83,** 1888–94 (1991).

63. Straus, S. M. J. M. *et al.* Prolonged QTc interval and risk of sudden cardiac death in a population of older adults. *J. Am. Coll. Cardiol.* **47,** 362–7 (2006).

64. Pfeufer, A. *et al.* Common Variants in Myocardial Ion Channel Genes Modify the QT Interval in the General Population. *Circ. Res.* **96,** 693–701 (2005).

65. Vicente, J., Johannesen, L., Galeotti, L. & Strauss, D. G. Mechanisms of sex and age differences in ventricular repolarization in humans. *Am. Heart J.* **168,** 749–756.e3 (2014).

66. de Bruyne, M. *et al.* Prolonged QT interval predicts cardiac and all-cause mortality in the elderly The Rotterdam Study. *Eur. Heart J.* **20,** 278–284 (1999).

67. Migdalovich, D. *et al.* Mutation and gender-specific risk in type 2 long QT syndrome: Implications for risk stratification for life-threatening cardiac events in patients with long QT syndrome. *Hear. Rhythm* **8,** 1537–1543 (2011).

68. Hong, Y. *et al.* Familial aggregation of QT-interval variability in a general population: results from the NHLBI Family Heart Study. *Clin. Genet.* **59,** 171–177 (2001).

69. Newton-Cheh, C. *et al.* QT interval is a heritable quantitative trait with evidence of linkage to chromosome 3 in a genome-wide linkage analysis: The Framingham Heart Study. *Hear. Rhythm* **2,** 277–284 (2005).

70. Yang, J. *et al.* Genome partitioning of genetic variation for complex traits using common SNPs. *Nat. Genet.* **43,** 519–525 (2011).

71. Pfeufer, A. *et al.* Common variants at ten loci modulate the QT interval duration in the QTSCD Study. *Nat. Genet.* **41,** 407–414 (2009).

72. Eijgelsheim, M. *et al.* Genetic variation in NOS1AP is associated with sudden cardiac death: evidence from the Rotterdam Study. *Hum. Mol. Genet.* **18,** 4213–8 (2009).

73. Arking, D. E. *et al.* A common genetic variant in the NOS1 regulator NOS1AP modulates cardiac repolarization. *Nat. Genet.* **38,** 644–651 (2006).

74. Tian, C., Gregersen, P. K. & Seldin, M. F. Accounting for ancestry: population substructure and genome-wide association studies. *Hum. Mol. Genet.* **17,** R143–R150 (2008).

75. Swerdlow, D. I. *et al.* Selecting instruments for Mendelian randomization in the wake of genome-wide association studies. *Int. J. Epidemiol.* **45,** 1600–1616 (2016).

76. Burgess, S. & Thompson, S. G. Use of allele scores as instrumental variables for Mendelian randomization. *Int. J. Epidemiol.* **42,** 1134–1144 (2013).

77. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48,** 1284–1287 (2016).

78. Algra, A., Tijssen, J. G., Roelandt, J. R., Pool, J. & Lubsen, J. QTc prolongation measured by standard 12-lead electrocardiography is an independent risk factor for sudden death due to cardiac arrest. *Circulation* **83,** 1888–94 (1991).

79. Zhang, Y. *et al.* Sex-Steroid Hormones and Electrocardiographic QT-Interval Duration: Findings From the Third National Health and Nutrition Examination Survey and the Multi-Ethnic Study of Atherosclerosis. *Am. J. Epidemiol.* **174,** 403–411 (2011).

80. NAKAGAWA, M. *et al.* Influence of Menstrual Cycle on QT Interval Dynamics. *Pacing Clin. Electrophysiol.* **29,** 607–613 (2006).

OXFORD UNIVERSITY PRESS LICENSE TERMS AND CONDITIONS

Aug 15, 2019

This Agreement between Rebecca Mitchell ("You") and Oxford University Press ("Oxford University Press") consists of your license details and the terms and conditions provided by Oxford University Press and Copyright Clearance Center.

