THE GENETICS OF QUANTITATIVE TRAITS ASSOCIATED WITH CARDIOVASCULAR DISEASE IN AFRICAN AMERICANS

By

Janina Maria Jeff

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements for

the degree of

DOCTOR OF PHILOSOPHY

in

Human Genetics

May, 2012

Nashville, Tennessee

Approved by:

Professor Scott Williams

Professor Dana C. Crawford

Professor Marylyn D. Ritchie

Professor Dan Roden

Professor Alyssa Hasty

To my family “it takes a village to raise a child”

To my wonderful circle of friends for their companionship and support

To my mother, Gretchen Jeff, you are my inspiration and I love you with all my heart

To my grandfather the late Dr. Morris F.X. Jeff, PhD, for teaching me at a young age the importance of education. Every day I strive to continue your legacy in all avenues of my life.

ACKNOWLEDGEMENTS

This work was supported by several resources. Genotyping for the Jackson Heart Study (JHS; Chapter II, section A) was provided National Institutes of Health (NIH) Pharmacogenetics Research Network U01 HL65962. The Jackson Heart Study is supported by NIH contracts N01-HC-95170, N01-HC-95171, and N01-HC-95172 that are provided by the National Heart, Lung, and Blood Institute (NHLBI) and the National Center on Minority Health and Health Disparities. The genome-wide association study performed in African Americans (Section B of Chapter II) was funded by two eMERGE network sites U01-HG-004609 (Northwestern University) and U01-HG-004603 (Vanderbilt University, also serving as the Coordinating Center). The eMERGE Network was initiated and funded by National Human Genome Research Institute (NHGRI), in conjunction with additional funding from National Institute of General Medical Sciences (NIGMS). Fibrinogen cluster variants genotyped in NHANES III (sections B of Chapters III and IV) were provided by the John Hopkins University under the federal contract number (N01-HV-48195) from the NHLBI. Genotyping for the replication of candidate gene results for ECG (section A of Chapters II and IV) and fibrinogen/hematological traits (sections B of Chapters III and IV) was funded by the NIH/National Human Genome Research Institute (NHGRI) U01HG004798 (EAGLE). Funding was also provided by the Genetic Variation and Human Phenotypes training grant (T32GM080178).

A special thanks to the Vanderbilt University Center for Human Genetics Research, Computational Genomics Core (CGC) for their analytical and computational support. I would like to personally acknowledge Kristin Brown-Gentry (member of the CGC) for her statistical expertise and personal support. Part of the genetic data for this work was generated by the Vanderbilt DNA Resources Core and the Haines Lab. I would also like to thank the Centers for Disease Control and Prevention (CDC) for access to genetic NHANES, as well as the National Center for Health Statistics and the Research Data Center for remote access to perform all analyses with NHANES data.

This work would not be possible without my thesis committee: Drs. Scott Williams (committee chair), Marylyn Ritchie (co-chair), Dana Crawford, , Dan Roden, and Alyssa Hasty for their insightful input, guidance, and for intellectually challenging me. Sincere thanks to my mentor Dana Crawford for recognizing my strengths and belief in my scientific ability. Her dedication and commitment to her students and their matriculation through the human genetics program is indescribable. I am very grateful to have been trained by someone with such expertise and talent in this field.

Special thanks to the current and previous members of the Crawford lab for their friendship, support, and creativity. I would also like to thank all (students and faculty) in the Center for Human Genetics in Research (CHGR) who helped me through difficult times with countless practice exams and reviewing my work; without your constructive criticism and honesty this work would not be possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

OVERVIEW

Cardiovascular disease (CVD) is the leading cause of death in most developed countries. In addition to environmental risk factors, such as diet and physical activity, genetic risk factors contribute to the CVD risk. Risk of CVD is not uniformly distributed across populations, as African Americans and Hispanics have more risk factors for CVD compared with European Americans, yet prevalence remins higher in Europeans. Although the genetics of CVD are well-studied in European Americans, few studies have been performed in minority populations.

CVD is common and complex, and while there has been progress determining the genetic risk factors, there is inconsistency across studies. One likely contributor to this discrepancy in the literature is probably due to phenotypic heterogeneity across studies. The use of quantitative traits to identify genetic risk factors is a potentially more powerful, informative and uniform approach compared with the use of binary or qualitative traits (such as CVD case status). The primary objective of this work is to identify genetic risk factors associated with the regulation of fibrinogen/hematological and electrocardiogram (ECG) traits in African Americans. Both fibrinogen/hematological and ECG traits are common clinical characteristics of CVD and have a strong genetic component. Using three approaches: candidate gene, genome-wide association studies (GWAS), and testing for genetic interactions (gene-gene and gene-environment), we identify novel and previously identified genetic associations with ECG traits and fibrinogen/hematological traits in African Americans and other global populations.

Chapter I gives an in-depth description of the biology and epidemiology of ECG and fibrinogen/hematological traits. The advantages of using quantitative traits compared to disease outcomes, as well as their role in CVD risk are highlighted. Additional background information about what is known concerning the genetics of ECG and fibrinogen/hematological traits and what has yet to be determined is also discussed. Lastly, the current state of genetics for ECG and fibrinogen/hematological traits in diverse populations is described, highlighting the potential advantages and obstacles.

Chapter II focuses on ECG traits in African Americans. In section A, we performed a candidate gene study with 65 variants in a sodium channel gene, *SCN5A*, in >3,000 African Americans from the Jackson Heart Study. In this section we also tested for replication of significant associations in an independent study population of non-Hispanic blacks from the Third National Health and Nutrition Examination Survey (NHANES III). In section B, we conducted a genome-wide association study (GWAS) in 455 African Americans from the Vanderbilt Genome-Electronic Records Project and Northwestern University NUgene Project. We identified novel associations, as well as determined the generalizability of previous GWAS findings, originally identified in European and/or Asian populations, in African Americans. In the last section of this Chapter, we tested for gene-gene interactions between variants in the *SCN5A* gene. Then we tested for gene-environment interactions with environmental risk factors age and sex with variants in the *SCN5A* gene.

Chapter III focuses on identifying genetic variants associated with fibrinogen and hematological traits in African Americans. In section A we tested 25 variants in the three genes that encoded the fibrinogen polypeptide, *FGA, FGB,* and *FGG* for an association

with fibrinogen levels in non-Hispanic blacks from NHANES III. We also tested these variants for pleiotropic effects with hematological traits: platelet count, platelet distribution width, mean platelet volume, white blood cell counts and serum triglycerides also recorded in NHANES III. Section B focuses on the generalization of previously identified GWAS associations for fibrinogen and hematological traits in the combined NHANES dataset of non-Hispanic blacks. In the last section of this chapter we investigate the role genetic modifiers (gene-gene and gene-environment interactions) have on fibrinogen levels. This section was limited to variants in the fibrinogen gene cluster in non-Hispanic blacks from NHANES III.

Replication and generalization of significant genetic associations is important in confirming genetic disease risk. In Chapter IV, we replicate and generalize findings in Chapter II and III in non-Hispanic whites and Mexican Americans from NHANES. In section A, we focus on the ECG traits. We tested *SCN5A* significant associations for generalization in non-Hispanic whites and Mexican Americans from NHANES III. In this section we also replicate previously identified GWAS variants for ECG traits in non-Hispanic whites and generalize these variants in Mexican Americans. Using the same approach in section B of this chapter, we performed all tests in non-Hispanic blacks (from Chapter III) with fibrinogen and hematological traits in non-Hispanic whites and Mexican Americans in NHANES III and 99-02.

Chapter V summarizes Chapters II-IV, specifically discussing the limitations and strengths of our findings. In this chapter we describe how these results contribute to the little known about the quantitative traits of cardiovascular disease in African Americans.

Additionally, we discuss future analyses using alternative approaches not described here for identifying genetic risk variants in African Americans.

CHAPTER I

INTRODUCTION

Cardiovascular Disease

Cardiovascular disease (CVD) is a class of diseases that involve the heart and/or blood vessels (Arking and Chakravarti, 2009; Mendis et al., 2011). Cardiovascular disease is not a single disease nor is it limited to diseases that affect the heart. Some of the most common CVDs include atherosclerosis, cerebrovascular disease (commonly referred to as stroke), diseases of the aorta and arteries (arterial thrombosis), hypertension, coronary heart disease, and myocardial infarction (MI, commonly referred to as a heart attack). Cardiovascular disease also includes aneurysms, angina (chest pains), peripheral vascular disease, cardiac arrhythmias, rheumatic heart disease, cardiomyopathies, pericarditis, cardiac tumors, myocarditis, and congenital heart disease (Mendis et al., 2011).

Clinicians often diagnosis patients with heart disease if the patient has recently had a heart attack, stroke, or has been previously diagnosed with hypertension or atherosclerosis. There are several clinical procedures to confirm cardiovascular disease in patients. Some of the most common procedures include an electrocardiogram, complete blood exam, coronary angiographs, echocardiographs, and nuclear scans(2009a).

Cardiovascular disease is the leading cause of death worldwide and is responsible for over 17.3 million deaths per year (2011). Greater than 80% of these deaths are in

middle to low income countries (2009b). While the highest rates for CVDs are in low and middle income countries, it is also the leading cause of death in the United States and most developed countries (Figure 1.1) (Lloyd-Jones et al., 2009; 2009b). The correlation with income and CVD status is also observed in the United States (US), particularly in the southeastern states which have the highest prevalence of CVD in the US (Lloyd-Jones et al., 2009). In developed countries, the majority of CVD cases in Americans are in relatively young individuals. Over 80 million Americans have one or more CVDs, but only 38 million are ≥ 60 years of age (Lloyd-Jones et al., 2009). In addition to age, there is a difference in CVD disease incidence between men and women. In previous years, CVD was higher in males compared to females, but recent data suggest that females have a slightly higher prevalence of CVD (Lloyd-Jones et al., 2009). Racial disparities also exist in the United States for cardiovascular diseases. Europeans have higher prevalence of CVD compared to other race/ethnicities in the US but African Americans have the highest mortality rates (Thom et al., 2006; Lloyd-Jones et al., 2009).

