THE ASSOCIATION OF SINGLE NUCLEOTIDE POLYMORPHISMS (SNPs) IN THE NITRIC OXIDE SYNTHASE 1 ADAPTOR PROTEIN GENE (NOS1AP) AND THE QT INTERVAL DURATION ACROSS RACE/ETHNICITY IN ADULTS WITHOUT CLINICAL CARDIOVASCULAR DISEASE

BY

SIDHARTH A. SHAH

A Thesis Submitted to the Graduate Faculty of WAKE FOREST UNIVERSITY GRADUATE SCHOOL OF ARTS AND SCIENCES in Partial Fulfillment of the Requirements for the Degree of MASTER OF SCIENCE in the Health Sciences Research Program of the Wake Forest University School of Medicine March 2010 Winston- Salem, North Carolina

Approved by:

David M. Herrington, M.D., M.H.S., Advisor

Examining Committee:

Donald W. Bowden, Ph.D., Chairman

Doug Case, Ph.D.

Timothy D. Howard, Ph.D
ACKNOWLEDGEMENTS

I would like to sincerely thank the following people for their guidance, support
and mentorship, without whom this thesis submission would not have been possible.

To Dr. Herrington, thank you for your unremitting resolve to making me better.
Your mentorship and guidance have been unparalleled and I will always be grateful for
all that you have done for me.

To Allie Richardson and Karen Blinson, the den mothers of the fellows, thank you
for keeping me in line with all my abstract and poster deadlines, appointment times,
and professional obligations. You were instrumental in getting me settled into
fellowship and life in Winston. To Georgia Saylor and Kurt Lohman, thank you for all
your statistical wizardry. You have spent countless hours helping me decode our
statistical software programs and datasets. To Drs. Fitzgerald and Applegate, I owe my
outstanding clinical experience at Wake to you both. It has been truly amazing and I look
forward to continuing my training with our Electrophysiology Department. Dr. Case,
thank you for your time and efforts in making this thesis a reality. To Drs. Bowden and
Howard, thank you so much for taking me under your wing to help me better
understand genetics. Your mentorship was not only instrumental in helping me develop
the tools to complete this research project but also forged a friendship through the
process.

I would also like to take an opportunity to sincerely thank my parents, Anil and
Rashmi, as well as my younger brothers, Sumit and Amar for their unconditional love.
They have taught me life’s most important lessons and have always been my biggest
supporters. Finally, the submission of this thesis would not have been possible without the unrelenting love and support of my wonderful wife, Dianna, and my beautiful daughter, Sonali. Words cannot express my eternal love for you and how much you have done to help me through life. Simply stated, you are my rock.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF ILLUSTRATIONS</td>
<td>vi</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>ix</td>
</tr>
<tr>
<td>Chapter</td>
<td></td>
</tr>
<tr>
<td>1. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>2. ASSOCIATIONS BETWEEN GENETIC VARIATIONS IN NOS1AP AND QT INTERVAL</td>
<td>18</td>
</tr>
<tr>
<td>DURATION IN FOUR RACIAL/ETHNIC GROUPS IN THE MULTI-ETHNIC STUDY OF</td>
<td></td>
</tr>
<tr>
<td>ATHEROSCLEROSIS (MESA)</td>
<td></td>
</tr>
<tr>
<td>3. USE OF PRINCIPAL COMPONENT ANALYSIS (PCA) ENHANCES THE ABILITY TO</td>
<td>48</td>
</tr>
<tr>
<td>STUDY ASSOCIATIONS BETWEEN SINGLE NUCLEOTIDE POLYMORPHISMS (SNPS)</td>
<td></td>
</tr>
<tr>
<td>IN NOS1AP AND QT INTERVAL IN THE MULTI-ETHNIC STUDY OF</td>
<td></td>
</tr>
<tr>
<td>ATHEROSCLEROSIS (MESA)</td>
<td></td>
</tr>
<tr>
<td>4. CURRICULUM VITAE</td>
<td>63</td>
</tr>
</tbody>
</table>
LIST OF ABBREVIATIONS

AIMS= Ancestry Informative Markers

AFA= African-American

BP= Blood Pressure

CAU= Caucasians

CVD= Cardiovascular Disease

CHN= Chinese

ECG= Electrocardiogram

GWAS= Genome wide association study

HIS= Hispanics

LD= Linkage disequilibrium

MESA= Multi-Ethnic Study of Atherosclerosis

NOS1AP= Nitric oxide synthase 1 adaptor protein

SCD= Sudden Cardiac Death

SNP= Single Nucleotide Polymorphisms

VF= Ventricular fibrillation

VT= Ventricular tachycardia
LIST OF ILLUSTRATIONS

TABLES PAGE

CHAPTER 2

I. MESA Participant Characteristics According to Race/ethnicity.........................28

II. Genotypic Association of SNPs in NOS1AP with QT Interval

Duration for African- American (AFA), Chinese (CHN), Caucasian (CAU),

and Hispanic (HIS) Groups in MESA.................................................................36

CHAPTER 3

I. Associations Between NOS1AP SNPs and QT Interval across Self Reported

Race/ethnicity in MESA Unadjusted and Adjusted for Population Stratification

(PS) With PCA Using AIMS: Parameter (Beta) Estimates and Associated P

Values..................................................................................................................59
LIST OF ILLUSTRATIONS

FIGURES

CHAPTER 1

1. The Incidence of SCD in Specific Populations and the Annual Numbers of Sudden Deaths in Those Populations.................................................................2

2. Normal Electrocardiogram (ECG) Tracing. ..........................................................3

3. Action Potential of the Ventricular Myocyte (top panel) and the Surface Electrocardiogram (ECG; bottom panel).................................................................5

4. Possible Mechanism Behind Arrhythmogenesis Related to the QT Interval.................................................................6

5. Ventricular Tachycardia (VT; Panel A) and Ventricular Fibrillation (VF; Panel B)......................................................................................................................7

CHAPTER 2

1. Race-Specific Negative Log_{10} P-value for Each SNP Across NOS1AP...............31

CHAPTER 3

1. Effects of Population Structure at a SNP Locus..................................................49

2. Classical Confounding (Panel A) and Population Stratification

   (PS; Panel B). ........................................................................................................50

3. Prevalence of T2DM by Age and the Presence of the Haplotype Gm^{2,5,13,14}

   Among Native Americans of Pima- Papago Descent. ..................................51
4. Age-adjusted Prevalence of Type II Diabetes (T2DM) by the Presence of the Haplotype Gm2;5,13,14, According to Indian Heritage, Among Native Americans Residents of Pima- Papago Descent

5. Ancestry Proportion Estimates Amongst Self-Reported African-American (AFA), European (CAU), Chinese (CHN), and Hispanic (HIS) Groups

6. Two Dimensional (2D) Graphical Representation of the First Two Principal Components (PCs) Derived from Ancestry Informative Markers (AIMS)
ABSTRACT

Sidharth Anil Shah

THE ASSOCIATION OF SINGLE NUCLEOTIDE POLYMORPHISMS (SNPs) IN THE NITRIC OXIDE SYNTHASE 1 ADAPTOR PROTEIN GENE (NOS1AP) AND THE QT INTERVAL DURATION ACROSS RACE/ETHNICITY IN ADULTS WITHOUT CLINICAL CARDIOVASCULAR DISEASE

Thesis under the direction of David M. Herrington, M.D., M.H.S., Professor of Medicine

Introduction: The QT interval is an independent risk factor for sudden cardiac death (SCD). A genome wide association study (GWAS) identified NOS1AP variants associated with QT, which has been replicated in predominantly Caucasian populations. The MESA study provides an opportunity to study single nucleotide polymorphisms (SNPs) in NOS1AP for association with QT interval in an ethnically diverse cohort.

Methods: Twenty-eight tagging SNPs, starting in the first intron and spanning the entire gene, were genotyped in 2847 men and women, age 45-84 years, without known CVD in MESA, including approximately equal numbers of Caucasians (CAU), African-Americans (AFA), Hispanics (HIS) and Chinese (CHN) US participants. QT interval was determined using a standard 12-lead ECG recording. Associations between QT and genotypes were evaluated using multiple linear regression, adjusted for heart rate (HR), age, gender, and field center stratified by racial/ethnic group, using an additive model of inheritance. Ancestry informative markers (AIMs) were genotyped in MESA participants.
and principal components (PCs) using AIMs were used as additional covariates in accounting for population stratification (PS).

**Results:** Associations between genotype and QT interval were stronger in CAU (11 SNPs, \( p=7.20 \times 10^{-7} - 4.3 \times 10^{-2} \)) than in AFA (2 SNPs, \( p=0.030 - 0.043 \)), CHN (5 SNPs, \( p=0.021 - 0.047 \)) or HIS (3 SNPs, \( p=0.03 - 0.003 \)). In CAU, each additional copy of the minor allele in rs1932933 was associated with a 4.5 msec increase in QT (\( p=7.20 \times 10^{-7} \)). The preponderance of significant CAU and HIS SNPs were at the 5’ end of NOS1AP (some within similar LD blocks), while significant associations in CHN were only at the 3’ end.

**Conclusions:** NOS1AP variants are associated with QT interval in AFA, CAU, HIS, and CHN. The location of significant SNPs varies across racial/ethnic groups. We have identified possible novel variants at the 3’ end of NOS1AP associated with QT in CHN only. Further genotyping within these regions should be pursued to determine functional genetic variants affecting QT interval and the potential risk for SCD.
CHAPTER 1

INTRODUCTION

I. Sudden Cardiac Death (SCD)

The term sudden cardiac death (SCD) is often used to describe abrupt cardiovascular collapse secondary to a life-threatening arrhythmia in an individual with or without known cardiac disease. SCD is a major public health problem, as it is responsible for approximately 15% of all deaths in the US per year with approximately 60% of SCDs occurring outside the hospital setting [1]. Unfortunately, despite advances in understanding the pathophysiology of cardiac diseases and the resultant improvement in resuscitation science, the incidence of SCD has largely remained unchanged, accounting for approximately 300,000 events per year in the US [2].

The predominant risk factor for SCD is the presence of coronary artery disease (CAD) in the setting of a compromised cardiac function. However, most deaths occur in the larger, lower-risk subgroups (Figure 1). It follows that SCD has been shown to be a very heterogeneous condition, and other precipitating factors include coronary ischemia, structural defects, myocardial scarring, and genetic mutations. Several large, randomized clinical trials, including SCD-HeFT, MADIT II, COMMIT, EPHESUS, and HOPE have shown mortality benefits of both mechanical and medical intervention in patients with either stable coronary artery disease (CAD) or acute coronary syndrome (ACS; [3-
However, identifying risk factors that predispose individuals in the general population for SCD has emerged as a topic for further scientific investigation.