License Number	4638280818523
License date	Jul 29, 2019
Licensed content publisher	Oxford University Press
Licensed content publication	European Heart Journal
Licensed content title	A comprehensive evaluation of the genetic architecture of sudden cardiac arrest
Licensed content author	Ashar, Foram N; Mitchell, Rebecca N
Licensed content date	Aug 28, 2018
Type of Use	Thesis/Dissertation
Institution name	
Title of your work	Evaluation of the genetic architecture of sudden cardiac arrest
Publisher of your work	Johns Hopkins University
Expected publication date	Aug 2019
Permissions cost	1.1 USD
Value added tax	0.00 USD
Total	0.00 USD
Title	Evaluation of the genetic architecture of sudden cardiac arrest
Institution name	Johns Hopkins University
Expected presentation	Aug 2019

date	
Portions	I will include the introduction, methods, results and discussion about the mendelian randomization analyses.
Requestor Location	Rebecca Mitchell 733 N Broadway BALTIMORE, MD 21205 United States Attn: Rebecca Mitchell
Publisher Tax ID	GB125506730
Total	0.00 USD

Terms and Conditions

STANDARD TERMS AND CONDITIONS FOR REPRODUCTION OF MATERIAL FROM AN OXFORD UNIVERSITY PRESS JOURNAL

1. Use of the material is restricted to the type of use specified in your order details.
2. This permission covers the use of the material in the English language in the following territory: world. If you have requested additional permission to translate this material, the terms and conditions of this reuse will be set out in clause 12.
3. This permission is limited to the particular use authorized in (1) above and does not allow you to sanction its use elsewhere in any other format other than specified above, nor does it apply to quotations, images, artistic works etc that have been reproduced from other sources which may be part of the material to be used.
4. No alteration, omission or addition is made to the material without our written consent. Permission must be re-cleared with Oxford University Press if/when you decide to reprint.
5. The following credit line appears wherever the material is used: author, title, journal, year, volume, issue number, pagination, by permission of Oxford University Press or the sponsoring society if the journal is a society journal. Where a journal is being published on behalf of a learned society, the details of that society must be included in the credit line.
6. For the reproduction of a full article from an Oxford University Press journal for whatever purpose, the corresponding author of the material concerned should be informed of the proposed use. Contact details for the corresponding authors of all Oxford University Press journal contact can be found alongside either the abstract or full text of the article concerned, accessible from www.oxfordjournals.org Should there be a problem clearing these rights, please contact journals.permissions@oup.com
7. If the credit line or acknowledgement in our publication indicates that any of the figures, images or photos was reproduced, drawn or modified from an earlier source it will be necessary for you to clear this permission with the original publisher as well. If this permission has not been obtained, please note that this material cannot be included in your publication/photocopies.
8. While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by Oxford University Press or by Copyright Clearance Center (CCC)) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and Oxford University Press reserves the right to take any and all action to protect its copyright in the materials.
9. This license is personal to you and may not be sublicensed, assigned or transferred by you to any other person

without Oxford University Press's written permission.
10. Oxford University Press reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.
11. You hereby indemnify and agree to hold harmless Oxford University Press and CCC, and their respective officers, directors, employs and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.
12. Other Terms and Conditions:
v1.4

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

Curriculum Vitae

Rebecca N. Mitchell
rebecca.mitchell5987@gmail.com I (814)883-3300
328 S Collington Ave
Baltimore, MD 21231

EDUCATION

Johns Hopkins University Expected September 2019
Ph.D., Human Genetics

Johns Hopkins University August 2011
M.S., Biotechnology
Concentration in Biodefense

The Pennsylvania State University May 2009
B.S., Forensic Science
Biochemistry and Molecular Biology Minor

RESEARCH EXPERIENCE

Doctoral Candidate August 2014-Present
Johns Hopkins University; Institute of Genetic Medicine; Laboratory of Dan Arking