There are several reasons that explain the increased CVD prevalence in recent years. Cardiovascular diseases have a complex etiology and are likely caused by a combination of several risk factors. Increased age has been consistently associated with CVD risk, and being male increases this risk among middle-aged individuals (Mendis et al., 2011; Mendis, 2005; Lloyd-Jones et al., 2009; Thom et al., 2006). Modifiable risk factors include diet, exercise, stress, and tobacco use, all of which are directly impacted by an individual's lifestyle (Yusuf et al., 1998; Grundy, 1980; Grundy et al., 1999). These risk factors can subsequently lead to elevated lipid levels, high blood pressure, and high

blood glucose, all of which are associated with CVD and affect over 34% of American adults (Lloyd-Jones et al., 2009; Kannel and McGee, 1979; Wilson et al., 1998).

In addition to these risk factors, CVD has a large genetic component. Heritability studies of CVD risk factors and other CVD related phenotypes report that anywhere from 20-80% of the phenotypic variance is explained by genetic factors (Rose et al., 1998; Hunt et al., 2002; Tholin et al., 2005; Hunt et al., 1989; Nelson et al., 2000; Poulsen et al., 1999; Feinleib et al., 1977). CVD outcome variables and traits have been described in familial studies and are said to "run in families" (Bhagavatula et al., 2004; Mjos et al., 1977; Thelle and Forde, 1979; Snowden et al., 1982). Additionally, linkage studies and small sequencing studies have identified rare genetic mutations associated with cardiovascular disease and related phenotypes (Benson et al., 1996; London et al., 2007; Best et al., 2008). More often than not, these mutations are rare and family-specific and are usually not observed at the population level. The work contained herein is focused on identifying the common genetic factors underlying cardiovascular disease risk in the general population.

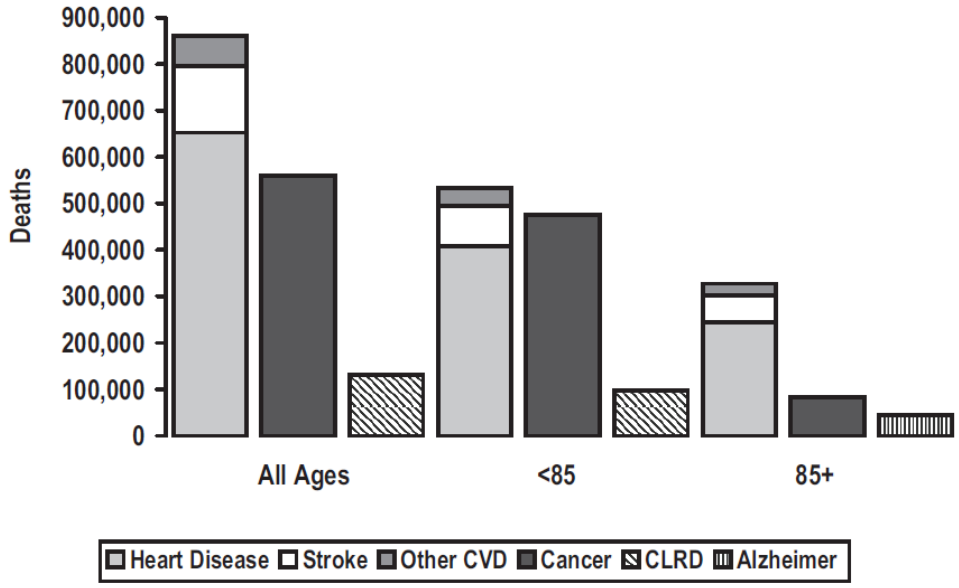


Figure 1.1 Major causes of death among Americans. Causes of deaths among both sexes in the United States in 2005. (Obtained from the National Center for Health Statistics (NCHS) and the National Heart, Lung, and Blood Institute (NHLBI) (Lloyd-Jones et al., 2009). Abbreviations: Cardiovascular Disease (CVD), Chronic Lower Respiratiry Disease (CLRD)

Quantitative Traits Associated with Cardiovascular Disease

More often than not, patients are diagnosed with CVD based on a clinical outcome event, such as myocardial infarction or stroke. Individuals diagnosed with CVD usually have other diseases that may have contributed to the development of CVD over time. Identifying and treating these diseases or conditions before diagnosis of CVD may reduce the likelihood of life threatening cardiovascular events such as myocardial infarction and stroke**.** The use of quantitative measurements or intermediate phenotypes is an accurate method of identifying patients at high-risk for CVD (Vasan, 2006). There are

several quantifiable biological parameters that are often used by clinicians to identify high-risk CVD patients. As previously mentioned, electrocardiographic (ECG) measurements, echocardiograms, inflammation markers, blood glucose measurements, and lipid levels are common procedures/measurements used to asses CVD status. An electrocardiogram measures and records the electrical activity of the heart, while an echocardiogram is a test that uses sound waves to create a visual image of the heart. Both tests provide information about the physical condition of the heart including: damage to the heart, heart rhythm, size and position of the heart chambers, abnormal heart valves, heart murmurs, infections around the heart or on the heart valves, and the source of blood clots. Recently, inflammation markers such as fibrinogen levels, white blood cell count, and C-reactive protein have been associated with CVD (Pearson et al., 2003). Elevated levels of inflammatory markers can lead to arterial thrombosis that is caused by a clot in the arteries thereby limiting the blood supply to the heart.

Using intermediate or quantitative traits is an excellent alternative to using clinical outcomes for complex diseases such as CVDs. Given the complexity of CVD, it is likely that different genetic etiologies explain several traits or intermediate phenotypes of CVD. Quantitative measurements are more robust to phenotypic heterogeneity and are often more uniform and informative compared to disease outcomes. Additionally, quantitative traits also have more statistical power compared to disease outcome variables. The research reported in succeeding chapters will focus on two groups of quantitative phenotypes associated with cardiovascular disease: electrocardiographic (ECG) traits and fibrinogen levels (and/or other hematological traits).

Electrocardiographic Traits

An electrocardiogram (ECG) is a common clinical procedure usually performed on patients who are at high risk for CVD or experience chest pain or dizziness (Micheal A Chizner, 2004). An ECG measures the electrical activity of the heart and is a useful tool in assessing cardiac arrhythmias, myocardial infarction, pericarditis, and other cardiac abnormalities (Micheal A Chizner, 2004). The ECG records electrical depolarization and repolarization of the heart and provides information about the electrical activity of the cardiac cycle. An ECG records the duration, amplitude, and direction of electrical activity during the cardiac cycle. The most commonly used ECG has twelve leads (a record of electrical activity between two electrodes), which provides measurements for both the frontal and horizontal planes and captures views of the left ventricle from twelve different angles (Aehlert, 2006). ECG duration measurements (P, PR, QRS, QT, ST and T durations) reflect the time spent for depolarization or repolarization in each phase of cardiac conduction (Micheal A Chizner, 2004). ECG axes (P, QRS and T) values represent the direction and position of myocardial cell depolarization. Usually the P wave, a QRS complex, and a T wave are visible on the ECG (Figure 1.2A). The ECG begins with the P wave which occurs during atrial depolarization, and represents the electrical impulse from the sinoatrial (SA) node towards the atrioventricular (AV) node that then spreads to the left and right atrium. The PR interval is the time the electrical impulse takes to go to the sinus node to the AV node then to the ventricles (lower chambers of the heart). It is measured from the P wave to the start of the QRS complex (Figure 1.2). The QRS marks the start of depolarization of the ventricles, and the ST segment represents the ventricle once they have fully depolarized (Figure 1.2). After depolarization, the left and right ventricles repolarize, represented as the T wave on the

ECG. The QT interval represents the time it takes the ventricles to depolarize and repolarize and is measured in the ECG from the start of the QRS complex to the end of the T wave (Figure 1.2).

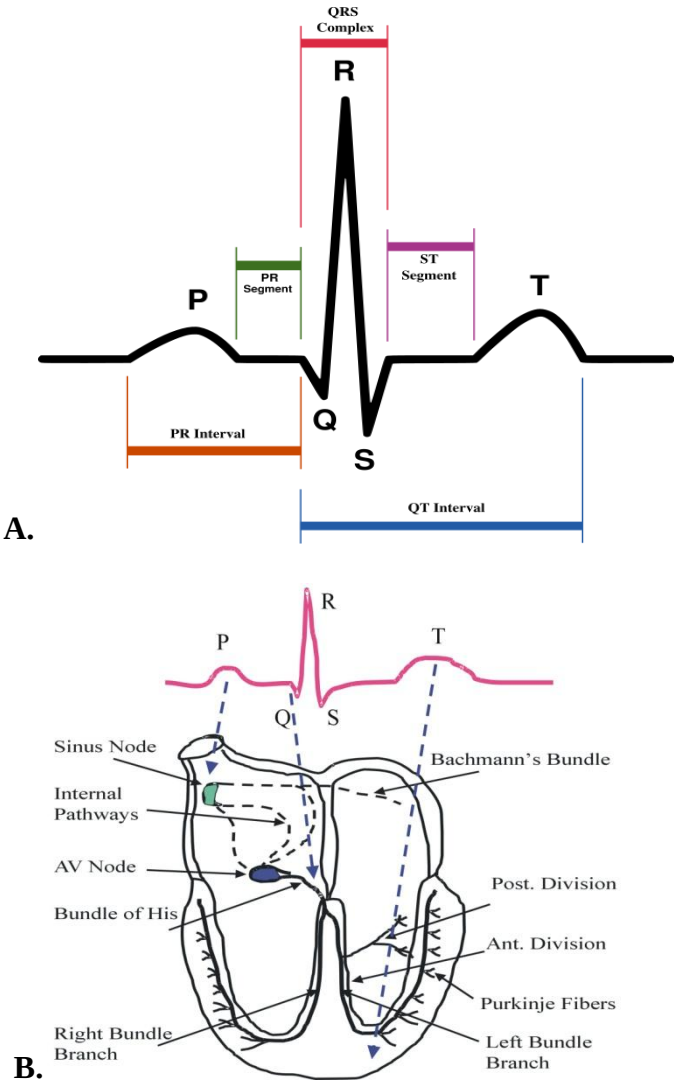


Figure 1.2 An example of a normal electrocardiograph. (A. Obtained from http://www.dentalarticles.com/visual/d/index.php B. http://www.meditech.cn/meditech-edu/ecg-2.asp)

The risk profile of an individual with an abnormal ECG is consistent with individuals diagnosed with other CVD outcomes such as myocardial infarction. In addition to traditional risk factors such as age and sex, hyperlipidemia, hyperinsulinaemia, and hypertension are also associated with ECG traits (Lind et al., 1997). As with other CVD outcome variables, a considerable amount of ECG traits are variability can be explained by genetics. Several twin studies report heritability estimates as much as 30-60% of the variability in ECG traits attributable to genetic factors (MATHERS et al., 1961; Perez and Keith, 1978; Havlik et al., 1980; Hanson et al., 1989; Kupper et al., 2004; Moller et al., 1982; Russell et al., 1998). These twin studies established the framework for subsequent studies designed to discover the genetic variants underlying the reported heritability.