![Figure 1. The Incidence of SCD in Specific Populations and the Annual Numbers of Sudden Deaths in Those Populations. Adapted from Myerburg et al. (1998) and Huikuri et al. (2001) [8, 9].](image)

II. The Electrocardiogram (ECG) and QT Interval Duration

The 12-lead surface electrocardiogram (ECG) is reflective of the behavior of transmembrane action potentials of the cardiac myocardium. The character of these membrane potentials varies with the four different cell types within the heart (ventricular myocytes, atrial myocytes, sinoatrial (SA)/atrioventricular (AV) nodal cells and Purkinje cells). Normal cardiac rhythm is dependent upon intact mechanisms of generation of cellular action potentials. Abnormally generated cellular action potentials can result in rhythmic aberrations, potentiating the genesis of SCD.
ECG waveforms (Figure 2) are the graphic representation of four electrophysiologic events: 1) impulse formation in the primary pacemaker of the heart (the SA node); 2) transmission of the impulse through specialized myocardial conduction fibers; 3) activation (depolarization) of the myocardium; and 4) repolarization (recovery) of the myocardium [10].

Figure 2. Normal Electrocardiogram (ECG) Tracing. P wave = depolarization of atria. QRS interval = depolarization of ventricles, consisting of the Q, R and S waves. T wave = ventricular repolarization. ST segment = time interval between ventricular depolarization and repolarization. PR interval = time interval between onset of atrial depolarization and onset of ventricular depolarization. QT interval = time interval between onset of ventricular depolarization and end of ventricular repolarization. R-R interval = time interval between 2 QRS complexes. Adapted from www.gene-test.org/[11]

The QT interval duration, as measured in milliseconds from the ECG, is reflective of the total duration of ventricular depolarization and repolarization (Figure 3; [11, 12]). The QT interval duration has been shown to be an independent risk factor for SCD in numerous community-based studies [13-15]. However, a more complete understanding
of the QT interval hinges upon a brief review of the electrophysiology behind the action potential of ventricular myocytes.

The resting membrane potential (RMP) represents the voltage potential between the inside and outside of a cell. The RMP of cardiac myocytes is predominantly determined by the gradient of the potassium ions (K$^+$) across the cell membrane. The intracellular (150 meq/L): extracellular (5 meq/L) concentration of K$^+$ results in a 30:1 concentration gradient. An opposite gradient exists for sodium (Na$^+$) ions, resulting in a high extracellular Na$^+$ concentration relative to intracellular Na$^+$ concentration. These concentration gradients are maintained by an active Na$^+$ transport channel. The resulting RMP in most cardiac cells is -80 to -90 millivolts (mV) [10].

Depolarization, or activation, of the ventricular myocyte, is triggered by an abrupt change in membrane permeability to Na$^+$ (phase 0, Figure 3). Na$^+$ and to a lesser extent, calcium ions (Ca$^{2+}$), enter the cell through their respective channels, resulting in a sharp rise in the membrane potential to approximately +20 mV. This phase of depolarization, termed “phase 0”, reflects the Na$^+$-dependent fast inward current typical of ventricular myocytes (as well as Purkinje cells). However, in certain pathologic conditions, such as ischemia, cells whose fast inward current of Na$^+$ is inhibited are depolarized by slow inward currents of Ca$^{2+}$ [10].

Following depolarization, there is a gradual return of the voltage potential to the resting potential (phases 1, 2, 3, and 4; Figure 3). This recovery, or repolarization, is divided into three phases: phase 1 is the initial rapid return of the intracellular potential to 0 mV due to the closing of the Na$^+$ channels; the plateau characteristic of phase 2 is a
result of the slow influx of Ca\(^{2+}\) into the cell; and in **phase 3**, the efflux of K\(^+\) results in the slow return of intracellular potential towards the RMP. At the end of the **phase 3**, the Na\(^+\) / K\(^+\) active transport system pumps Na\(^+\) from the cell into the extracellular matrix in exchange of K\(^+\), thereby reestablishing RMP (Figure 3).

![Figure 3](image)

**Figure 3. Action Potential of the Ventricular Myocyte (top panel) and the Surface Electrocardiogram (ECG; bottom panel).** The QT interval is a surface recording of the myocardial action membrane potential. The resting membrane potential (RMP) of ventricular myocytes is approximately -80 to -90 millivolts (mV). **Phase 0** = depolarization; **Phases 1,2,3,4** = repolarization, and **Phase 4** = diastolic phase. **Phase 0**: rapid depolarization due to Na\(^+\) (and Ca\(^{2+}\)) influx into the cell. **Phase 1**: initial phase of repolarization due to the closing of Na\(^+\) channels; **Phase 2**: plateau phase of repolarization due to slow Ca\(^{2+}\) influx. **Phase 3**: final phase of repolarization in which the efflux of K\(^+\) results in the return of the intracellular potential to RMP. At the end of **phase 3**, the Na\(^+\) / K\(^+\) active transport extrudes Na\(^+\) from within the cell and pumps K\(^+\) into the cell. Adapted from Libby et al. [12] and Goldschlager et al. [10].
The clinical manifestation of a lengthened QT interval duration represents prolonging of the action potential duration within at least some of the ventricular myocytes (Figure 4 panel B) [16]. It is proposed that prolongation of the action potential increases Ca$^{2+}$ influx into myocytes during the cardiac cycle, causing excessive Ca$^{2+}$ accumulation in the sarcoplasmic reticulum (SR) and spontaneous SR Ca$^{2+}$ release [12]. The resulting elevation of intracellular free Ca$^{2+}$ is thought to depolarize cardiac myocyte membranes, subsequently evoking abnormal cardiac depolarizations that may disrupt the normal cardiac action potential. These aberrations in cardiomyocyte depolarization are termed “early afterdepolarizations (EADs; Figure 4 panel C)”. EADs can increase repolarization time and can generate a propagated response, or “upstroke”, thereby eliciting an extra ventricular beat (Figure 4 panel D) and potentially launching into a ventricular tachyarrhythmia [10, 12, 16].

**Figure 4. Possible Mechanism Behind Arrhythmogenesis Related to the QT Interval.** In certain conditions, a normal action potential of the ventricular myocytes (panel A) can lead to a pathologic prolonging of the QT (panel B) and a deformity in the trajectory of repolarization termed an early after-depolarization (EAD). EADs can generate a triggered beat, potentially launching into a tachyarrhythmia. Adapted from Roden et al. [16].
The two malignant tachyarrhythmias that are most often associated with QT prolongation and lead to SCD are ventricular tachycardia (VT) and ventricular fibrillation (VF; Figure 5; [17]). Prolongation of the absolute QT interval beyond 500 msec is commonly regarded as conferring an increased risk for arrhythmogenesis [16, 17]. Various physiological conditions can alter intracellular Ca\(^{2+}\), thereby causing a pathologic prolongation of the QT interval. Such states include ischemia during acute myocardial infarction (AMI) and pharmacotherapy. However, strategies in identifying a genetic predisposition for SCD have provided a substrate for new investigations into phenotyping this complex heterogenous condition. As such, the ultimate goal for these studies has been to facilitate preventive approaches to this significant public health concern.

**Figure 5. Ventricular tachycardia (VT; Panel A) and Ventricular fibrillation (VF; Panel B).** VT and VF are two arrhythmias that are associated with QT prolongation and can lead to SCD. Note the loss of waveform morphology that was easily distinguishable in the normal tracing in Figure 2. Adapted from www.frca.co.uk/images_main/[11].

### III. Landmark Genetic QT Study: Arking et al. 2006

Previous studies by Busjahn (1999), Carter (2000), and Newton-Cheh (2005) have established the QT interval as a trait influenced by genetics, with a heritability of approximately 30% [18-20]. However, Arking et al. (2006) used a genome wide association study (GWAS) to identify previously unknown genetic variants that could
potentially influence the QT interval [21]. Measurement of the QT directly, as opposed to the SCD phenotype, was chosen for several reasons: 1) as the QT interval is an inheritable trait implicated in the pathogenesis of SCD, identification of genetic variants responsible for QT variation could provide insight into genetic influences on SCD; 2) unlike SCD, the QT interval is an easily obtainable marker that can be measured in individuals with and without CAD; and 3) large community-based studies of SCD are unavailable [21].

The study by Arking et al. possessed four key features in its design. The first was that study participants came from a population-based survey of volunteers aged 25-75 years from the German cohort, KORA S4 survey. Second, GWAS was performed on only a subset of 200 women from this survey, thereby reducing any variability that would be due to gender. Finally, a three-staged study was performed by examining samples with phenotypic means of decreasing deviations from the population average but with increasing sample size. This was done to minimize the number of false positive associations yet maximize power and efficiency. Finally, replication studies were performed in two population-based samples of 2,646 subjects from Germany and 1,805 subjects from the Framingham Heart Study [21].

The GWAS performed by Arking et al. identified noncoding variants in the 5’ end of the nitric oxide synthase 1 adaptor protein gene (NOS1AP) associated with QT interval duration. In this study, approximately 60% of the European-descended participants studied carried at least one minor allele of the NOS1AP genetic variant. While a major quantitative trait locus (QTL) was not found, the involvement of the NOS1AP single
nucleotide polymorphism (SNP) was found to explain 1.5% of the variance in QT in the combined sample of approximately 6,600 German participants and approximately 0.6% in the Framingham Heart Sample [21].

The GWAS performed by Arking and colleagues was a landmark study in which NOS1AP was first identified as a modulator of cardiac repolarization in a predominantly all-Caucasian cohort. Since that study, these associations have been replicated in several predominantly all-Caucasian cohorts [22-25].

IV. Replication Studies--Additional Insights and Exposure of Limitations

Since the GWAS performed by Arking et al. in 2006, several studies have attempted to evaluate and replicate the association between NOS1AP and QT interval variation. While these studies have provided additional insight into the genetic influences on cardiac repolarization, there are important limitations in their design that have served as challenges in the attempt to further characterize the role of genetics on QT interval variation in the general population.

In 2007, Post et al. performed a GWAS to discover an association of the QT interval with four SNPs in NOS1AP in the Old Order Amish [22]. The purpose of this study was to replicate the findings of Arking et al. in a genetically isolated population. The Old Order Amish of Lancaster County, PA are a closed, rural, genetically isolated community whose ancestry can be traced back to European immigrants from the 1700s. Two of the four NOS1AP SNPs selected were significantly associated with QT variation in the 763 subjects studied [22]. The findings of Post et al. validated the results of Arking et al. and underscored the importance of GWAS by testing these associations in a more genetically
homogenous population. While genetically homogenous populations reduce the potential of spurious associations, an unfortunate consequence is that the findings are largely limited to such populations. In a society where population admixture and substructure can potentially confound associations between complex traits and genetic polymorphisms, such limitations can obscure our full understanding of the functional components of these complex phenotypes.