- Lead analyst of the largest genome-wide association study for sudden cardiac arrest (SCA) in collaboration of 17 different national and international cohorts as a part of the CHARGE consortium.
- Performed pre-processing, quality control and statistical modeling and analyses of data from different genotyping platforms, including next-generation sequencing, Agena massARRAY, microarray and TaqMan.
- Used various variant databases (i.e. NCBI,OMIM, UCSC Genome Browser) and software to functionally annotate genetic variants (i.e. Annovar and SnpEff) to prioritize sequence variants for further validation.
- Designed project in predictive analysis of metabolomics data for SCA using the ARIC study.
- Mentored younger graduate students in all technical aspects of the laboratory, including coding, experimental design, methods and data interpretation.
- Attended a weekly department journal club on new research in human genetics and molecular biology.

Biostatistics Consultant January 2019-Present
Personal Genome Diagnostics (PGDx); Baltimore, MD

- Assisted in data wrangling and presentation for 510(k) submission to the FDA

Research Associate I July 2012-August 2014
Leidos Biomedical Research, Inc.; Cancer Genomics Research Laboratory; Department of Cancer Epidemiology and Genetics; National Cancer Institute; National Institutes of Health; Gaithersburg, MD

- Performed exome, targeted, and amplicon sequencing on various sequencers, including Illumina HiSeq 2500, MiSeq and Ion Torrent Personal Genome Machine (PGM)
- Assisted in various research, validation and troubleshooting projects
- Utilized Biomek FX automation robotic system for high throughput processing of samples
- Wrote, reviewed, and edited standard operating procedures (SOPs)
- Trained and mentored new employees in all wet laboratory processes

Forensic Scientist 1 June 2009-April 2012
The Armed Forces DNA Identification Laboratory; Mitochondrial DNA Unit; Office of the Armed Forces Medical Examiner; Rockville, MD

- Performed laboratory methods for mtDNA sequencing using Sanger sequencing on skeletal remains
- Analyzed mtDNA sequence data using the software program Sequencer
- Trained and mentored all new laboratory technicians
- Performed 12s species identification laboratory work and data analysis
- Designed and implemented various validation projects

TECHNICAL EXPERIENCE

Programming: Fluent in R and UNIX languages; Experience with Python, Stata and awk; Proficient in working in Unix environment on computing cluster
Software packages: PLINK; samtools; VCFtools; PicardTools; GATK; Hail; GeneMapper® ID-X Software Version 3.2; GeneMarker HID Version 1.7; Mutation Surveyor Version 3.23; Sequencer
Variant Annotation software/databases: Annovar; SnpEff; NCBI; OMIM; UCSC Genome Browser
Instrumentation: Applied Biosystems 3130xl Genetic Analyzer; Biomek FX 96/Span-8 Robot; Covaris E-Series; Illumina HiSeq 2500; Illumina MiSeq; Roche LightCycler 480 Real Time PCR System; Illumina cBot Cluster Station; Ion Torrent Personal Genome Machine (PGM); Ion Torrent Proton
Next Generation Library Preparation and Sequencing Kits: TruSeq Sample Prep Kit v2, Set A; SeqCap EZ Human Exome, v3.0; NimbleGen SeqCap EZ Hybridization and Wash Kit; TruSeq PE Cluster Kit v3; Ion Ampliseq Library kit; BIOO NEXTflex DNA Sequencing kit; Kapa Biosystems HyperPrep kit (Illumina and Ion)

HONORS AND AWARDS

Maryland Genetics, Epidemiology, and Medicine (MD-GEM) Training Fellow
Fall 2015-Present
Johns Hopkins University

- Cross-training in Johns Hopkins Bloomberg School of Public Health Genetic Epidemiology program
- Additional coursework in biostatistics and genetic epidemiology to further training in statistical and epidemiological methods
- Participated in MD-GEM events including MD-GEM faculty-student lunches and Genetics Research Day
- Attended Models: at the Intersection of Discovery and Data conference in Ann Arbor, MI from August 10-12, 2017