Several genetic association studies of ECG traits have been focused on genes that encode for proteins in voltage-gated ion channels(George, Jr. et al., 1995; Gouas et al., 2005b; Newton-Cheh et al., 2009; George, Jr., 2005; Lai et al., 1994; Gellens et al., 1992; Abriel and Kass, 2005). Voltage-gated ion channels are integral membrane proteins that allow the certain ions to pass across the cell membrane. Ion channels play an important role in generating electrical impulses thereby producing action potentials in myocytes (heart cells). Candidate gene studies have identified several genes throughout the genome that have an effect on heart depolarization and conduction measured by ECG traits. Several candidate gene and genome-wide association studies (GWAS) have identified polymorphisms in sodium and potassium channel genes that affect ECG measurements (Gouas et al., 2005b; Newton-Cheh et al., 2009; Schott et al., 1999; Wang et al., 1995a; Plant et al., 2006; Bezzina et al., 2006; Sotoodehnia et al., 2010; Baroudi et al., 2002;

Chambers et al., 2010; Smith et al., 2009b; Pfeufer et al., 2010; Rook et al., 1999; Baroudi et al., 2001). However, collectively these associations do not account for the all of the variability expected to be explained by genetic factors (<10% of the trait variability total). There have been few studies that have found associations in genes other than voltage-gated ion channels but the role these genes play in ECG trait variability is poorly understood.

Fibrinogen Levels

Fibrinogen is a glycoprotein made in the liver and is composed of three pairs of polypeptides: two Aα, two Bβ, and two γ (Hermans and McDonagh, 1982). Each polypeptide is encoded by genes (*FGA, FGB, FGG*) that make up the fibrinogen gene cluster on chromosome 4. All three polypeptides are covalently linked to each other by disulfide bonds in an ordered process at the N-terminals. Inflammatory markers, such as fibrinogen, are major risk factors for cardiovascular disease (Meade et al., 1986; Kannel et al., 1987; Green et al., 1994; Barasch et al., 1995; Lee et al., 1993; Heinrich et al., 1994). Circulating fibrinogen plays a major role in the coagulation pathway. The coagulation pathway is a series of events that usually occurs following vascular injury resulting in the production of fibrin clots (Figure 1.3). Circulating fibrinogen is converted to fibrin, the main protein component of a fibrin clot, in the last stage of the coagulation pathway (Figure 1.3) (Collet et al., 1993; Voetsch and Loscalzo.J, 2003; Blomback et al., 1978). Fibrin clots can be formed in the veins or the arteries resulting in venous or arterial thrombosis, respectively (Mackman, 2008). Arterial thrombosis is responsible for known CVD outcomes, myocardial infarction and stroke, and has a strong association with fibrinogen levels (Feinbloom and Bauer, 2005; Voetsch and Loscalzo.J, 2003).

Fibrinogen levels can be measured from blood plasma and elevated levels is indicative of an acute infection, cancer, coronary heart disease, myocardial infarction, recent trauma, or an inflammatory disease (Abrahamsen et al., 2008; David Ginsburg, 2005; Davi and Patrono, 2007; Fries et al., 2009). Normal plasma fibrinogen levels range from 1.5-4.0 g/L of blood. There are two common procedures used to measure levels of plasma fibrinogen: Clauss clotting assay and immunoassay(Mackie et al., 2003). The Clauss clotting assay is a functional assay that determines the time it takes for fibrin clot formation whereas the immunoassay measures the fibrinogen antigen rather than functional fibrinogen. The Clauss clotting method is preferred to diagnose diseases related to thrombosis, such as arterial thrombosis (or atherosclerosis), compared with the immunoassay.

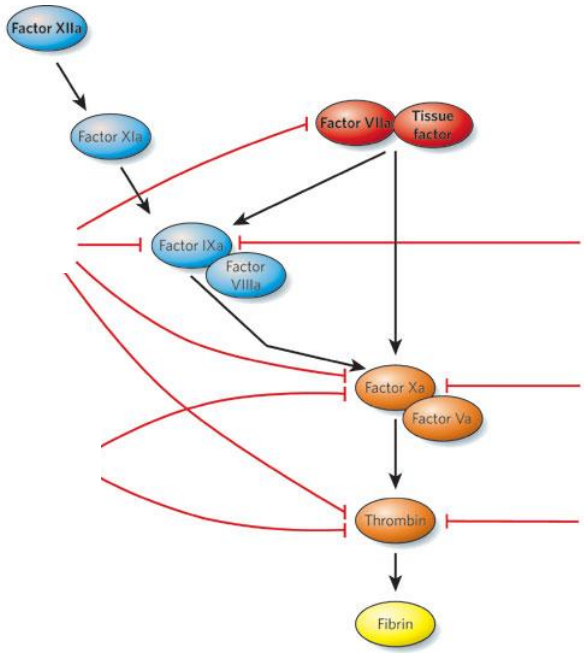


Figure 1.3. Coagulation Pathway. Adapted from Mackman et al. (Mackman, 2008).

Similar to ECG traits, CVD risk factors, such as sex (females), smoking, increased age, lack of exercise, and poor diet are associated with elevated fibrinogen levels

(Barasch et al., 1995; Lee et al., 1993; Heinrich et al., 1994). However, after adjusting for these risk factors, a significant amount of the variance in fibrinogen remains unexplained. Heritability estimates for fibrinogen suggest that remaining variance could be explained by genetics. Twin studies report up to 50% of the trait variability for fibrinogen attributable to genetic factors (Voetsch and Loscalzo.J, 2003). To date most of the genetics studies have been focused on genes that encode hemostatic factors in the beginning stages of the coagulation pathway (Figure 1.3). Other studies have found several associations between variation in fibrinogen genes (*FBA, FGB,* and *FGG*) and fibrinogen levels (Siegerink et al., 2009; Boekholdt et al., 2001; Folsom et al., 2001; Koch et al., 2008; Hamsten et al., 1987). There has not been strong evidence for genes that alter fibrinogen levels besides those involved in the coagulation pathway.

Hematological Traits

Unlike fibrinogen and other well-studied coagulation factors, little is known about other factors measured in the blood and their association with cardiovascular disease risk. As previously mentioned, individuals with CVD are also diagnosed with arterial thrombosis or atherosclerosis. Unlike venous thrombosis, arterial thrombosis generally requires the rupture of the atherosclerotic plaque, which happens over time due to accrual of lipids in the arterial wall (Figure 1.4). Several factors such as platelets, which are recruited at the site of the ruptured plaque (platelet aggregation), are involved in the pathogenesis of arterial thrombosis and may ultimately increase CVD risk.

Common hematological measurements include white blood cells (WBC), red blood cells (RBC), and platelets (PLT). These measurements are highly regulated in the body and can vary between individuals (Mackman, 2008; Soranzo et al., 2009). Hematological

measurements are derived from a complete blood count (CBC) exam. A CBC exam is a common test ordered by clinicians to measure WBC, RBC, and PLT counts. Several other factors such as the concentration of hemoglobin (Hb), hematocrit (Hct) and the volumes of red blood cells and platelets (MCV and MPV, respectively) can also be derived from a CBC exam.

There have been reports that hematological measurements have a genetic component. Heritability estimates for other traits such as Hb, WBC, and PLT range from 0.37-0.89 (Pennington et al., 2001; Evans et al., 1999) . Genetics also accounts for about 20-30% of the variance in platelet aggregation (Lane and Grant, 2000). There have been several candidate gene studies that report association throughout the genome with hematological traits. More recently there have also been a few GWAS with hematological traits (Meisinger et al., 2009; Nalls et al., 2011; Soranzo et al., 2009). One study identified three common loci that were associated with mean platelet volume (Meisinger et al., 2009). Another GWAS from the Haem Gen Consortium identified 22 loci that are associated with eight hematological traits in Europeans: Hb, RBC, WBC, PLT, MPV, MCV, mean corpuscular hemoglobin concentration (MCHC) and mean corpuscular content (MCH) (Soranzo et al., 2009).

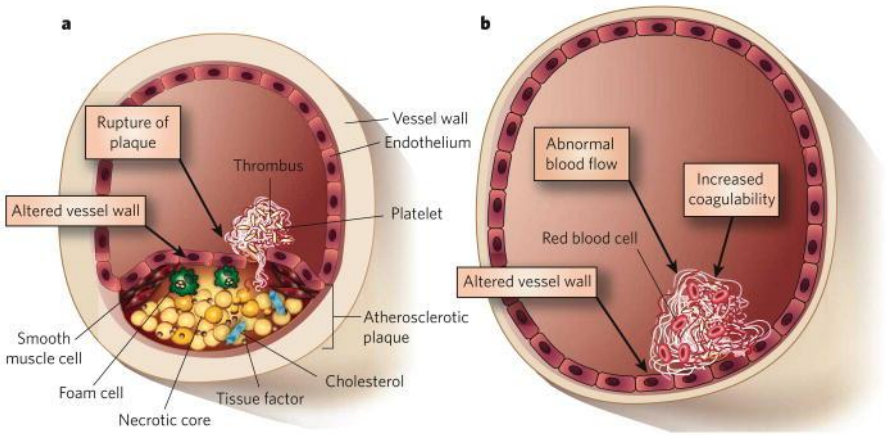


Figure 1.4 Comparison of an arterial clot (a) compared to venous clot (b). Obtained from Mackman et al (Mackman, 2008).

State of Cardiovascular Research in African Americans

Over the past 30 years mortality rates for CVD has been higher in African Americans compared to Europeans (Liao and Cooper, 1995; Sempos et al., 1988; Lloyd-Jones et al., 2009). Interestingly, the prevalence of CVD still remains higher in European Americans (Caucasians) compared to other ethnic groups in the United States, with the exception of southeastern states (Lloyd-Jones et al., 2009). Prevalence for CVD risk factors such as being hypertensive, diabetic, overweight, and having high levels of lipoprotein Lp (a) is higher in African Americans (Burt et al., 1995; Sorrentino et al., 1992; Lloyd-Jones et al., 2009). There has been considerable debate as to whether this disparity is due to differences in social risk factors such as education and income status, since these factors are associated with CVD risk regardless of race (Lloyd-Jones et al., 2009; Sempos et al., 1999). Another explanation could be differences in genetic risk factors for CVDs and other related traits across populations (Ioannidis et al., 2004).