The studies of Arking et al. and Post et al. further substantiated the importance of GWAS in identifying genomic regions, such as \textit{NOS1AP}, that previously would not have been implicated to influence traits, including the QT interval. Subsequent studies by Aarnoudse et al. (2007) and Tobin et al. (2008) have investigated the associations between \textit{NOS1AP} variants and QT interval variation in predominantly all-Caucasian, population-based cohorts [23, 24]. However, Kao et al. (2009) attempted to validate the association between \textit{NOS1AP} variants and the QT interval using 19 tagging SNPs genotyped in 14,737 white and 4558 black adults from the Atherosclerosis Risk in Communities (ARIC) Study and the Cardiovascular Health Study (CHS [26]). Three of the 19 tagging SNPs in \textit{NOS1AP} were significantly associated with the adjusted QT interval in whites (p<0.0001); however, no significant associations between \textit{NOS1AP} variants and the QT interval were observed in blacks. The discordance of findings between the blacks and whites may reflect reduced statistical power, as approximately 75% of the study participants were white. Furthermore, because a majority of the study participants were white, SNPs were selected to tag the linkage disequilibrium (LD) block that contained the most significant SNPs from fine mapping in a previous study performed in whites.
Thus inappropriate tagging SNPs could have further reduced statistical power to detect a SNP-trait association in blacks.

The above studies have identified and validated an association of NOS1AP variants and QT interval variation in population-based, predominantly Caucasian cohorts. The identification of a panel of susceptible alleles in the general population is tantamount in assessing SCD risk and consequently developing prevention strategies and therapeutic targets for cardiac repolarization. Hence, further testing in large, ethnically heterogeneous populations will be important in determining the role of NOS1AP variants on cardiac repolarization, as reflected by the QT interval.

V. **NOS1AP: An Ion Channel Modulator Implicated in Cardiac Repolarization**

Investigations of Mendelian disorders of the QT interval, including congenital Long QT Syndrome, have identified the role of ion channel gene mutations on altering cardiac membrane potentials and resulting in pathologic QT variation and an elevated risk of SCD [27-30]. As an adapter protein, the gene product of NOS1AP in modulating cardiac repolarization suggests an association between NOS1AP and the risk for SCD.

The NOS1AP gene spans over 299kb of DNA on Chromosome 1 (1q23.3) and comprises 10 exons [24]. NOS1AP encodes a cytosolic carboxy-terminal PDZ ligand of neuronal nitric oxide synthase (nNOS) and serves to physically bridge nNOS and its targets as well as modulator proteins [26]. In this sense, NOS1AP may regulate nitric oxide production by competitively inhibiting access to Ca\(^{2+}\) influx within ion channels via this PDZ domain. While ubiquitously expressed, overexpression on NOS1AP could result in shortening of the cardiac action potential, a decrease in L-type Ca\(^{2+}\) current, and a
smaller increase in the delayed rectifier potassium current $I_K$, thereby prolonging the QT interval and potentially increasing the susceptibility for arrhythmogenesis [26, 31]. Others have speculated that NOS1AP modulates the balance between nitric oxide and superoxide production, thereby directly influencing cardiac repolarization and the propensity for SCD [32-34].

VI. The Future: Discovery of Genetic Pathways that Influence SCD Risk

SCD remains a major public health concern. The QT interval is a quantitative trait that can accurately and reliably be obtained and measured from the 12-lead ECG; however, as with many surrogate markers, the association of the QT and to the event of interest is imperfect, as the risk of ventricular arrhythmogenesis is not a linear function of the QT interval [16]. Despite this limitation, there is a growing body of evidence that has defined a relationship between NOS1AP mutations and the QT interval [21-26, 35-37]. Previous studies have established these associations in predominantly Caucasian cohorts; consequently little research has explored these gene-trait associations in non-Caucasian groups. This has subsequently limited the generalizability of these findings. Given that the QT interval is an intermediate trait for SCD with moderate heritability, the identification of novel NOS1AP genetic variants associated with the QT interval in a large, heterogenous population is warranted. Furthermore, linkage disequilibrium (LD) can help narrow region(s) of interest, thereby facilitating the identification of associated SNPs within the NOS1AP gene. The purpose of this thesis is to explore the variance in the NOS1AP gene across different ethnicities and the potential associations between NOS1AP SNPs and the QT interval. We hope that our findings may expand our
understanding of cardiac repolarization and facilitate the development of new prevention strategies and therapeutic targets to prevent SCD.
References


11. [cited; Available from: www.gene-test.org/].


CHAPTER 2

Associations Between Genetic Variations in NOS1AP and QT Interval Duration in Four Racial/Ethnic Groups in the Multi-Ethnic Study of Atherosclerosis (MESA)

Shah SA¹, Herrington DM¹,³, Howard TD², Divers J², Arnett DK⁴, Burke GL³, Kao WH⁵, Guo X⁶, Siscovick DS⁷, Chakravarti A⁸, Lima JA⁹, Psaty BM⁷, Tomaselli GF⁹, Rich SS¹⁰, Bowden DW², and Post W⁵,⁹

¹Section on Cardiology, ²Center for Human Genomics, ³Public Health Sciences, Wake Forest University School of Medicine; Winston-Salem, NC
⁴Department of Epidemiology, University of Alabama; Birmingham, AL
⁵Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health; Baltimore, MD
⁶Mathematical Genetics, Cedars-Sinai Medical Center; Los Angeles, CA
⁷Cardiovascular Health Research Unit, University of Washington; Seattle, WA
⁸McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, Maryland
⁹Division of Cardiology, John Hopkins University; Baltimore, MD
¹⁰Center for Public Health Genomics, University of Virginia; Charlottesville, VA
ACKNOWLEDGMENTS:

This research was supported in part by NIH R01HL076438 and by NIH T32 Grant: HL076132 as well as N01-HC-95159 through N01-HC-95169 (MESA) and RO1 HL071205 (MESA Family) from the National Heart, Lung, and Blood Institute. The authors thank the other investigators, the staff, and the participants of the MESA study for their valuable contributions. A full list of participating MESA investigators and institutions can be found at http://www.mesa-nhlbi.org.

Introduction: The electrocardiographic QT interval is an independent risk factor for sudden cardiac death (SCD). A genome wide association study (GWAS) identified \textit{NOS1AP} variants associated with QT, which has been replicated in predominantly Caucasian populations. The MESA study provides an opportunity to study single nucleotide polymorphisms (SNPs) in \textit{NOS1AP} for association with QT interval in an ethnically diverse cohort.

Methods: Twenty-eight tagging SNPs, starting in the first intron and spanning the entire gene, were genotyped using the Illumina GoldenGate assay in 2847 men and women, age 45-84 years, without known CVD in MESA, including approximately equal numbers of Caucasians (CAU), African-Americans (AFA), Hispanics (HIS) and Chinese (CHN) US participants. QT interval was determined using a standard 12-lead ECG recording. Associations between QT interval and genotypes were evaluated using multiple linear regression, adjusted for heart rate, age, gender, and field center stratified by ethnic group, using an additive model of inheritance. Ancestry informative markers (AIMs)
were genotyped in MESA participants and principal components (PCs) using AIMs were used as additional covariates in accounting for population stratification (PS).

Results: Associations between genotype and QT interval were stronger in CAU (11 SNPs, p=7.20x10⁻⁷ - 4.3x10⁻²) than in AFA (2 SNPs, p=0.030 - 0.043), CHN (5 SNPs, p=0.021 - 0.047) or HIS (3 SNPs, p= 0.03 - 0.003). In CAU, each additional copy of the minor allele in rs1932933 was associated with a 4.5 msec increase in QT (p= 7.20x10⁻⁷). The preponderance of significant CAU and HIS SNPs were at the 5' end of NOS1AP (some within similar LD blocks), while significant associations in CHN were only at the 3' end.

Conclusions: NOS1AP variants are associated with QT interval in AFA, CAU, HIS, and CHN. However, the location of significant SNPs varies across racial/ethnic groups. We have identified possible novel variants at the 3' end of NOS1AP associated with QT in CHN only. Further genotyping within these regions should be pursued to determine functional genetic variants affecting QT interval and the potential risk for SCD.
Introduction

Sudden cardiac death (SCD) is a major health concern, with approximately 300,000 events per year in the US [1]. The QT interval duration, as measured from the 12-lead electrocardiogram (ECG), has been shown to be an independent risk factor for SCD in numerous community-based studies[2-5]. Through a genome wide association study (GWAS), Arking et al. (2006) identified noncoding variants in the 5’ end of the nitric oxide synthase 1 adaptor protein gene (NOS1AP) associated with QT interval duration in Caucasians [6]. These associations have been replicated in several predominantly all-Caucasian cohorts [7-12]. However, these associations have not been studied extensively in a large, ethnically diverse cohort.

The NOS1AP gene spans over 299kb of DNA on Chromosome 1 (1q23.3) and comprises 10 exons[9]. NOS1AP encodes a cytosolic carboxy-terminal PDZ ligand of neuronal nitric oxide synthase (nNOS) and serves to physically bridge nNOS and its targets as well as modulator proteins[10]. In this sense, NOS1AP may regulate nitric oxide production by competitively inhibiting access to calcium (Ca+) ion influx within ion channels via this PDZ domain. While ubiquitously expressed, overexpression on NOS1AP could result in shortening of the cardiac action potential, a decrease in L-type Ca+ current, and a smaller increase in the delayed rectifier potassium current I\textsubscript{K}, thereby prolonging the QT interval and potentially increasing the susceptibility for arrhythmogenesis [10, 13]. Others have speculated that NOS1AP modulates the balance between nitric oxide and superoxide production, thereby directly influencing cardiac repolarization and the propensity for SCD [14-16].
At this time, associations between \textit{NOS1AP} genetic variants and QT interval duration have been identified and extensively replicated in predominantly Caucasian cohorts and found to be associated with SCD \cite{6-8, 10, 11, 17}. Associations in other ethnicities, including Hispanics and African Americans, have only been identified in smaller cohort populations \cite{17}. This has subsequently limited the applicability of previous findings to Caucasian populations. The Multi-Ethnic Study of Atherosclerosis (MESA) includes participants without clinical cardiovascular disease ages 45 to 84 years old recruited in six US communities from four racial/ethnic groups—Caucasian (CAU), African-American (AFA), Hispanic (HIS), and Asian (of Chinese descent; CHN).

Given the association between QT interval duration and sudden cardiac death (SCD), the identification of novel genetic variants in a large, heterogeneous population may expand our understanding of myocardial conduction and facilitate the development of new clinical therapies to prevent SCD. The purpose of the current study was to evaluate genetic variants in \textit{NOS1AP} in previously unstudied ethnicities that may be associated with the QT interval duration in the MESA cohort.