Most Outstanding Poster October 2018
Cohorts for the Heart & Aging Research in Genomic Medicine (CHARGE) Consortium Meeting

TEACHING AND LEADERSHIP EXPERIENCE

Teaching Assistant June 2019
Johns Hopkins University
Center for Computational Genomics
- Practical Genomics Workshop

Teaching Assistant Winter 2016
Johns Hopkins University
School of Medicine
- Bioinformatics for first-year graduate students

Graduate Student Tutor April 2016-Present
- Tutored second year graduate students in epidemiological and biostatical methods, including association and linkage studies for their comprehensive exams

Course Assistant June 2008-May 2009
Pennsylvania State University
Forensic Science Program
- Criminalistics: Biological Evidence Screening

President April 2008-May 2009
Vice President August 2005-April 2008
Pennsylvania State University Forensics Club
- Ran monthly meetings and events
- Organized a forensic science symposium for 100 high school and college students including speakers from the Pennsylvania State Police, New Jersey State Police, the FBI and other law enforcement agencies

SCIENTIFIC PUBLICATIONS

In progress:
- **Mitchell, R. N.**, Ashar, F. N., Jarvelin, M.-R., Froguel, P., Sotoodehnia, N., Brody, J. A., *et al.* The effect of sex and underlying disease on the genetic association of QT interval and sudden cardiac death. *bioRxiv* 2019; 664300. doi:10.1101/664300. **Under review at the Journal of the American Heart Association**.

Published:
- Ashar, F.N.*, **Mitchell, R.N.***, Albert, C.M., Newton-Cheh, C., Brody, J.A., Müller-Nurasyid, M., Moes, A., Meitinger, T., Mak, A., Huikuri, H., et al. (2018). A comprehensive evaluation of the genetic architecture of sudden cardiac arrest. Eur. Heart J.
- Dean, M.,…,**Eggebeen,R.**,…,Nahleh, Z. (2015). Addressing health disparities in Hispanic breast cancer: accurate and inexpensive sequencing of BRCA1 and BRCA2. Gigascience 4, 50.
- Kocak, H., Ballew, B.J., Bisht, K., **Eggebeen, R.,** Hicks, B.D., Suman, S., O'Neil, A., Giri, N., NCI DCEG Cancer Genomics Research Laboratory, N.D.C.G.R., NCI DCEG

Cancer Sequencing Working Group, N.D.C.S.W., et al. (2014). Hoyeraal-Hreidarsson syndrome caused by a germline mutation in the TEL patch of the telomere protein TPP1. Genes Dev. 28, 2090–2102.

PRESENTATIONS

- **Mitchell, R.**, Ashar, F.N., SCD CHARGE Working Group, Sotoodehnia, N., Arking, D.E. A Comprehensive Evaluation of the Genetic Architecture of Sudden Cardiac Arrest. Poster session at American Society of Human Genetics meeting. October 16-20, 2018. San Diego, CA.
- **Mitchell, R.**, Ashar, F.N., SCD CHARGE Working Group, Sotoodehnia, N., Arking, D.E. A Comprehensive Evaluation of the Genetic Architecture of Sudden Cardiac Arrest. Poster session at CHARGE Consortium Investigator meeting. October 11-12, 2018. Baltimore, MD.
- **Mitchell, R.**, Ashar, F.N., SCD CHARGE Working Group, Sotoodehnia, N., Arking, D.E. A Comprehensive Evaluation of the Genetic Architecture of Sudden Cardiac Arrest. Poster session at MD-GEM Genetics Research Day. February 9, 2018. Baltimore, MD.
- **Mitchell, R.**, Ashar, F.N., Junttila, J., Kuikuri, H., Arking, D.E. Investigating the Role of the *SCD1* Locus on Sudden Cardiac Death Risk. Poster session presented at MD-GEM Genetics Research Day, February 17, 2017. Baltimore, MD.