Several genetic association studies on rare CVDs and other related phenotypes such as diabetes have been performed in African Americans (Adeyemo et al., 2009; Cheng et al., 2009; Helgadottir et al., 2006; Hooper et al., 1999; Norris et al., 2005; Plant et al., 2006; Van Norstrand et al., 2008). Genetic association studies have not been extensively performed in African Americans for ECG and fibrinogen/hematological traits. Furthermore, the extent to which ancestry affects genetic risk factors is poorly understood for ECG traits and fibrinogen/hematological measurements. There are several genetic differences observed across populations such as differences in allele frequency, linkage disequilibrium (LD) patterns, and the amount of genetic variation (Rosenberg et al., 2010). Linkage disequilibrium is the non-random association of alleles in the genome. Numerous factors can influence LD patterns such as mutations and recombination. It is believed that since African populations were one of the first human populations to exist; Africans are more likely to have gone through more recombination events and have accumulated more mutations overtime (Rosenberg et al., 2010; Teo et al., 2010; Teo et al., 2009). Thus, in general African descent populations have more genetic variation and lower levels of LD in the genome compared to Europeans.

Although genetic association studies have been performed in African descent populations, most studies have been performed in European populations. Furthermore, these studies use LD patterns in European populations (which have high LD) to select single nucleotide polymorphisms (SNPs) for genotyping. Identifying SNPs for genotyping using this approach is cost-effective but reduces the likelihood of identifying the true casual variant(s) in populations outside of Europeans. One solution is to use African descent populations to identify the true casual variant in the region where an

association has been identified in a European population. This is often achieved by densely genotyping or sequencing a region of interest in an African descent population, where more genetic variation exists. Another benefit to using diverse populations such as African Americans is to assess the extent to which results can be generalized to other populations. Generalization is defined here as a statically significant association with a consistent direction of effect observed across multiple populations. The generalization of results provides population-specific information about genetic risk variants, which can be useful in identifying the true casual variant for common diseases. Furthermore, determining whether an association generalizes is important for the clinical use of genetic risk variants for the diagnosis and treatment of common diseases for certain populations.

Disparities for quantitative traits associated with CVD such as cardiac arrhythmias are also observed with African Americans (Go et al., 2001). Genetic risk factors for CVDs such as sudden cardiac death (SCD) and long QT syndrome in African Americans have been suggested (Noseworthy and Newton-Cheh, 2008; Akylbekova et al., 2010; Plant et al., 2006; Van Norstrand et al., 2008). Multiple studies have suggested associations between SNPs in the sodium channel gene *SCN5A* with SCD. Interestingly, some of these studies have identified allelic associations that are population-specific and are not observed in other populations. For example, *SCN5A* S1103Y, a polymorphism unique to African descent populations, is associated with prolonged QT interval and SCD (Splawski et al., 2002; Burke et al., 2005).

With the increase of GWAS in recent years, there has been an increase of genetic associations identified with ECG traits. An association between another sodium channel gene, *SCN10A*, has been identified in several European and some Asian populations for

PR and QRS intervals. (Chambers et al., 2010; Holm et al., 2010; Pfeufer et al., 2010; Denny et al., 2010b). There is one GWAS in African Americans for PR interval (an ECG trait), but this study did not identify any novel associations nor did it generalize the *SCN10A* association (Smith et al., 2011). In addition to sodium channel genes, GWAS have identified several associations with other genes throughout the genome, such as *NOS1AP* (Arking et al., 2009; Newton-Cheh et al., 2009). One study successfully generalized the *NOS1AP* association in African Americans but only examined tagSNPs based from a European population; thus, the casual variant has yet to be identified (Smith et al., 2011). Further characterization of these associations, as well as the identification of novel genetic associations in African Americans, may explain the inter-individual trait variability observed for ECG traits across populations.

Similarly to genetic research for ECG traits, there is also a lack of genetic association studies for fibrinogen /hematological traits in African Americans. There have been both GWAS and candidate gene studies that have identified associations with white blood cell count (WBC) and fibrinogen levels in African Americans (Reiner et al., 2011; Wassel et al., 2010j; Reiner et al., 2006b; Nalls et al., 2011; Crosslin et al., 2011). For fibrinogen, European-identified variants in the fibrinogen gene cluster (*FBA*, *FGB* and *FGG*) and other inflammatory gene loci (i.e. *IL6R, IL1RN,* and *NLRP3T)* are also associated in African Americans (Wassel et al., 2010i). These associations have yet to be replicated in other African American populations. Two GWAS have recently been performed in African Americans with white blood cell count (Nalls et al., 2011; Reiner et al., 2011; Crosslin et al., 2011). Both studies confirm an association with SNPs in the *DARC* gene with white blood cell count and other WBC related phenotypes. Aside from

this association, collectively these studies have identified an additional twelve loci associated with WBC but unfortunately these associations do not overlap across studies. One genetic association study identified genetic variants for twelve hematological traits (Lo et al., 2011). In this study, pleiotropic effects were identified between *G6PD*, originally associated with malaria resistance, and RBC, Hct, and Hb in African Americans (Lo et al., 2011). Additionally, a novel association in both European and African Americans was identified for *TPM4* and platelet count. These associations have not been replicated in an independent African American population.

CHAPTER II

GENETIC VARIANTS ASSOCIATED WITH ELECTROCARDIOGRAPHIC TRAITS

A. Identification and Replication of *SCN5A* Variants with ECG traits in African Americans[1]

Introduction

Many factors are known to influence electrocardiograph (ECG) measurements, including underlying heart disease such as coronary heart disease or hypertension, drug therapy, body mass index, and age. These factors, however, do not explain the majority of the trait variance observed in ECG traits, and a large portion of the variance remains unknown(Zheng et al., 2001; Roberts and Day, 1985). Studies in twins and populations have reported that >35% of the variance observed for specific ECG traits is heritable(Akylbekova et al., 2009; Kao et al., 2009; Friedlander et al., 1998; Jouven et al., 1999; Dekker et al., 2006; George, Jr., 2005). Given the heritability of ECG traits, it is possible that common genetic variation in genes that modulate the electrical activity of the heart can explain a proportion of the unknown trait variance observed in the general population. Insights into such normal variability may in turn provide the starting point for predicting serious cardiovascular events such as drug-induced arrhythmias or Sudden Cardiac Death (SCD)(Kao et al., 2009; Friedlander et al., 1998; Jouven et al., 1999; Dekker et al., 2006). SCD accounts for 10-20% of all deaths in adults, and risk factors

[1] Adapted from: Jeff JM, Brown-Gentry K, Buxbaum SG, Sarpong DF, Taylor HA, George Jr AL., Roden DM, Crawford DC. *SCN5A* variation is associated with electrocardiogram traits in the Jackson Heart Study. *Circulation: Cardiovascular Genetics*. April 2011. 4(2):139-144. PMC3080430

include underlying heart disease and a family history of SCD(Friedlander et al., 1998; Jouven et al., 1999; Dekker et al., 2006).

The opening of sodium channels is responsible for initiating and propagating action potentials in cardiac and other excitable cells(George, Jr., 2005; Abriel and Kass, 2005; Bezzina et al., 1999; Gellens et al., 1992; George, Jr. et al., 1993; Gouas et al., 2007; Gouas et al., 2005b; Probst et al., 2009; Schott et al., 1999; Wang et al., 1995c). Sodium channels are encoded by nine pore-forming alpha subunit genes(George, Jr., 2005), and each channel consists of a single pore-forming α-subunit and a variable number of function-modifying β-subunits and other interacting proteins(Abriel and Kass, 2005). The $NA_v1.5$ sodium channel α subunit (encoded by *SCN5A*) is the predominant α-subunit expressed in cardiac muscle(George, Jr. et al., 1993; Gellens et al., 1992). Mutations in *SCN5A* have been identified in cardiac conduction disease, long QT syndrome, Brugada syndrome, and other life threatening arrhythmias(Wang et al., 1995c; Bezzina et al., 1999; Schott et al., 1999; Chen et al., 1998).

In addition to rare variants, other studies have implicated common variants in *SCN5A* with arrhythmia susceptibility or with variable ECG traits(Gouas et al., 2007; Gouas et al., 2005a; Probst et al., 2009; Bezzina et al., 2006; Tan et al., 2005). One non-synonymous *SCN5A* variant, rs7626962 (S1103Y, which is sometimes reported as S1102Y depending whether the reference sequence includes a common splice variant that eliminates a single residue)(Makielski et al., 2003), is common in African populations and rare in others. This SNP has been implicated in susceptibility to SCD, drug-induced arrhythmias, and SIDS in African American populations(Splawski et al., 2002; Plant et al., 2006; Van Norstrand et al., 2008). With this exception, however, there is very little

information on ion channel variation in African Americans. These lack of data cannot be underscored as epidemiologic studies have demonstrated that the risk factor burden for CVD differs across race/ethnicities, with African Americans typically having a greater burden compared with European and Mexican Americans(Ford et al., 2009). The rural south region of the US, which has the largest African American population, has a higher rate of CVD compared to other regions(Rosamond et al., 1998; Crook and Taylor, 2003).

Disparity of CVD in African Americans is an emergent concern to clinicians and epidemiologists. Admixed populations such as Hispanics and African Americans have different genetic backgrounds due to the mixture of distinct ancestral populations(Reich et al., 2001a; Lohmueller et al., 2008) and patterns of linkage disequilibrium (LD) compared with European-descent populations(Bush et al., 2009). Thus, it is possible that genetic determinants associated with ECG traits can vary depending on the ancestral populations. Given these potential differences, we genotyped 72 common single nucleotide polymorphisms (SNPs) in the *SCN5A* gene in 4,558 African Americans from the Jackson Heart Study (JHS) to test for an association with nine ECG traits: P, PR, QT, QTc and QRS durations, heart rate, and the P, QRS and T axes. We identified 14 significant associations at $p<1.0x10^{-4}$, 13 of which have not been described in populations of European-descent. Overall, *SCN5A* variations reported here explain up to 2% of the variation in ECG traits. For replication, we genotyped five of the 14 significant *SCN5A* associated SNPs in the Third National Health and Nutrition Examination Survey (NHANES III). This study represents an important first step in the identification and characterization of genetic variants associated with ECG traits in African-descent populations.