\textbf{Methods}

\textbf{Sample}: MESA was initiated in July 2000 to investigate the prevalence, correlates, and progression of subclinical cardiovascular disease. Participants had no evidence of clinical cardiovascular disease at enrollment. The MESA cohort consists of 6,814 men and women ages 45 to 84 years old recruited in six US communities: Baltimore, MD; Chicago, IL, Forsyth County, NC; Los Angeles County, CA, Northern Manhattan, NY; and St. Paul, MN. Details of the sampling, recruitment, and data
collection have been reported elsewhere[18]. MESA comprises 47% men with a racial/ethnic distribution of 38% Caucasian, 28% African-American, 22% Hispanic, and 12% Asian (of Chinese descent) individuals. The study participants for this genetic study were 2847 MESA subjects genotyped as part of a candidate gene Illumina panel (approximately 720 participants in each of four racial/ethnic groups): Caucasian (CAU), African-American (AFA), Hispanic-American (HIS), Asian (of Chinese descent; CHN). The Institutional Review Board at each site approved the study and these participants consented to use of their DNA for genetic analyses.

Exclusion criteria for this analysis included participants with pacemaker (PM) dependency, underlying atrial fibrillation, bundle branch block (BBB; as defined as QRS duration ≥ 120msec), and those on Class III and IV antiarrhythmics. A total of 125 MESA participants were excluded from subsequent analysis.

**DNA Extraction**: DNA was extracted from peripheral leukocytes isolated from packed cells of anticoagulated blood by use of a commercially available DNA isolation kit (Puregene; Gentra Systems, Minneapolis, MN). The DNA was quantified by determination of absorbance at 260 nm followed by PicoGreen analysis (Molecular Probes, Inc., Eugene, OR). Two vials of DNA were stored per participant at -70 degrees centigrade and subsequently aliquoted for use.

**Exposure Variables and Selection of Single-Nucleotide Polymorphisms (tagSNPs)**: Twenty-eight NOS1AP tagging SNPs beginning in the first intron & spanning the entire NOS1P gene, including 3 synonymous SNPs and 1 nonsynonymous SNP, were included for analysis. SNPs were selected in candidate gene loci according to the following
criteria: 1) within the proximal and distal 10 kbase regions 5’ and 3’ to the given
candidate gene (NCBI Build 35); 2) compatibility with the Illumina GoldenGate
technology as determined by the Assay Design Tool (TechSupport, Illumina, San Diego,
CA [19, 20]); 3) minor allele frequency (MAF) ≥0.05 or a tag ($r^2$ value ≥0.8) for another
SNP with MAF≥0.05 as determined by applying the multilocus or “aggressive” “Tagger”
option of Haploview v3 using International HapMap project data for CEPH and Yoruban
populations (release 19) [21-23]. MESA SNPs were chosen to better tag linkage
disequilibrium (LD) in African-Americans.

**Genotyping and Data Quality Control:** Genotyping was performed by Illumina
Genotyping Services (Illumina Inc., San Diego, CA) using their proprietary GoldenGate
assay. The SNPs were typed in two separate panels of 1536 markers, selected to assay
multiple phenotype x gene hypotheses. Illumina performed initial quality control in their
laboratory to identify samples and SNPs that failed genotyping according to proprietary
protocols, and sporadic failed genotypes with gencall quality score <0.25. Of 156
duplicate pairs included in 33 plates of samples typed, Illumina were blinded to 92 pairs.
Both unblinded and blinded sample replicate concordance rates were > 99.99%. After
removal of failed SNPs and samples, the genotype calling rate in MESA CG Panel 1,
which included NOS1AP was 99.93%, with maximum missing data rate per sample of
2.1%, and maximum missing data per SNP of 4.98%. The cohort genetic data were
checked for cryptic sample duplicates and discrepancies in genetically predicted sex
(using X markers) versus study database reported sex. Samples with unresolved
duplicate and sex discrepancies were removed from the genetic study database.
Ancestry Informative Markers (AIMS): AIMs were genotyped in MESA participants in two separate panels and were chosen to maximize allele frequency differences among racial/ethnic groups. In Panel 1, 97 SNPs were selected from an Illumina proprietary SNP database to maximize the difference in allele frequencies between CAU and AFA, CAU and CHN, or AFA and CHN. One SNP failed Illumina QC, resulting in 96 AIMs. In Panel 2, an additional 112 AIMs were selected from published lists that were informative for Hispanic ancestry [24, 25]. Six of these SNPs failed Illumina QC and an additional three SNPs were eliminated for technical reasons, resulting in 103 AIMs.

Outcome Variables: The QT interval duration was determined from the resting 12-lead ECG obtained for each participant. Standard 12-lead ECGs were digitally acquired using a Marquette MAC-PC electrocardiograph (Marquette Electronics, Milwaukee, Wisconsin) at 10mm/mV calibration and speed of 25 mm/sec. ECGs were performed at the six MESA clinical sites & electronically transmitted to be read centrally at Wake Forest University central ECG reading center (EPICARE). All ECGs were visually inspected for technical errors and inadequate quality. ECG processing was done by the 2001 version of the GE Marquette 12-SL program (GE Marquette, Milwaukee, Wisconsin). The QT algorithm 12SL ECG Analysis software (12SL) constructs a super lead from a median beat of all 12 leads, and the end of the T wave is defined as the point at which the slope of the downward portion of the T wave is less than 25% of maximum slope. This QT interval essentially corresponds to earliest onset of Q to latest offset of T in any lead.
Other participant characteristics that were obtained included: age, gender, race/ethnicity, cardiovascular history, and QT prolonging medications, including Class I-IV antiarrhythmics [18]. Hypertension was defined as systolic blood pressure (SBP) > 140 mm Hg or diastolic blood pressure (DBP) > 90 mmHg or taking BP meds. Race/ethnicity was determined by self report.

Statistical Methods:

Analysis of variance (ANOVA) and Fisher exact test were used to compare the baseline distributions of variables among the 4 race/ethnicity groups. Minor allele frequencies (MAFs) were calculated in the combined cohort and in each of the four race/ethnicity groups. The MAF in CAU was used as the reference allele frequencies for subsequent analyses. Allelic and genotypic frequencies were used to test for departures from Hardy-Weinberg equilibrium (HWE), stratified by race/ethnicity. A Bonferroni-adjusted p value of 0.002 (0.05/28 tagging SNPs) was used to determine departures from HWE. LD patterns were determined using standard estimates of D’ and r².

To address population stratification (PS) within each race/ethnicity, principal component analysis (PCA) with the AIMs was used to identify PCs that accounted for approximately 30% of the variation in allele frequency amongst individuals of that population subgroup. Race specific PCs were used as additional covariates in determining the associations between NOS1AP SNPs and the QT in the combined MESA cohort.

Associations between each SNP & the QT interval duration were determined using multiple linear regression analyses. QT interval duration, heart rate (HR), age, and
the race specific principal components were analyzed as continuous variables. Gender and race/ethnicity were treated as categorical variables. Covariates included age, gender, MESA field center, HR, QT altering medications, and race specific principal components. QT altering medications were defined as beta blockers, diltiazem, and verapamil (total n=338). Interactions between NOS1AP variants and race/ethnicity were formally tested. Analyses for SNP-QT interval associations were then stratified by race/ethnicity. SNP associations with the QT interval were tested using an additive model of inheritance to estimate association and effect. Concerns associated with multiple comparisons were addressed by evaluating consistency between the four race/ethnic groups.

Additional multiple comparison procedures, including Bonferroni adjustment, were used to ensure adequate type I error protection. The Bonferroni- adjusted p value for this analysis was 0.002 = 0.05/(28 tagging SNPs).

Haplotype structure was evaluated using the program Haploview. Sliding haplotype analysis was then performed within the four MESA racial/ethnic groups. A sliding window of four SNPs with a three SNP overlap was used for haplotype analysis. Associations between each haplotype and QT interval duration were evaluated by multivariate regression for common haplotypes assuming an additive genetic model.

Results

The clinical characteristics of the study participants from each of the four ethnicities in MESA are displayed in Table 1. Of the 2847 MESA participants genotyped for this study, 109 had a pre-existing bundle branch block and 16 were on Class III and IV
antiarrhythmics. These 125 participants were excluded from subsequent analyses. The final study group included 669 Caucasians (CAU), 672 African-Americans (AFA), 703 Chinese (CHN) and 678 Hispanic (HIS) participants (total n=2722). The mean age of the study participants was approximately 61 years of age. There were no significant differences among the groups with respect to mean age, gender distribution of participants, mean QT interval, or mean HR. There were significant differences between AFA and the other three race/ethnicities with respect to the prevalence of hypertension (p <0.001; Table 1).

<table>
<thead>
<tr>
<th>Variable</th>
<th>CAU (n=669)</th>
<th>AFA (n=672)</th>
<th>CHN (n=703)</th>
<th>HIS (n=678)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>61±10</td>
<td>61±10</td>
<td>62±10</td>
<td>61±10</td>
</tr>
<tr>
<td>Female (%)</td>
<td>52</td>
<td>54</td>
<td>51</td>
<td>54</td>
</tr>
<tr>
<td>HTN (% on meds)</td>
<td>35</td>
<td>54</td>
<td>37</td>
<td>41</td>
</tr>
<tr>
<td>QT interval (msec)</td>
<td>412±29</td>
<td>410±30</td>
<td>412±30</td>
<td>408±30</td>
</tr>
<tr>
<td>HR (bpm)</td>
<td>63±10</td>
<td>63±10</td>
<td>63±9</td>
<td>64±9</td>
</tr>
<tr>
<td>LBBB</td>
<td>41</td>
<td>32</td>
<td>12</td>
<td>24</td>
</tr>
</tbody>
</table>

Table 1. MESA participant characteristics according to race/ethnicity.