Methods

Population characteristics and ECG measurements

The Jackson Heart Study (JHS) was developed to help resolve the disparity of cardiovascular disease among African Americans(Wilson et al., 2005). The JHS is a longitudinal study established in 2000 to characterize the determinants of CVD in 5,301 African Americans ascertained in Jackson, Mississippi(Campbell-Jenkins et al., 2009). For the purpose of genetic analysis, consenting family members age 21 or older were ascertained(Wilson et al., 2005). The participants' ages ranged from 21-85 years old at the time of ascertainment (Appendix A), and 62% of the participants were female. Each participant received a clinical examination and interview on CVD status and other environmental factors during enrollment. ECG measurements were collected during the clinical examination using the Marquette MAC/PC digital electrocardiograph(Carpenter et al., 2004). All measurements were documented and sent by phone to the Electrocardiographic Reading Center (ECGRC) in Minnesota. The Minnesota Code Modular ECG Analysis System (MC-MEANS) computer program was used by the ECGRC to generate representative averaged measurements of ECG waves simultaneously over all leads. Population demographics and summary statistics of ECG measurements are given in Table 2.1 for JHS participant.

All replication analyses were performed in non-Hispanic blacks from the Third National Health and Nutrition Examination Survey (NHANES III). Ascertainment of NHANES III and method of DNA collection have been previously described (Crawford et al., 2006; Chang et al., 2009; Steinberg et al., 1997). The National Health and

Nutritional Examination Surveys are cross-sectional surveys conducted across the United States by the National Center for Health Statistics (NCHS) at the Centers for Disease Control and Prevention (CDC). NHANES III was conducted between 1988-1990 (phase 1) and 1991-1994 (phase 2) (1996; 2004) as a complex survey design that over-sampled minorities, the young, and the elderly. All NHANES have interviews that collect demographic, socioeconomic, dietary, and health-related data. Also, all NHANES study participants undergo a detailed medical examination at a central location known as the Mobile Examination Center (MEC). Beginning with phase 2 of NHANES III, DNA samples were collected from study participants aged 12 years and older.

Electrocardiograms (ECGs) were recorded on men and women in the mobile examination center (MEC) using a standard 12-lead resting ECG. ECGs were recorded using the Marquette MAC 12 (Marquette Medical Systems, Inc, Milwaukee, Wisconsin (U.S. DHHS, 1996). NHANES III 12-lead ECG data were recorded with eight independent components of the 12 standard leads simultaneously. ECG data were also sampled at 250 samples per second per channel; giving the availability of multiple simultaneous ECG leads for analysis. ECGs were only performed on individuals 40 years of age or greater and NHANES III was the last NHANES that collected ECG data. Population demographics and summary statistics of ECG measurements are given in Table 2.3.

Genotyping and SNP selection

For the Jackson Heart Study, blood samples were collected from consented individuals during enrollment for future genetic analysis. DNA was isolated from blood samples and genotyped for 72 single nucleotide polymorphisms (SNPs) using the

Sequenom iPlex Gold assay on the MassARRAY platform (San Diego, CA). SNPs were selected based on the linkage disequilibrium patterns in African Americans (Bush et al., 2009). The location of the 72 genotyped SNPs for JHS participants, minor allele, minor allele frequencies, Hardy-Weinberg p-values, and genotyping call rates are given in Appendix B.

For NHANES III participants, DNA was extracted from crude cell lysates from lymphoblastoid cell lines established for NHANES III participants aged 12 and over as part of Genetic NHANES (Steinberg et al., 1997). We chose 12 SNPs that were significant at p 1.0×10^{-4} in Jackson Heart Study for genotyping in NHANES III. To date five SNPs have been successfully genotyped for this study, the others have yet to be genotyped. All genotyping was performed in the Center for Human Genetics Research DNA Resources Core and the laboratory of Dr. Jonathan Haines using either Sequenom's iPLEX Gold assay on the MassARRAY platform (San Diego, CA) or Illumina's BeadXpress. All genotype data reported here passed CDC quality control metrics and are available for secondary analysis through CDC.

Statistical Methods

The JHS in this analysis is comprised of 3,071 unrelated participants and 1,487 related participants from 263 families selected for genotyping. Unrelated individuals from 263 pedigrees with two or more successfully genotyped relatives were randomly selected and analyzed. Individuals selected by this random algorithm had to have all phenotype information and have at least 95% genotyping efficiency for targeted SNPs. In two families, only one participant was genotyped; as a result, these participants were

always selected. We repeated the random selection several times (five times), and data suggest that our results are robust to the random selection process (data not shown).

To determine if ECG traits were strongly correlated (i.e., redundant), we calculated pair-wise correlations for 11 ECG traits (Appendix C). Prolonged QRS duration (>120 m/sec) is an indication of congestive heart failure. Patients with abnormal QRS duration measurements (>120 m/sec) were excluded from the QRS analysis (n = 169; Appendix D); however, these patients were included in subsequent analysis with the other ECG traits. The exclusion of these patients in the analysis of QRS duration or other ECG traits yielded similar results compared with the inclusion of these patients (data not shown) due to the relatively small sample size of patients with prolonged QRS duration compared with the overall sample size. Of all ECG traits, T and QT durations and PR duration in leads II and VI were highly correlated (r^2=0.93 and 0.87 respectively, Appendix C). To minimize redundancy, we chose to analyze QT duration instead of both QT and T durations and PR in lead II instead of both PR in leads II and VI. Although modest, a correlation was also observed for P wave and PR durations, QTc and QT duration, as well as QTc and heart rate (r^2 >0.40, Appendix C). All of these moderately correlated traits were retained for subsequent analyses.

For the analysis, we excluded SNPs with a minor allele frequency (MAF) <0.05, genotyping efficiency <95% or Hardy Weinberg equilibrium (HWE) p-value <0.0001. Using these criteria, seven SNPs were excluded from the analysis. We also excluded 280 DNA samples due to poor genotyping efficiency, resulting in a total study population of 3,054 participants.

Global admixture estimates were previously calculated using genome-wide data (Deo et al., 2009) to determine the amount European ancestry using ANCESTRYMAP Version 2.0(Patterson et al., 2004a) and expressed as "mean percent European ancestry" based on a probabilistic term representing the genome-wide mean percentage of European ancestry as determined from markers on the autosomes. European ancestry was associated with durations (P, QTc, and QRS) and T axes ($p<0.05$).

We calculated skewness and kurtosis and performed formal tests of normality for all ECG traits. For QRS duration, the skewness (0.34) and kurtosis (2.4) suggested that this ECG trait is normally distributed. However, QRS duration failed a formal test of normality while all other ECG traits followed a normal distribution. Based on these results, we performed tests of association for QRS duration untransformed and log transformed (data not shown). Because the results did not differ between the untransformed and transformed analyses, we chose to present the untransformed results for ease of interpretation.

Using linear regression, we performed single SNP tests of association assuming an additive genetic model for 65 SNPs that passed QC in the unrelated sample with nine ECG traits using PLINK(Purcell et al., 2007). Analyses were performed unadjusted and adjusted for age, sex, and European ancestry, and results were plotted using Synthesis-View (Pendergrass et al., 2010a). To account for multiple testing, we employed a significance threshold of $p<1.0 \times 10^{-4}$. We also performed unadjusted family-based tests of association for all ECG traits using the QFAM procedure from PLINK(Purcell et al., 2007). This method uses traditional linear regression but uses permutation to correct for family structure. We employed this method with all genotyped samples that passed

quality control [5 families (n=19) were removed due to Mendelian errors] and tested for an association with each SNP for all ECG traits. The results from this analysis are consistent with results from the unrelated samples (data not shown). To determine the amount of the variance in ECG traits explained by each SNP, R^2 was computed using STATA version 10. Linkage disequilibrium (r^2) was calculated using Haploview (Barrett et al., 2005).

All replication analyses in NHANES III were performed using the Statistical Analysis Software (SAS v.9.2; SAS Institute, Cary, NC) either locally or via the Analytic Data Research by Email (ANDRE) portal of the CDC Research Data Center (RDC) in Hyattsville, MD. All traits followed a normal distribution and descriptive statistics are reported in Table 2.3. Participants with QRS duration >120 m/sec were excluded from all analysis. Using standard linear regression, assuming an additive model, we tested each SNP for an association with ECG traits: P duration, QRS duration, QT interval and heart rate. We did not test *SCN5A* variants with heart rate and QT interval since these SNPs were not associated with these traits in the original analysis in JHS. All tests were limited to self-identified non-Hispanic blacks and adjusted for age and sex. We did not formally adjust for genetic ancestry; however all analysis were stratified by race/ethnicity. We declared significance at $p<0.05$ uncorrected for multiple testing. Power calculations were performed using QUANTO (Gauderman and Morrison, 2006)using effect sizes from the original JHS study. Fst was calculated between JHS and NHANES III NHB using the Platform for the Analysis, Translation, and Organization of large-scale data (PLATO) (Grady et al., 2010).

Results

ECG interval durations

The P, PR, QRS, QT and QTc durations were tested for an association with 65 SNPs in *SCN5A* that passed quality control measures. Overall there were 14 significant associations ($p<1.0 \times 10^{-4}$) amongst 65 SNPs for P, PR, QRS and QT durations. PR duration had the most significant SNP associations, with eight SNPs significantly associated with decreased PR duration: nonsynonymous rs7626962 (S1103Y) and seven intronic SNPs (Table 2.2). Three intronic SNPs were associated with increased PR duration but not with any other ECG trait. Four SNPs were associated with decreased P wave duration, one with decreased QRS duration. A single intronic SNP was associated with increased QRS duration, but no other ECG trait. Unique to QT duration, rs9311195 was associated with decreased QT duration (Table 2.2). There were no significant associations observed with QTc duration at $p<1.0 \times 10^{-4}$ (Appendix E).