^a = P<0.05 vs. CAU, CHN, and HIS groups
Patterns of LD for each race/ethnicity are illustrated in Figure 1. The differences in LD patterns between the four race/ethnicity groups are reflected in the estimates of D’ and r². LD patterns show less LD in AFA compared to CHN, CAU, and HIS (Figure 1). The association between each SNP and the QT interval duration for each of the four races/ethnicities are shown in Table 2. The distribution of two SNPs in HIS (rs1572495 and rs164181) and one SNP in AFA (rs6427669) deviated from Hardy-Weinberg equilibrium (HWE; p<0.002). While SNP rs164181 was nominally associated with the QT interval duration in HIS (p=3.8x10⁻²), SNP rs1572495 in HIS and SNP rs6427669 in AFA were not associated with QT variation. Formal interaction tests between NOS1AP variants and race/ethnicity revealed eight nominally significant SNP-race interactions (data not shown). Of SNPs in HWE, race stratified analyses revealed nominally significant associations between NOS1AP variants and QT interval variation within each of the four race/ethnicities at the alpha= 0.05 level. Eleven SNPs were associated with the QT interval duration in the CAU group at the alpha= 0.05 level; four of these SNP-trait associations were significant after Bonferroni correction (rs1415259, rs4657154, rs1932933, and rs7412698). Across the SNPs, the range of QT interval increase ranged from 1.5 to 5.0 msec with each additional copy of the minor allele (p= 7.20x10⁻⁷-4.3x10⁻²). Of note, each additional copy of the minor allele in rs1932933 was associated with a 4.5 msec increase in QT in CAU (p= 7.20x10⁻⁷). There were two nominally statistically significant associations between genotype and QT interval variation in AFA (rs347298 and rs7512126); however, none of the SNP-trait associations were significant at the Bonferroni- adjusted p value of 0.002. Amongst AFA, the minor allele of rs347298 was
associated with shorter QT interval duration (p= 0.04). Conversely, the most significant SNP- trait association was found in SNP rs7512126 in which each additional copy of the minor allele conferred a 4.5 msec increase in the QT interval duration (p=0.03). Among HIS, three SNPs were nominally significantly associated with the QT (rs1415259, rs4657178, and rs6427664), with rs1415259 approaching significance after Bonferroni correction (p=3.2 x 10^{-3}). Each additional copy of the minor allele within these SNPs conferred a 3 msec increase in QT (p-value range 0.003- 0.03). In the CHN participants, five SNPs were nominally associated with the QT interval duration (rs2819328, rs6427669, rs10800465, rs3751284, and rs449908); however, no SNP-trait associations were significant at the Bonferroni-adjusted p value of 0.002. The p-value range for the SNP-trait associations was from 0.021- 0.047. Amongst CHN, the minor allele of SNPs rs2819328, rs6427669, and rs449908 were associated with shorter QT interval duration (p= 0.041, 0.047, 0.043, respectively). In contrast, the minor allele of SNPs rs10800465 and rs3751284 were associated with an increase in QT interval duration (p= 0.021- 0.028, respectively). The most significant SNP- trait association was found in SNP rs10800465 in which each additional copy of the minor allele conferred a 2.5msec increase in QT duration (p=0.021).

Figure 1 displays the race- specific association results for each SNP across NOS1AP. The red markers represent the p values for each SNP-trait association with respect to the nominal and Bonferroni-adjusted alpha levels (0.05 and 0.002, respectively). The LD structure is shown below. For AFA, significant SNP-trait associations at the nominal p value (< 0.05) spanned across the gene. In contrast, the
preponderance of significant CAU and HIS SNPs at the nominal (p< 0.05) and Bonferroni-adjusted (p< 0.002) p-values occurred at the 5’ end of NOS1AP (some within similar LD blocks). This differed in location from nominally significant SNP-trait associations in CHN, which were observed only at the 3’ end of the gene (Figure 1).

**Figure 1.** Race-Specific Negative Log\(_{10}\) P-value for Each SNP Across NOS1AP. The red markers (adjusted model) represent the p value for each SNP-trait association with respect to the nominal and Bonferroni-adjusted alpha levels (0.05 and 0.002, respectively). LD structure is shown below.

**Discussion**

Sudden cardiac death (SCD) is responsible for approximately 15% of all deaths in the US per year with approximately 60% of SCDs occurring outside the hospital setting [26]. Unfortunately, despite advances in understanding the pathophysiology of cardiac diseases and the resultant improvement in therapeutic management, the incidence of
SCD has largely remained unchanged, accounting for approximately 300,000 events per year in US [1].

Previous studies have established the QT interval as an independent risk factor for SCD[2, 5, 27-29]. While the heritability of the QT interval duration in the general population has been estimated to be approximately 30%[30-32], the associations of specific genes, such as NOS1AP, with QT interval duration have been identified and extensively replicated in predominantly Caucasian cohorts. Associations in other ethnicities, including Hispanics and African Americans, have only been identified in smaller cohort populations[6-8, 11, 17, 33].

We observed significant associations at the alpha= 0.05 level between NOS1AP SNPs and the QT interval duration in the four race/ethnicities of the MESA Study-- a large, well-studied ethnically diverse cohort without clinical cardiovascular disease (CVD). In participants with European ancestry (CAU), specific NOS1AP genetic variants were associated with QT interval variation at both the nominal and Bonferroni- adjusted p values. Sliding haplotype analysis did not reveal additional associations or improve significance levels (data not shown). It follows that our results provide three significant contributions that help expand our current understanding of the genetic influences of the NOS1AP gene on QT interval duration and cardiac repolarization.

Based upon the GWAS completed by Arking et al, subsequent studies have genotyped only the 5’ end of the NOS1AP gene in an attempt to evaluate the association between specific NOS1AP variants and QT variation [6-8, 10-12]. Subsequently, significant SNP-trait associations were limited to this part of the NOS1AP gene.
Conversely, MESA investigators used 28 tagging SNPs that spanned the entire NOS1P gene to be included for analysis. While there are inherent limitations in attempting to cover a 299kb gene with only 28 tagging SNPs, our results reveal SNP-trait associations at both the 5’ and 3’ end of NOS1AP and QT variation across race/ethnicity.

Second, our results further validate the association between variants in the NOS1AP gene and the QT interval in individuals of European ancestry. Of SNPs in HWE, more significant associations between NOS1AP variants and QT interval variation were found within Caucasians (CAU) than in any other race/ethnicity (11 of the 28 tagging SNPs; \(p = 7.20 \times 10^{-7} \text{ to } 4.3 \times 10^{-2}\)). Of note, each additional copy of the minor allele in rs1932933 was associated with a 4.5 msec increase in QT in CAU (\(p = 7.20 \times 10^{-7}\)). The fact that the MAF for the 11 significant SNPs within CAU ranged from 0.09- 0.47 suggests that the significance of these SNP-trait associations were not due only to exceedingly rare alleles in the CAU population. Furthermore, 3 SNPs were shown to be significantly associated with the QT interval even after multiple comparison adjustment with Bonferroni correction (rs1415259, rs4657154, and rs1932933 with \(p = 3.2 \times 10^{-5} \), \(5.0 \times 10^{-4}\), and \(7.20 \times 10^{-7}\), respectively). With respect to the current literature, SNP rs1415259 and rs1932933 in MESA are in LD with rs10494366 (\(r^2 = 1.00 \text{ and } 0.61 \) respectively), a SNP that has previously been shown to be associated with QT interval duration in several studies [7, 8, 10-12, 34]. Notably, after adjusting for population stratification with PCA using AIMs, neither of the 2 nominally significant SNP-trait associations in AFA was in LD with a previously reported SNP (rs16856785) that has been shown to be associated with
the QT interval duration in the Dallas Heart Study (DHS), a multi-ethnic population study with 1,506 non-Hispanic Blacks, 942 non-Hispanic Whites, and 501 Hispanics[17].

Finally, our study is the first to show associations between NOS1AP genetic variants and the QT interval in a large, established cohort that includes participants from four different race/ethnicities. Critics may argue that the credibility of gene-trait association studies can be undermined by the presence of population stratification, defined as the mixture of individuals from different genetic backgrounds [35, 36]. As a result, it could be argued that SNP-QT associations could be due to different allele frequencies between populations from different geographical regions. However, what is notable about our findings is the fact that SNP-trait associations were observed after adjusting for population stratification using PCA with AIMs. Furthermore, our findings revealed that the location of significant SNPs varies across the four racial/ethnic groups of MESA. Nominally significant CAU and HIS SNPs were at the 5′ end of NOS1AP (some within similar LD blocks); in contrast, associations in CHN were only at the 3′ end of NOS1AP. The segregation of significant SNPs across the different race/ethnicities may address the role of genetic variation at the NOS1AP locus in myocardial repolarization in these population groups. Our findings suggest a role of NOS1AP on influencing the QT interval in ethnic populations with different genetic backgrounds. The preponderance of significant CAU and HIS SNPs were at the 5′ end of NOS1AP versus the significant associations in CHN at the 3′ end suggesting a heterogenous influence of NOS1AP variants on the QT in different race/ethnicities which needs to be studied in additional cohorts for replication.
The potential clinical implications of our findings remain to be established. However, previous studies have shown that lengthening or shortening of the QT interval within even what is considered the “normal range” for QT is independently associated with a higher incidence of coronary heart disease (CHD), CHD mortality, and all-cause mortality [2-4]. As such, the US Food and Drug Administration has recommended that a 5msec prolongation in the QT interval duration in a drug trial in healthy volunteers warrants an expanded ECG safety evaluation in later stages of drug development [9].

The replication of previous findings in CAU group within the MESA cohort further substantiates the potential influence of NOS1AP on QT interval in individuals of European ancestry. Furthermore, our results observed in CHN underscore the utility of GWAS to identify common SNP-trait associations within the spectrum of previously unanticipated associations.

Limitations of our study include the fact that false positive results could be due to inadequate alpha error protection. However, multiple comparison adjustments, including Bonferroni adjustment, have been viewed as too conservative in that it assumes that individual SNPs within the same gene are not independent. Furthermore, while SNP-trait associations may be statistically significant, the clinical implications of such associations have yet to be fully established at this time. Finally, because the MESA cohort consists of individuals without CVD, it remains unclear whether our findings extend to patients with established CVD.

The genetic determinants of the QT interval and their subsequent influence on SCD in the general population are complex. The identification of genetic variants in a
large, heterogenous population may expand our understanding of cardiac repolarization and facilitate the development of new prevention strategies and therapeutic targets to prevent SCD. We have identified variations in the location and strength of association between SNPs and QT duration in CAU, HIS, and CHN individuals from a large, ethnically diverse cohort. Our data confirm previously established associations between selected SNPs in the 5’ end of *NOS1AP* and QT duration in CAU and provide suggestive evidence for novel variants at the 3’ end of *NOS1AP* associated with QT in CHN only. Further genotyping within these regions should be pursued to determine functional genetic variants affecting QT interval and their potential impact on risk for SCD.

**Table 2.** Genotypic association of SNPs in *NOS1AP* with QT interval duration for African (AFA), Chinese (CHN), Caucasian (CAU), and Hispanic (HIS) groups in MESA.