Heart Rate

Of the *SCN5A* SNPs tested for an association with heart rate, none of the associations survived the significance threshold of $p<1.0 \times 10^{-4}$. Two SNPs that were also associated with decreased P and PR duration trended towards significance with increased heart rate: SNPs rs7629265 ($\beta= 0.97$, $p= 0.04$) and rs7626962 ($\beta= 0.91$, $p= 0.05$) (Appendix E). These two SNPs are in high linkage disequilibrium (LD) with one another ($r^2=0.87$); thus, these associations are not independent.

ECG axes

In addition to ECG durations, the P, QRS and T axes were also tested for an association with the same 65 SNPs. There were no significant associations with our SNPs and ECG axes measurements using a significance threshold of $p<1.0x10^{-4}$ (Appendix E). One SNP that was associated with decreased PR duration trended towards significance with P axis ($\beta= 1.95$, $p= 0.05$).

SNPs associated with multiple ECG traits

Intronic SNP rs3922844 was associated with increased P, PR, and QRS. SNPs rs7374138, rs7626962, rs76229265, rs7637849, rs7627552, and rs6763048 consistently trended towards decreased PR, QRS and P wave durations (Figure 2.1). Although most SNPs did not survive our significance threshold for QRS duration, SNPs that were associated with PR and P wave duration trended towards significance with QRS duration compared with other ECG traits.

Table 2.1 Study population demographics and ECG trait descriptive statistics for the unrelated (n= 3,054) JHS participants.

Variable	Mean or %	Standard Deviation (minimum, maximum)
Age (yrs)	56.5	±11.73 (21, 85)
% Female	62	-
% Cardiovascular disease	11	-
P duration (msec)	118.5	±13.08 (80, 170)
PR duration (msec)	171.6	±33.02 (0,338)
QRS duration (msec)	92.3	±10.12 (64, 120)
QT duration (msec)	414.7	±27.64 (290, 580)
QTc duration (msec)	426.4	±27.64 (334, 594)
P axis (degrees)	48	±21.06 (-136, 151)
QRS axis (degrees)	16.9	±31.18 (-137, 157)
T axis (degrees)	30.6	±40.03 (-179, 179)
Heart rate (beats/min)	64.6	±10.72 (30, 118)

Table 2.2 Beta coefficients and p-values for SNPs significantly associated with at least one ECG trait in unrelated JHS participants. Linear regression adjusting for age, sex, and European ancestry was performed for 65 SNPs assuming an additive model for nine quantitative ECG traits (for QRS duration only patients with <120 msec were included, n= 2,878). Fifteen significant SNPs that met our significance threshold were associated with at least one ECG trait and are listed above. In bold are significant associations at $p< 1.0x10^{-4}$.

SNP	Location	P duration		QRS duration		PR lead II		QT duration	
		β (95% CI)	p-value	β (95% CI)	p-value	β (95% CI)	p-value	β (95% CI)	p-value
rs11129796	Intron	-0.81 (-1.6, 0.7)	1.8E-1	-1.1 (-1.6, -0.3)	2.5E-3	-3.5 (-5.0, -0.2)	**2.4E-4**	-1.4 (-4.1, 0.3)	2.3E-1
rs13084981	Intron	1.6 (-0.0, 2.9)	2.7E-2	1.4 (0.4, 2.2)	**9.4E-4**	3.5 (0.8, 6.5)	**1.1E-2**	-0.61 (-3.1, 2.1)	6.5E-1
rs3922844	Intron	1.5 (0.4, 2.0)	**6.3E-4**	1.3 (0.8, 1.8)	**7.5E-8**	4.7 (3.1, 6.4)	**3.9E-9**	-0.68 (-2.3, 0.8)	3.9E-1
rs6763048	Intron	-1.7 (-2.5, 0.6)	**6.6E-4**	-0.91 (-1.3, -0.2)	1.4E-3	-4.3 (-6.1, -2.3)	**3.5E-6**	-0.54 (-2.8, 1.4)	5.5E-1
rs6768664	Intron	0.60 (-0.2, 1.5)	1.8E-1	0.67 (0.2, 1.2)	9.5E-3	3.0 (1.5, 4.9)	**2.3E-4**	2.4 (0.46, 3.7)	4.0E-3
rs7373102	Intron	0.81 (-0.2, 1.6)	8.1E-2	0.71 (0.13, 1.2)	1.1E-2	2.8 (1.5, 5.1)	**1.4E-4**	2.1 (-0.1, 3.3)	2.1E-2
rs7374138	Intron	-1.5 (-2.3, -0.4)	3.6E-3	-1.3 (-1.8, -0.6)	**6.2E-6**	-4.0 (-6.0, -2.1)	**2.4E-5**	-0.73 (-3.5, 0.1)	4.4E-1
rs7374605	Intron	1.1 (-0.4,1.6)	3.6E-2	0.76 (0.2, 1.4)	9.9E-3	4.0 (2.0, 6.0)	**3.3E-5**	-1.4 (-3.0, 0.7)	1.3E-1
rs7626962	Ser->Tyr	-3.3 (-4.9,-1.8)	**2.5E-5**	-1.4 (-2.2, -0.4)	2.3E-3	-6.9 (-9.9, -4.0)	**2.5E-6**	-0.74 (-3.1, 2.6)	6.1E-1
rs7627552	Intron	-2.2 (-2.9,-0.8)	**5.8E-5**	-1.0 (-1.7, -0.5)	1.5E-3	-6.4 (-3.7, -5.5)	**4.6E-10**	-2.1 (3.5, 0.5)	3.6E-2
rs7629265	Intron	-3.2	**7.5E-5**	-1.5	1.8E-3	-7.8	**2.4E-7**	-1.3	3.8E-1

		(-4.6,-1.6)		(-2.4, -0.6)		(-10, -4.4)		(-3.2, 2.5)	
rs7637849	Intron	-1.7 (-2.9, -0.9)	1.1E-3	-0.76 (-1.2, 0.0)	1.2E-2	-4.2 (-6.1, -2.1)	**2.3E-5**	-0.43 (-1.9, 1.9)	6.6E-1
rs9311195	Intron	0.29 (-0.5, 1.7)	6.2E-1	0.21 (-0.6, 0.7)	5.3E-1	-0.58 (-2.5, 2.0)	6.0E-1	-3.5 (-5.0, -0.7)	**1.6E-4**
rs9832586	Intron	-1.3 (-2.6,0.59)	1.1E-1	-0.58 (-1.5, 0.3)	2.1E-1	-5.2 (-7.8, -1.4)	**7.3E-4**	-2.9 (-7.4, -1.5)	4.8E-2

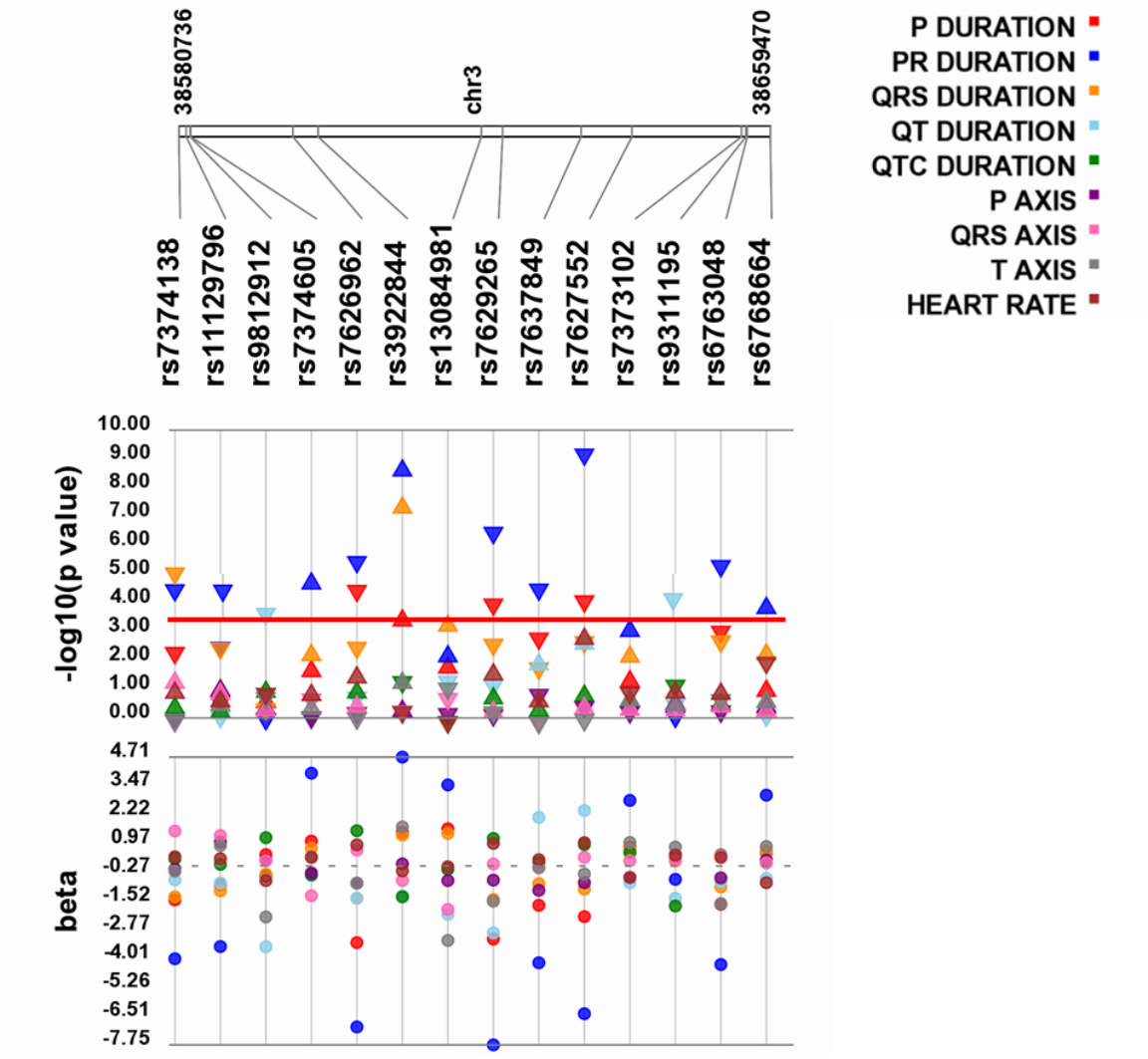


Figure 2.1 Significant associations across ECG traits. Each SNP was tested for an association with each ECG trait assuming an additive genetic model adjusted for age, sex, and European ancestry. P-values are $-\log_{10}$ transformed along the y-axis and corresponding location for each SNP is located on the x-axis. Each point represents a p-value for each trait indicated by color (see legend). The direction of the arrows corresponds to the direction of the beta coefficient. The exact beta coefficients are reported on the bottom panel. The significance threshold is indicated by the red line at $p= 1.0x10^{-4}$.