<table>
<thead>
<tr>
<th>SNP</th>
<th>Minor Allele Frequency</th>
<th>Genotypic Means (QT in msec)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>rs6664501</strong></td>
<td>C</td>
<td>GG, GC, CC</td>
<td>Add</td>
</tr>
<tr>
<td>AFA</td>
<td>0.305</td>
<td>410, 410, 411</td>
<td>0.72</td>
</tr>
<tr>
<td>CHN</td>
<td>NP</td>
<td>409</td>
<td></td>
</tr>
<tr>
<td>CAU</td>
<td>0.0007</td>
<td>413, 394 --</td>
<td>0.29</td>
</tr>
<tr>
<td>HIS</td>
<td>0.054</td>
<td>410, 411, 417</td>
<td>0.29</td>
</tr>
<tr>
<td>rs1415259</td>
<td>G</td>
<td>GG, GA, AA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AFA</td>
<td>CHN</td>
<td>CAU</td>
</tr>
<tr>
<td>----</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td></td>
<td>0.631</td>
<td>0.653</td>
<td>0.391</td>
</tr>
<tr>
<td></td>
<td>411</td>
<td>410</td>
<td>418</td>
</tr>
<tr>
<td></td>
<td>412</td>
<td>407</td>
<td>414</td>
</tr>
<tr>
<td></td>
<td>0.91</td>
<td>0.17</td>
<td>3.20x10^{-3}</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.567</td>
<td>0.197</td>
<td>0.087</td>
</tr>
<tr>
<td></td>
<td>409</td>
<td>409</td>
<td>412</td>
</tr>
<tr>
<td></td>
<td>412</td>
<td>407</td>
<td>416</td>
</tr>
<tr>
<td></td>
<td>410</td>
<td>411</td>
<td>422</td>
</tr>
<tr>
<td></td>
<td>0.18</td>
<td>0.48</td>
<td>4.3x10^{2}</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.279</td>
<td>0.304</td>
<td>0.304</td>
</tr>
<tr>
<td></td>
<td>412</td>
<td>411</td>
<td>411</td>
</tr>
<tr>
<td></td>
<td>407</td>
<td>407</td>
<td>414</td>
</tr>
<tr>
<td></td>
<td>409</td>
<td>411</td>
<td>414</td>
</tr>
<tr>
<td></td>
<td>0.07</td>
<td>0.51</td>
<td>5.0x10^{-4}</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.725</td>
<td>0.75</td>
<td>0.401</td>
</tr>
<tr>
<td></td>
<td>411</td>
<td>409</td>
<td>418</td>
</tr>
<tr>
<td></td>
<td>409</td>
<td>408</td>
<td>414</td>
</tr>
<tr>
<td></td>
<td>412</td>
<td>410</td>
<td>409</td>
</tr>
<tr>
<td></td>
<td>0.37</td>
<td>0.84</td>
<td>7.20E-07</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.46</td>
<td>0.46</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>410</td>
<td>410</td>
<td>410</td>
</tr>
<tr>
<td></td>
<td>411</td>
<td>411</td>
<td>411</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.401</td>
<td>0.401</td>
<td>0.401</td>
</tr>
<tr>
<td></td>
<td>418</td>
<td>418</td>
<td>418</td>
</tr>
<tr>
<td></td>
<td>414</td>
<td>414</td>
<td>414</td>
</tr>
<tr>
<td></td>
<td>409</td>
<td>409</td>
<td>409</td>
</tr>
<tr>
<td></td>
<td>0.17</td>
<td>0.17</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.46</td>
<td>0.46</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>410</td>
<td>410</td>
<td>410</td>
</tr>
<tr>
<td></td>
<td>411</td>
<td>411</td>
<td>411</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.401</td>
<td>0.401</td>
<td>0.401</td>
</tr>
<tr>
<td></td>
<td>418</td>
<td>418</td>
<td>418</td>
</tr>
<tr>
<td></td>
<td>414</td>
<td>414</td>
<td>414</td>
</tr>
<tr>
<td></td>
<td>409</td>
<td>409</td>
<td>409</td>
</tr>
<tr>
<td></td>
<td>0.17</td>
<td>0.17</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.46</td>
<td>0.46</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>410</td>
<td>410</td>
<td>410</td>
</tr>
<tr>
<td></td>
<td>411</td>
<td>411</td>
<td>411</td>
</tr>
<tr>
<td></td>
<td>CHN</td>
<td>0.208</td>
<td>408</td>
</tr>
<tr>
<td>---</td>
<td>-----</td>
<td>-------</td>
<td>-----</td>
</tr>
<tr>
<td></td>
<td>CAU</td>
<td>0.175</td>
<td>411</td>
</tr>
<tr>
<td></td>
<td>HIS</td>
<td>0.161</td>
<td>410</td>
</tr>
<tr>
<td>rs1538018</td>
<td>C</td>
<td>GG</td>
<td>GC</td>
</tr>
<tr>
<td></td>
<td>AFA</td>
<td>0.201</td>
<td>410</td>
</tr>
<tr>
<td></td>
<td>CHN</td>
<td>0.191</td>
<td>409</td>
</tr>
<tr>
<td></td>
<td>CAU</td>
<td>0.256</td>
<td>414</td>
</tr>
<tr>
<td></td>
<td>HIS</td>
<td>0.218</td>
<td>411</td>
</tr>
<tr>
<td>rs7412698</td>
<td>A</td>
<td>AA</td>
<td>AG</td>
</tr>
<tr>
<td></td>
<td>AFA</td>
<td>0.732</td>
<td>411</td>
</tr>
<tr>
<td></td>
<td>CHN</td>
<td>0.781</td>
<td>409</td>
</tr>
<tr>
<td></td>
<td>CAU</td>
<td>0.377</td>
<td>415</td>
</tr>
<tr>
<td></td>
<td>HIS</td>
<td>0.462</td>
<td>412</td>
</tr>
<tr>
<td>rs4531275</td>
<td>A</td>
<td>GG</td>
<td>GA</td>
</tr>
<tr>
<td></td>
<td>AFA</td>
<td>0.539</td>
<td>411</td>
</tr>
<tr>
<td></td>
<td>CHN</td>
<td>0.501</td>
<td>410</td>
</tr>
<tr>
<td></td>
<td>CAU</td>
<td>0.347</td>
<td>411</td>
</tr>
<tr>
<td></td>
<td>HIS</td>
<td>0.424</td>
<td>409</td>
</tr>
<tr>
<td>rs10918936</td>
<td>A</td>
<td>AA</td>
<td>AG</td>
</tr>
<tr>
<td></td>
<td>AFA</td>
<td>0.641</td>
<td>411</td>
</tr>
<tr>
<td></td>
<td>CHN</td>
<td>0.506</td>
<td>409</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>CAU</td>
<td>0.436</td>
<td>415</td>
<td>414</td>
</tr>
<tr>
<td>HIS</td>
<td>0.498</td>
<td>411</td>
<td>410</td>
</tr>
<tr>
<td>rs4657178</td>
<td>A</td>
<td>GG</td>
<td>GA</td>
</tr>
<tr>
<td>AFA</td>
<td>0.383</td>
<td>410</td>
<td>410</td>
</tr>
<tr>
<td>CHN</td>
<td>0.523</td>
<td>409</td>
<td>409</td>
</tr>
<tr>
<td>CAU</td>
<td>0.284</td>
<td>411</td>
<td>415</td>
</tr>
<tr>
<td>HIS</td>
<td>0.338</td>
<td>409</td>
<td>410</td>
</tr>
<tr>
<td>rs6427664</td>
<td>A</td>
<td>GG</td>
<td>GA</td>
</tr>
<tr>
<td>AFA</td>
<td>0.276</td>
<td>409</td>
<td>411</td>
</tr>
<tr>
<td>CHN</td>
<td>0.523</td>
<td>409</td>
<td>409</td>
</tr>
<tr>
<td>CAU</td>
<td>0.267</td>
<td>411</td>
<td>415</td>
</tr>
<tr>
<td>HIS</td>
<td>0.315</td>
<td>409</td>
<td>410</td>
</tr>
<tr>
<td>rs4145621</td>
<td>G</td>
<td>GG</td>
<td>GA</td>
</tr>
<tr>
<td>AFA</td>
<td>0.867</td>
<td>410</td>
<td>410</td>
</tr>
<tr>
<td>CHN</td>
<td>0.792</td>
<td>409</td>
<td>408</td>
</tr>
<tr>
<td>CAU</td>
<td>0.473</td>
<td>414</td>
<td>414</td>
</tr>
<tr>
<td>HIS</td>
<td>0.589</td>
<td>411</td>
<td>409</td>
</tr>
<tr>
<td>rs2341744</td>
<td>G</td>
<td>GG</td>
<td>GA</td>
</tr>
<tr>
<td>AFA</td>
<td>0.579</td>
<td>412</td>
<td>410</td>
</tr>
<tr>
<td>CHN</td>
<td>0.616</td>
<td>409</td>
<td>409</td>
</tr>
<tr>
<td>CAU</td>
<td>0.459</td>
<td>413</td>
<td>413</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>HIS</td>
<td>0.541</td>
<td>411</td>
<td>409</td>
</tr>
<tr>
<td>rs2661810</td>
<td>A</td>
<td>TT</td>
<td>TA</td>
</tr>
<tr>
<td>AFA</td>
<td>0.481</td>
<td>408</td>
<td>412</td>
</tr>
<tr>
<td>CHN</td>
<td>0.389</td>
<td>409</td>
<td>409</td>
</tr>
<tr>
<td>CAU</td>
<td>0.464</td>
<td>414</td>
<td>413</td>
</tr>
<tr>
<td>HIS</td>
<td>0.437</td>
<td>412</td>
<td>409</td>
</tr>
<tr>
<td>rs2819318</td>
<td>A</td>
<td>GG</td>
<td>GA</td>
</tr>
<tr>
<td>AFA</td>
<td>0.445</td>
<td>409</td>
<td>411</td>
</tr>
<tr>
<td>CHN</td>
<td>0.395</td>
<td>409</td>
<td>409</td>
</tr>
<tr>
<td>CAU</td>
<td>0.462</td>
<td>413</td>
<td>413</td>
</tr>
<tr>
<td>HIS</td>
<td>0.43</td>
<td>412</td>
<td>408</td>
</tr>
<tr>
<td>rs2819328</td>
<td>A</td>
<td>GG</td>
<td>GA</td>
</tr>
<tr>
<td>AFA</td>
<td>0.317</td>
<td>410</td>
<td>410</td>
</tr>
<tr>
<td>CHN</td>
<td>0.067</td>
<td>410</td>
<td>406</td>
</tr>
<tr>
<td>CAU</td>
<td>0.039</td>
<td>413</td>
<td>413</td>
</tr>
<tr>
<td>HIS</td>
<td>0.081</td>
<td>409</td>
<td>412</td>
</tr>
<tr>
<td>rs6427669</td>
<td>C</td>
<td>AA</td>
<td>AC</td>
</tr>
<tr>
<td>AFA</td>
<td>0.174</td>
<td>411</td>
<td>410</td>
</tr>
<tr>
<td>CHN</td>
<td>0.066</td>
<td>410</td>
<td>406</td>
</tr>
<tr>
<td>CAU</td>
<td>0.001</td>
<td>413</td>
<td>417</td>
</tr>
<tr>
<td>HIS</td>
<td>0.021</td>
<td>410</td>
<td>409</td>
</tr>
<tr>
<td>rs365522</td>
<td>G</td>
<td>GG</td>
<td>GC</td>
</tr>
<tr>
<td>----------</td>
<td>---</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>409</td>
<td>412</td>
</tr>
<tr>
<td>rs347298</td>
<td>A</td>
<td>CC</td>
<td>CA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>412</td>
<td>412</td>
</tr>
<tr>
<td>rs10800465</td>
<td>G</td>
<td>AA</td>
<td>AG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>410</td>
<td>411</td>
</tr>
<tr>
<td>rs3751284</td>
<td>A</td>
<td>GG</td>
<td>GA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>409</td>
<td>411</td>
</tr>
<tr>
<td>rs347273</td>
<td>G</td>
<td>AA</td>
<td>AG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haplotype</td>
<td>Frequency</td>
<td>Start</td>
<td>End</td>
</tr>
<tr>
<td>-----------</td>
<td>-----------</td>
<td>-------</td>
<td>-----</td>
</tr>
<tr>
<td>AFA</td>
<td>0.254</td>
<td>411</td>
<td>410</td>
</tr>
<tr>
<td>CHN</td>
<td>0.057</td>
<td>409</td>
<td>415</td>
</tr>
<tr>
<td>CAU</td>
<td>0.141</td>
<td>413</td>
<td>416</td>
</tr>
<tr>
<td>HIS</td>
<td>0.11</td>
<td>410</td>
<td>422</td>
</tr>
<tr>
<td>rs449908</td>
<td>A</td>
<td>GG</td>
<td>GA</td>
</tr>
<tr>
<td>AFA</td>
<td>0.268</td>
<td>411</td>
<td>416</td>
</tr>
<tr>
<td>CHN</td>
<td>0.24</td>
<td>411</td>
<td>405</td>
</tr>
<tr>
<td>CAU</td>
<td>0.271</td>
<td>412</td>
<td>414</td>
</tr>
<tr>
<td>HIS</td>
<td>0.239</td>
<td>411</td>
<td>413</td>
</tr>
<tr>
<td>rs446569</td>
<td>G</td>
<td>AA</td>
<td>AG</td>
</tr>
<tr>
<td>AFA</td>
<td>0.162</td>
<td>410</td>
<td>402</td>
</tr>
<tr>
<td>CHN</td>
<td>8E-04</td>
<td>409</td>
<td>416</td>
</tr>
<tr>
<td>CAU</td>
<td>7E-04</td>
<td>413</td>
<td>402</td>
</tr>
<tr>
<td>HIS</td>
<td>0.028</td>
<td>410</td>
<td>393</td>
</tr>
<tr>
<td>rs348624</td>
<td>A</td>
<td>GG</td>
<td>GA</td>
</tr>
<tr>
<td>AFA</td>
<td>0.384</td>
<td>411</td>
<td>409</td>
</tr>
<tr>
<td>CHN</td>
<td>0.136</td>
<td>409</td>
<td>404</td>
</tr>
<tr>
<td>CAU</td>
<td>0.126</td>
<td>413</td>
<td>420</td>
</tr>
<tr>
<td>HIS</td>
<td>0.185</td>
<td>410</td>
<td>408</td>
</tr>
<tr>
<td>rs7512126</td>
<td>G</td>
<td>AA</td>
<td>AG</td>
</tr>
<tr>
<td>AFA</td>
<td>0.229</td>
<td>410</td>
<td>419</td>
</tr>
</tbody>
</table>
SNPs shaded in gray indicate those that deviated from Hardy-Weinberg proportions. P-values in bold indicate an association at the p<0.05 level. P-values are adjusted for age, gender, heart rate, MESA site, QT altering medications, and population stratification. NP= not polymorphic.
References