Replication in NHANES III non-Hispanic blacks

Using non-Hispanic blacks from NHANES III, we sought to replicate significant *SCN5A* associations. ECG trait measurements between NHANES III non-Hispanic blacks and JHS African Americans are described in Table 2.3. Mean age, P duration, PR interval, and QRS duration differed between NHANES III NHB and JHS African Americans (Table 2.3). On average, NHANES III participants were younger compared with JHS study individuals (53.9, 56.5). Average PR interval was shorter and QRS duration was longer, consistent with the lower average age of NHANES III participants compared with JHS individuals (Ramirez et al., 2011).

Although both JHS individuals and NHANES III NHB participants are self-described African Americans, it is possible that the genetic ancestry of the two populations differ due to admixture, which is known to varying widely by geography (Rosenberg et al., 2010; Hammer et al., 2011; Tishkoff et al., 2009). To examine this, we examined the coded allele frequencies of the five SNPs in NHANES III NHB with JHS (Table 2.4). Two SNPs had higher coded allele frequencies in JHS compared with NHANES III NHB: rs7374138 (0.23 versus 0.15) and rs11129796 (0.15 and 0.08). One SNP, rs7637849, had a lower coded allele frequency in JHS compared with NHANES III (0.79 versus 0.86). We then calculated Fst to quantitate possible population differences. The Fst values for these five SNPs ranged from <0.0001 to 0.019; therefore, there was no evidence for significant population differentiation at these loci (Table 2.4).

We tested five previously identified SNPs from the JHS for associations with ECG traits in NHANES III non-Hispanic blacks. A total of five tests of association

involving three ECG traits (P duration, PR interval, and QRS duration), equivalent to fifteen tests total, was performed. None of the SNP-trait associations identified in JHS replicated in NHANES III non-Hispanic blacks at $p<0.05$ (Table 2.5). Regardless of significance, nine out of 15 (60%) tests of association resulted have a consistent direction of effect with the original study in JHS.

Table 2.3 Population characteristics comparisons. Mean and Standard deviation was calculated ECG traits tested and age for all NHANES III non-Hispanic blacks compared to JHS African American. Student's T-test was performed to assess the difference in means for each trait between JHS and non-Hispanic blacks.

Trait	JHS African Americans (n= 3,054)		NHANES III Non-Hispanic Blacks (n=530)	
	Mean	**SD**	**Mean**	**SD**
Age (yrs)	56.5	±11.73	53.9	±11.61
P Duration (msec)	118.5	±13.08	113.1	±14.57
PR Interval (msec)	171.6	±33.02	164.6	±25.56
QRS duration (msec)	92.3	±10.12	95.52	±10.90

Table 2.4 Comparison of minor allele frequencies *SCN5A* variants between NHANES III non-Hispanic blacks and Jackson Heart Study African Americans. To compare allele frequency differences between NHANES III NHB and JHS, we calculated Fst for all *SCN5A* SNPs. For each SNP, the alleles, the minor allele frequencies (MAF) for each study, and F_{ST} are given.

SNP	Major Allele/Minor Allele	MAF Jackson Heart Study (n~3,054)	MAF Non-Hispanic blacks (n~530)	F_{ST}
rs7374138	C/G	0.23	0.15	<0.0001
rs7637849	G/A	0.79	0.86	0.014
rs11129796	C/T	0.15	0.08	0.019
rs7629265	G/T	0.08	0.09	<0.0001
rs6768664	A/G	0.36	0.36	<0.0001

Table 2.5 Replication of *SCN5A* variants in non-Hispanic blacks from NHANES III. Linear regression adjusted for age and sex was performed for five JHS identified genetic variants and three ECG traits in non-Hispanic blacks. All genetic variants were coded assuming an additive genetic model, and coding was consistent between studies.

SNP	Phenotype	NHANES III Non-Hispanic Blacks (n =552)		JHS African Americans (n = 3,054)	
		Effect size	P-value	Effect size	P-value
rs7637849	P Duration	0.94	0.41	-1.70	1.1E-3
rs7374138	P Duration	-1.37	0.24	-1.50	3.6E-3
rs11129796	P Duration	-0.64	0.66	-0.81	0.18
rs7629265	P Duration	-1.51	0.38	-3.20	7.5E-5
rs6768664	P Duration	-1.09	0.23	0.60	0.18
rs7637849	PR Interval	0.11	0.96	-4.20	2.3E-5
rs7374138	PR Interval	-1.65	0.40	-4.0	2.4E-5
rs11129796	PR Interval	0.12	0.96	-3.50	2.4E-4
rs7629265	PR Interval	-4.13	0.17	-7.80	2.4E-7
rs6768664	PR Interval	0.16	0.92	3.00	2.4E-4
rs7637849	QRS Duration	-0.39	0.66	-0.76	0.012
rs7374138	QRS Duration	-0.69	0.41	-1.30	6.2E-6
rs11129796	QRS Duration	0.09	0.94	-1.10	0.0025
rs7629265	QRS Duration	1.07	0.82	-1.50	1.8E-3
rs6768664	QRS Duration	1.93	0.73	0.67	9.5E-3

Discussion

Using a candidate gene approach, we sought to identify genetic variations within *SCN5A* associated with ECG measurements in African Americans from the Jackson Heart Study. We were able to detect both previously known as well as novel genetic associations with four ECG measurements (Table 2.2). Notably, to our knowledge, we are the first to report an association with the genetic variant S1103Y and atrial ECG traits, PR and P wave durations. We were also able to detect novel associations between intronic rs3922844 and QRS, P wave, and PR durations. SNP rs3922844 explains approximately 2% of the variability in these traits. Consistent with previous studies, no other SNP explained greater than 2% of variance for any ECG trait(Smith et al., 2009a). We also attempted to replicate five SNPs originally identified as associated with ECG traits in JHS participants in NHANES III non-Hispanic blacks; none replicated (Table 2.5).

To minimize redundancy in our JHS results, we calculated linkage disequilibrium (LD) among SNPs in the unrelated participants (Appendix F). Of the 14 significant associations, 13 SNPs represent independent associations. Correlated SNPs rs7626962/rs7627552 are in high LD with each other (r^2= 0.87), and each likely represents the same effect. SNP pairs rs6793245/rs3935472 and rs9833086/rs4130467 are also in LD; however, they were not significantly associated with any ECG trait.

Several genome-wide association studies (GWAS) have been published for the ECG traits in populations of non-African descent. For example, two studies report associations within *SCN5A* and the QT interval (Newton-Cheh et al., 2007b; Newton-

Cheh et al., 2009; Newton-Cheh et al., 2007a; Newton-Cheh and Shah, 2007). A recent GWAS in the isolated Kosrae population from the Federated States of Micronesia identified an association between rs7638909, a SNP located in intron 27 of *SCN5A*, and the PR interval, P wave duration and PR segment(Smith et al., 2009c). Other GWAS studies on ECG traits report an association with rs7638909 (or correlated SNPs rs12053903, rs1805126 and rs1805124) with the QT interval, PR interval, and QRS duration in Icelandic and European populations(Holm et al., 2010; Newton-Cheh et al., 2009). We tested for these associations and all ECG traits in the JHS but failed to replicate these findings. The lack of replication across studies is perhaps not surprising given the different linkage disequilibrium patterns and allele frequencies in these populations. The SNPs we choose for genotyping for this study are common in African Americans such as rs7626962 and rs7629265 but are monomorphic or rare in other populations. GWAS for ECG traits in African-descent populations has yet to be reported in the literature.

Unlike previous studies, we were unable to detect an association between S1103Y and ventricular related traits QRS, T, and QT durations in JHS at our significance threshold. Given previous studies, we had expected to detect a strong association with S1103Y (rs7626962) and our ECG traits. The effect of 1103Y and QTc duration is consistent with previous studies ($\beta = 1.53$, $p = 0.20$; Appendix E), but given the minor allele frequency of this SNP (0.08), our study was underpowered to detect this association (20%). In our study population, there were only 20 homozygotes for the 1103Y allele; thus, it is not surprising that mean QTc is not associated with S1103Y genotype (data not shown). S1103Y was associated with P wave and PR durations

($p<1.0x10^{-5}$), but there was no other association at $p<1.0x10^{-4}$ with other ECG traits. A significant effect was also observed with rs7629265, which is in LD with S1103Y (r^2= 0.87). Intronic rs7629265 ranked higher than S1103Y in four ECG traits: PR duration, QRS duration, QT duration, and heart rate (Table 2, Appendix E). Because of the strong LD, the association with intronic rs7629265 could be a capturing the effect of S1103Y.

As opposed to mutations, the effects of common variation are less clear. Our most significant finding in JHS, rs7627552, is in LD with rs7626962 (S1103Y), a nonsynonymous missense mutation that changes the amino acid serine to tyrosine. Studies of this variant in heterologous expression systems have reported differences between variant and wild-type channels especially under stress conditions such as low pH(Plant et al., 2006; Splawski et al., 2002). The other significant *SCN5A* SNPs from the present study are all located in the 3' region of the gene. While none of them have obvious function, it is noteworthy that this gene region is highly conserved with mouse. Also, it may be that these associated SNPs are in linkage disequilibrium with the "causal" or functional SNPs not genotyped in this study. However, examination of the Yoruba International HapMap data in this region (100kb upstream and downstream of rs7627552) revealed only one SNP with minor allele frequency >1% in LD at $r^2>0.20$ in this region (rs7629265).

We were unable to replicate *SCN5A* associations identified in JHS in non-Hispanic blacks from NHANES III. The lack of replication is most likely due to power given we had 552 participants with ECG trait data available for analysis. To further explore this possibility, we calculated the power to detect the originally reported

association assuming the effect size reported in JHS, average/SD for each trait, the coded allele frequency observed in NHANES III NHB, and the sample size observed in NHANES III NHB. At a liberal significance threshold of $p<0.05$, we had 20%, 15%, and 20% power to detect the original JHS associations for P duration, PR interval, and QRS duration, respectively. Thus, we were under-powered to replicate JHS associations in NHANES III NHB. Large samples in diverse populations are sorely needed to replicate, establish, and elucidate genetic associations relevant to normal cardiac conduction.