CHAPTER 3

Use of Principal Component Analysis (PCA) Enhances the Ability to Study Associations Between Single Nucleotide Polymorphisms (SNPs) in NOS1AP and QT Interval in the Multi-Ethnic Study of Atherosclerosis (MESA)

In Chapter 2, we identified variations in the location and strength of association between single nucleotide polymorphisms (SNPs) and QT duration in African-American (AFA), Caucasian (CAU), Hispanic (HIS), and Asian (of Chinese descent; CHN) individuals in MESA. However, critics may argue that the credibility of gene-trait association studies may be undermined by the presence of population stratification (PS), defined as the mixture of individuals from different genetic backgrounds [1, 2]. The third and final chapter of this thesis discusses the potential influences of PS on determining the association between NOS1AP variants and the QT duration.

I. Population Stratification (PS)

We studied SNP-trait associations between NOS1AP genetic variants and the QT duration in MESA—a large, ethnically diverse, population-based cohort of participants without clinical cardiovascular disease ages 45 to 84 years old recruited in six US communities from four racial/ethnic groups (CAU, AFA, HIS, and CHN). However, a significant drawback to such a population is the fact that allele frequency differences among ethnic groups and subgroups can result in false associations at markers whose
frequency differs across subpopulations. Thus differences in subpopulation structure
due to systematic differences in ancestry can mimic the signal of association in that
there is a significant difference in allele and genotype frequencies between cases and
controls (Figure 1). This population substructure, or stratification (PS), can confound the
validity of association studies and warrants strategies in accounting for ancestry [1, 3-6].
PS is especially problematic for diseases that have different prevalence rates across
ancestral populations [3].

**Figure 1. Effects of Population Structure at a SNP Locus.** The figure above shows an
example in which the case population has an excess of individuals from population 2, a
population in which the frequency of allele A is lower than that of population 1.
Population structure can mimic the signal of association in that there is a significant
difference in allele and genotype frequencies between cases and controls. Adapted from
Marchini et al. [4]
II. Population Stratification (PS): An Example of Confounding

Classically, confounding is the distortion of the relationship between an exposure of interest and the disease due to the presence of an extraneous factor that is not a causal intermediary ("Factor X" in Figure 1, panel A). The confounding variable ("Factor X") must be statistically associated with both the exposure and the disease but is not a result of the exposure variable [7]. It follows that a confounder can either mask a real association or create a false one between the exposure of interest and disease.

In similar fashion, PS is a manifestation of confounding that may bias estimates of the SNP-trait association due to the presence of a factor that is associated with the genotype(s) of interest (Figure 2, panel B). In PS, ethnicity can serve as a surrogate for the confounder, which may be environmental or genetic (Figure 2, panel B). It follows that controlling for ethnicity can reduce bias due to PS. Adjusting for a known confounder will effectively eliminate bias; however, if the confounder is unknown, adjustment for ethnicity can reduce bias from confounding to the extent that ethnicity correlates with the genotype of interest and the true risk factor [1].

![Figure 2. Classical Confounding (Panel A) and Population Stratification (PS; Panel B).](image)

Solid bidirectional arrows indicate correlations that are not causal. Solid unidirectional arrows indicate direction of causal relationship. Dotted unidirectional arrows indicate confounded association. Panel A: In classical confounding, an exposure of interest...
correlated to an alternate risk factor but not a causal intermediary (Factor X) can be falsely associated with the risk of disease. Panel B: In population stratification (PS), the genotype of interest can be incorrectly associated with the risk of disease due to its correlation with both ethnicity and the alternate risk factor (Factor X). Adapted from Wacholder et al. [1]

A landmark study performed by Knowler et al (1988) exposing the influences of population admixture on the interpretation of disease association between the Gm$^{2,5,13,14}$ haplotype and the incidence of type 2 diabetes (T2DM) among Native Americans of Pima- Papago descent with varying degrees of European ancestry has been frequently cited in the literature as the classic example of PS [8]. Among 4,920 residents in the studied Native American community, there was a very strong negative association between the presence of the haplotype Gm$^{2,5,13,14}$ and T2DM (Figure 3; prevalence ratio = 0.27, 95% CI 0.18-0.40).

Figure 3. Prevalence of T2DM by Age and the Presence of the Haplotype Gm$^{2,5,13,14}$ Among Native Americans of Pima- Papago Descent. Adapted from Knowler et al. [8]
However, previously published studies of the distributions of Gm antigens in Pima, Papago, and other Native-American tribes had shown that the haplotype Gm$^{2;5,13,14}$ was a very sensitive marker for Caucasian admixture in American Indians. Accounting for Indian heritage demonstrated that the estimates of the OR between haplotype status and T2DM were much weaker (0.73 95% CI= 0.17 - 1.8); moreover, it revealed the direct relationship of Indian heritage on T2Dm prevalence (Figure 4) [1, 8]. Thus it was most likely the difference in frequency of Caucasian and Indian alleles that influenced the risk for T2DM, rather than the direct action of the Gm haplotype [8].

Figure 4. Age-adjusted Prevalence of Type II Diabetes (T2DM) by the Presence of the Haplotype Gm$^{2;5,13,14}$, According to Indian Heritage, Among Native Americans Residents of Pima- Papago Descent. Adapted from Knowler et al. [8]

The study by Knowler et al (1988) was a pivotal study in that it exposed the potential confounding effect of admixture on the interpretation of gene-disease association
studies; furthermore, it raised awareness in addressing PS in studies of genetic markers of disease. The impact of this study on future investigations is evident in that subsequent association studies have drawn upon predominantly all-Caucasian, population-based cohorts in an attempt to reduce bias due to PS [9-15]. Unfortunately however, while genetically homogenous populations reduce the potential of spurious associations due to PS, an unfortunate consequence is that the findings are largely limited to such populations. In a society where population admixture can potentially confound associations between complex traits and genetic polymorphisms, such limitations can obscure our complete understanding of the functional components underlying these complex phenotypes.

III. Ancestry Informative Markers (AIMS)

Ancestry-informative markers (AIMs) can be used to account for hidden PS and improve analysis [3, 5]. AIMs are a set of polymorphisms which occur at different allele frequencies between populations from different geographical regions. In this sense, AIMs can be used to infer ancestry; furthermore, these markers can be used to examine population substructure within ethnic groups.

In MESA, AIMs were genotyped in participants in two separate panels and chosen to maximize allele frequency differences among racial/ethnic groups. In Panel 1, 97 SNPs were selected from an Illumina proprietary SNP database to maximize the difference in allele frequencies between CAU and AFA, CAU and CHN, or AFA and CHN. One SNP failed Illumina QC, resulting in 96 AIMs. In Panel 2, an additional 112 AIMs were selected from published lists that were informative for Hispanic ancestry [16, 17]. Six of these
SNPs failed Illumina QC and an additional three SNPs were eliminated for technical reasons, resulting in 103 AIMs.

Figure 5 demonstrates the comparison between self-reported ethnicity and estimated ancestry derived from AIMs. Without population admixture, the ancestry proportion estimates for each self-reported ethnicity would equal 1 and thereby be represented by a single bar. Figure 5 reflects the heterogeneity within each self-reported race/ethnicity. Furthermore, the bar graph reveals the heterogeneity of the ancestry proportion estimates within the HIS group, thereby suggesting that self-reported ancestry may be less reliable for the HIS population subgroup (Figure 5).

Unfortunately, AIMs are not perfectly informative, as the allele frequency difference at each marker between the parental populations may not equal to one. As it is quite difficult to determine the allele frequency of each marker, the quality of estimates is highly dependent on the accuracy of the allele frequency estimates.
Figure 5. Ancestry Proportion Estimates Amongst Self-Reported African-American (AFA), European (CAU), Chinese (CHN), and Hispanic (HIS) Groups. Shown is the comparison between self-reported ancestry and the estimated ancestry proportion.

IV. Principal Component Analysis (PCA)

Principal component analysis (PCA) is a mathematical procedure in which a number of possibly correlated variables are transformed into a smaller number of uncorrelated variables called principal components (PCs). By seeking a few linear combination of variables which can be used to summarize the original data, PCA reduces dimensionality through a parsimonious summarization of the data [18]. It follows that PCs have two unique properties: 1) PCs are orthogonal, that is, they are not correlated with each other; 2) and each, in turn, has maximum variance, given that all variables are mutually uncorrelated [19]. The variances of the PCs are termed “eigenvalues”. Larger eigenvalues reflect greater importance of its associated PC in representing the
information in the predictors. Eigenvalues that approach zero are thus reflective of possible collinearity amongst the original predictors. The number of zero or near-zero eigenvalues reflects the number of colinearities or near-colinearities among the predictors. Thus, summarizing the variability in the data using only the PCs with the highest variances will consequently result in reducing the dimensionality of the original predictors. It follows that PCA can be particularly useful for data sets in multiple dimensions, such as gene association studies. Specifically, PCA provides an opportunity to address the variability amongst allele differences within each race/ethnicity in MESA.

Figure 6 displays PCA in 2D from the MESA AIMS. The green, blue, black and red dots represent individuals from CHN, CAU, HIS, and AFA race/ethnic groups, respectively, as defined by self-report. The scatter plot nicely illustrates two underlying properties regarding the genetic ancestry of the four race/ethnic groups in MESA — 1) the genetic ancestry of CHN (green) and the CAU (blue) race/ethnic groups in MESA are fairly homogenous. These findings corroborate the findings of previous studies and reflect why these subpopulations have traditionally been viewed as “ideal study populations” for genetic association studies [9-13]; 2) conversely, there is more heterogeneity observed in both the AFA (red) and HIS (black) race/ethnic groups. In fact, amongst 677 self reported HIS, there are 102 individuals whose PCA suggested that they would be better re-classified in one of the other ethnic groups predominately comprised of AFA or CAU. This heterogeneity in genetic ancestry reflects how a mixture of individuals from different genetic backgrounds could potentially confound the validity of SNP-trait
associations within population subgroups, especially in HIS or AFA. It follows that PCA provides one strategy in accounting for PS [1, 3-6].

Figure 6. Two Dimensional (2D) Graphical Representation of the First Two Principal Components (PCs) Derived from Ancestry Informative Markers (AIMS). Above is a 2D scatter plot of the first 2 principal components (PCs) derived from genetic ancestry as defined by AIMS. The green, blue, yellow and red dots represent individuals from CHN, CAU, HIS, and AFA ethnicity, respectively, as defined by self-report.

To address population stratification (PS) within each race/ethnicity, PCA with AIMS was used to identify PCs that accounted for approximately 30% of the variation in allele frequency amongst individuals of that population subgroup. In applying PCA to the MESA cohort, 10 PCs in the AFA, CHN, and HIS subgroups and 14 PCs in the CAU subgroup were needed to explain 30% of the allele variability in those respective subpopulations. The race specific PCs were then used as additional covariates in
determining the associations between NOS1AP SNPs and the QT in the combined MESA cohort.

Thus, the regression models used to describe the SNP-trait associations were modified as follows:

AFA: \(QT = HR + \text{age} + \text{gender} + \text{QT_med} + \text{MESA site} + \text{SNP} + \text{PC1}_{\text{AFA}} + \text{PC2}_{\text{AFA}} + \ldots + \text{PC10}_{\text{AFA}}\)

HIS: \(QT = HR + \text{age} + \text{gender} + \text{QT_med} + \text{MESA site} + \text{SNP} + \text{PC1}_{\text{HIS}} + \text{PC2}_{\text{HIS}} + \ldots + \text{PC10}_{\text{HIS}}\)

CHN: \(QT = HR + \text{age} + \text{gender} + \text{QT_med} + \text{MESA site} + \text{SNP} + \text{PC1}_{\text{CHN}} + \text{PC2}_{\text{CHN}} + \ldots + \text{PC10}_{\text{CHN}}\)

CAU: \(QT = HR + \text{age} + \text{gender} + \text{QT_med} + \text{MESA site} + \text{SNP} + \text{PC1}_{\text{CAU}} + \text{PC2}_{\text{CAU}} + \ldots + \text{PC14}_{\text{CAU}}\)

Where:

\(QT\) = QT interval in msec

\(HR\) = heart rate in beats per min (bpm)

\(QT\_med\) = QT altering medications, as defined as beta blockers, diltiazem, and verapamil

\(MESA\_site\) = MESA field center

\(PC\) = Principal Component

Table 1 compares the unadjusted and adjusted parameter (beta) estimates and associated P values for the SNP-trait associations between NOS1AP SNPs and QT interval across self reported race/ethnicity in MESA when accounting for PS using PCA with AIMS. What is notable about our findings is the fact that SNP-trait associations were observed after adjusting for PS using PCA with AIMS. Furthermore, the parameter (beta) estimates and associated P values for the associations between NOS1AP SNPs and QT interval were similar between the two models. Thus, while it could be argued that
SNP-QT associations could be due to different allele frequencies between populations from different geographical regions, our findings suggest that PS may not significantly confound SNP-trait associations as previously estimated.

### Table 1. Associations Between NOS1AP SNPs and QT Interval across Self Reported Race/ethnicity in MESA Unadjusted and Adjusted for Population Stratification (PS)


Population stratification (PS), even within subsets stratified by self-reported ethnicity, can confound gene-trait associations. Undetected PS can lead to both false positive results (Type I error) and failures to detect real associations [4]. The potential for ancestral differences to distort the relationship between NOS1AP SNPs and the QT interval across MESA race/ethnic groups warranted strategies to account for population
substructure. We have shown that the application of PCA with AIMs underscores the value of examining genetic ancestry in genetic epidemiology studies by reducing genetic heterogeneity. As the global health burden of common diseases, including sudden cardiac arrest, continues to be shouldered by the individuals of low and middle income countries of predominantly non-European descent, future studies will be needed to investigate the association of genetic markers for various diseases in multiple ethnic groups. Such methods will help further our understanding of the functional components underlying these complex phenotypes.
References


Curriculum Vitae  
Sidharth Anil Shah, M.D.  
sshah@wfubmc.edu  

EDUCATION AND TRAINING  

1995-1999 Duke University  
Durham, North Carolina  
Bachelor of Science, Cum Laude, May 1999  

1999-2000 Duke University Clinical Research Institute  
Durham, North Carolina  
Clinical Research Assistant  

2000  
Aga Khan Hospital  
Nairobi, Kenya  
Medical Internship, June 2000  

2000-2004 University of Kansas School of Medicine (KUSM)  
Kansas City/Wichita, Kansas  
M.D. – May 2004  

2004-2005 Internship, Internal Medicine  
University of North Carolina- Chapel Hill  
Chapel Hill, North Carolina  

2005-2007 Residency, Internal Medicine  
University of North Carolina Hospitals  
Chapel Hill, North Carolina  

2007  
Diplomate in Internal Medicine  
American Board of Internal Medicine, October 2007  

2007-2008 Graduate Student  
Clinical Epidemiology and Health Services Research  
Wake Forest University Graduate School of Arts and Sciences Winston-Salem, North Carolina  

2007-2008 Cardiology Fellow  
Wake Forest Baptist Hospital, Section on Cardiology  
Winston-Salem, North Carolina  

AWARDS AND FUNDING  

1996-1999  
Dean’s List  
Duke University  

1998  
Howard Hughes Research Fellowship
Duke University

1999 Graduation Honors, Cum Laude
Duke University

2001 Lose Medical Scholarship
University of Kansas School of Medicine (KUSM)

2001, 2002 American Medical Association- Medical Student Sections (AMA- MSS) National Recruitment Award

2002 Irvin R. Medical Scholarship
KUSM

2003 Gladys Medical Scholarship
KUSM

2003 Gouldner, R. Gouldner Medical Scholarship
KUSM

2004 Grant and Helen Wagner Scholarship
KUSM

2005 Semi-finalist- National ACP Poster Competition:
Evaluating Heart Failure Treatment and Follow-up by Medical Specialty. George Martinez, Sidharth Shah, and H Farhoud.KUSM

2006 Kaiser Permanente Excellence in Teaching Award
University of North Carolina School of Medicine

2007- pres. Recipient T32 Grant—CVD Epidemiology Training Program Wake Forest University Health Sciences

2009 Research Silver Award (Fellows’ Division) Internal Medicine Resident and Fellow Research Day Wake Forest Baptist Hospital

2009 AM Nomeir Excellence in Echocardiography Award.
Wake Forest Baptist Hospital/Section on Cardiology

PUBLICATIONS


**RESEARCH PRESENTATIONS**

“Evaluating Heart Failure Treatment and Follow-up by Medical Specialty”. George Martinez, Sidharth Shah, H. Farhoud. KUSM. ACP Kansas Chapter Meeting- 2004- Poster Presentation

“Evaluating Heart Failure Treatment and Follow-up by Medical Specialty”. George Martinez, Sidharth Shah, H. Farhoud. KUSM. National ACP Meeting - 2005- Poster Presentation.

Regional ACP Meetings- Poster Presentation


“Incidental Pulmonary Nodules Found on CT Angiograms: Characteristics and Recommended Follow Up”. SG Truitt, WB Hall, MP Rivera, LP Scheunemann, SA Shah, DA Polisner, LA Parker, SS Carson. ATS National Conference- 2008- Poster Presentation


“The Identification and Pathophysiologic Implications of Ang-(1-12) in Cardiovascular Disease States”. Wake Forest University School of Medicine Cardiology Research Conference- January 2008- Oral Presentation

“Cardiac Transplantation: Evaluation of the Donor”. Wake Forest University School of Medicine Cardiology Fellows’ Conference- August 2008- Oral Presentation

”EP Review: Wolff-Parkinson-White (WPW)”. Wake Forest University School of Medicine Cardiology Fellows’ Conference- October 2008- Oral Presentation

“Cardiac Arrest Syndrome”. Wake Forest University School of
PROFESSIONAL EXPERIENCES

Attended 33rd Ten-Day Seminar on the Epidemiology and Prevention of Cardiovascular Disease- August 2007

Attended 2nd Annual Summer Workshop on Genetic Epidemiology of Complex Traits- May28-30, 2008

American Heart Association (AHA)

American Heart Association- Epidemiology Counsel
American College of Cardiology Fellows in Training (ACC-FIT)

AHA- Council on Functional Genomics and Translational Biology

PROFESSIONAL MEMBERSHIP

Member, American Heart Association Council on Epidemiology and Prevention (AHA)

Member, American College of Cardiology (ACC)