Genetic variation in the normal electrical activity of the heart, assessed by the ECG, may provide a starting point for studies of genetic susceptibility to serious arrhythmias, such as SCD during myocardial infarction or drug therapy. We identified several novel associations between *SCN5A* common genetic variants and the ECG traits in JHS African Americans, and we report here the direction and magnitude of effect for all tests of association. Of the 14 significant associations, rs7627552, which is in LD with a common missense mutation (S1103Y) observed in African Americans previously associated with arrhythmia susceptibility, was the SNP most strongly associated with ECG traits P wave and PR durations. We were also able to identify 13 novel associations, one of which (rs3922844) explains as much as 2% of the variance in P wave, QRS and PR durations. Collectively, these data suggest that multiple SNPs rather than one (S1103Y) in *SCN5A* have an effect on ECG traits. The results of this study may offer insight relevant to future genetic studies of cardiac diseases in African-descent populations**.**

B. Using Genome-Wide Association Studies (GWAS) to Identify Novel Variants and Generalize Previously-Identified Variants Associated with ECG traits in African Americans[2]

Introduction

Candidate gene studies have identified several genes that affect cardiac depolarization and conduction measured by ECG traits. One set of candidates is genes that encode voltage gated potassium and sodium channels, which also play an important role in the pathophysiology of long QT syndrome (George, Jr. et al., 1995; Wang et al., 1995b; Wang et al., 1995a; Lai et al., 1994; Splawski et al., 2002; Splawski et al., 1997; Wang et al., 1996). These associations have been confirmed by molecular genetic studies, and, in the past five years, GWAS have also identified these associations with various ECG phenotypes in Europeans, Korsae, and Indian Asian populations (Newton-Cheh et al., 2009; Smith et al., 2009b; Chambers et al., 2010). These GWAS have not only implicated common variation in congenital long QT syndrome genes but also have identified associations between genes not previously implicated in cardiac electrophysiology and ECG traits, such as *NOS1AP* with QT interval (Arking et al., 2006; Post et al., 2007). Genome-wide association studies of variation in ECG traits have also revealed pleiotropy: for example, the cardiac sodium channel gene *SCN5A,* has been associated with variation across multiple ECG traits (Smith et al., 2009b; Holm et al., 2010; Holm et al., 2010; Newton-Cheh et al., 2007b; Marroni et al., 2009; Newton-Cheh et al., 2007b; Pfeufer et al., 2010). Variants in *SCN10A,* a gene previously known to play

[2]Jeff JM, Ritchie MD, Ramirez AH, Denny J, Kho AN, Hayes MG, Armstrong L, Basford M, Wolf WA, Pacheco JA, Chisholm RL, Roden DM, Crawford DC Generalization of Variants Identified by Genome-wide Association Studies for ECG Traits in African Americans. (In preparation)

a role in the action potentials in nociceptive nerve fibers(Akopian et al., 1999; Zimmermann et al., 2007; Abrahamsen et al., 2008), have been identified by GWAS for an association with PR and QRS interval (Chambers et al., 2010; Pfeufer et al., 2010; Holm et al., 2010; Sotoodehnia et al., 2010; Denny et al., 2010b).

As of October 2011, there were 127 unique trait-SNP associations reported in the National Human Genome Research Institute's (NHGRI) GWAS catalog for electrocardiographic (ECG) traits (http://www.genome.gov/gwastudies/). Most of these studies were limited to European (118 associations) or Asian (9 associations) descent populations, and it is unclear if these GWAS-identified variants generalize across populations, particularly the extent to which these associations generalize to African Americans. There is considerable variability in ECG traits across populations. African Americans have a shorter QRS duration, QTc interval, and a longer PR interval compared to Europeans (Ramirez et al., 2011). From a genetic standpoint there are also significant differences in allele frequencies and linkage disequilibrium patterns between Africans and Europeans. For example, rs7626962 (S1103Y) in *SCN5A* is rare in European and Asian descent populations but is common in African Americans (Jeff et al., 2011b; Splawski et al., 2002) and has been associated with arrhythmia susceptibility.

Assessing the utility of electronic medical records (EMR) systems coupled to DNA repositories as a tool for genome science is one of the primary objectives for the National Human Genome Research Institute's electronic MEdical Records and GEnomics (eMERGE) Network (Roden et al., 2008). Studies from eMERGE, including those on PR duration, have demonstrated that EMR-based genetic studies replicate

existing findings and discover new ones (Denny et al., 2011; Denny et al., 2010b; Crosslin et al., 2011; Turner et al., 2011). Here we performed a GWAS of ECG measurements to identify novel genetic associations and to describe the extent to which previous associations generalize in African Americans within the eMERGE network. We also examine factors that contribute to non-generalization of associations in our African American study population.

Methods

Study population

African American subjects were collected from the Vanderbilt or Northwestern University biobanks. The Vanderbilt biobank, BioVU, is a collection of DNA samples from discarded blood samples collected for routine clinical care linked to de-identified electronic medical records (EMRs) (McCarty et al., 2011). The Northwestern biobank, NUgene, combines DNA samples from consented participants with enrollment questionnaire and longitudinal data from the EMR (McCarty et al., 2011). Both biobanks were approved by Institutional Review Boards at their respective sites.

Study population demographics and characteristics are described in Table 3.6. Individuals with a normal ECG without evidence of cardiac disease before or within one month following the ECG, without concurrent use of medications that interfere with QRS duration, and who did not have abnormal electrolyte values at the time of the ECG were included. Using natural language processing combined with EMR (structured database queries) to query (specific to BioVU), we excluded individuals with any of the following before or within one month of the ECG: indication of heart failure, arrhythmia,

cardiomyopathy, cardiac conduction defect, or myocardial ischemia/infarct based on clinical notes, unstructured text, billing codes, and labs. BioVU subjects were African American as indicated by observer reported ancestry, which is highly concordant with genetic ancestry (Dumitrescu et al., 2010). African American ancestry was self-reported for Northwestern subjects (McCarty et al., 2011), which is also known to be highly concordant with genetic ancestry. All ECGs had normal Bazett's corrected QT intervals (<450ms), heart rates (between 50-100 bpm), and QRS duration (65-120 ms).

Genotyping

Genotyping for the eMERGE network was performed by the Center for Inherited Disease Research (CIDR) and the Broad Institute. All individuals that met the inclusion criteria (n = 501) were genotyped for >1.1 million SNPs using the Illumina 1M BeadChip at the Broad Institute. Data were cleaned by the eMERGE QC pipeline (Zuvich et al., 2011). Individuals with cryptic relatedness, ancestry inconsistent with observer- or self-reported ancestry, anomalous X-chromosome heterozygosity or poor genotyping efficiency were removed from further analysis (n = 46). All markers that were intensity probes had technical failure, minor allele frequency ≤ 0.05 genotyping efficiency <99%, discordant calls with duplicates, and Mendelian errors >0 were removed. Tests of association were performed with and without SNPs that deviated from Hardy Weinberg Equilibrium (p-value $<1.0 \times 10^{-4}$); consistent findings were observed for both tests (data not shown).

Statistical methods

Greater than 930K SNPs from the Illumina 1M BeadChip were tested for an association with PR interval, QRS duration, QTc interval, and heart rate using linear regression assuming an additive genetic model. Tests of association were performed unadjusted and adjusted for age, sex, PR/QT drug usage (for PR and QT interval only), and principal components (PCs; two to four PCs depending on the trait) using PLINK (Purcell et al., 2007). Power calculations were performed using QUANTO (Gauderman and Morrison, 2006) using effect size from the original, published GWAS. The fixation index Fst, a measure of population differentiation, was calculated using the Weir and Cockerham algorithm (Weir and Cockerham, 1984). We calculated Fst between European descent populations and our African American study population using the Platform for the Analysis, Translation, and Organization of large-scale data (PLATO) (Grady et al., 2010). Pair-wise linkage disequilibrium was calculated (r^2) around the *SCN5A/SCN10A* and *NOS1AP* regions using the SeattleSNPs Genome Variation Server (http://gvs.gs.washington.edu/GVS/index.jsp).

Generalization

Using the NHGRI GWAS catalog (Hindorff et al., 2009), we identified SNPs associated with PR interval, QTc interval, QRS duration, and heart rate at significance threshold of $p<10^{-6}$ from European, Indian or Asian-descent populations. We then examined the significance level and direction of effect for the same SNP-trait pair in this African American dataset.

Results

Discovery

None of the SNPs tested for an association met genome-wide significance ($p<5.0x10^{-8}$) with any ECG trait. Although we did not detect associations at genome-wide significance, we were able to detect associations at $p<10^{-6}$ (Table 3.7, Appendix G). Nine common variants were associated with increased heart rate, three of which were in complete linkage disequilibrium with each other ($r^2= 1.0$) in *SERPINI1* on chromosome 3 (Appendix G). Aside from SNPs associated with heart rate, nine associations at $p<10^{-6}$ were observed for QT interval, QRS duration, and PR interval, separately (27 associations total, Table 3.7). None of these associations have been reported by previous GWAS for ECG traits in any population.

Table 2.6. Descriptive statistics of the study population. Mean and standard deviation were calculated in the study population for all covariates and ECG traits.

Study Population Descriptive Statistics, n =455		
Variable	**Mean /%**	**SD**
Female	77%	--
Age (yr)	46	15
PR duration (msec**)**	159	21
QRS duration (msec**)**	82	8
QTc duration (msec**)**	410	21
Heart Rate (msec**)**	74	11
On QT drug	10%	--
On PR drug	21%	--

Table 2.7 Most significant GWAS results in African Americans (n = 455). We performed linear regression assuming an additive genetic model for over 930 SNPs in the genome. All tests were adjusted by age, sex, and principal components. SNP-trait. significant associations at $p \leq 10^{-6}$ are shown.

Heart Rate					
CHR	SNP	LOCATION	MINOR ALLELE	BETA	P
8	rs1015003	Intergenic	G	3.94	1.21E-06
8	rs6468401	Intergenic	A	3.93	1.55E-06
5	rs816475	Intergenic	T	3.69	2.83E-06
3	rs13090836	*SERPINI1* (intron)	T	3.58	6.86E-06
3	rs9815034	*SERPINI1* (intron)	A	3.59	7.26E-06
3	rs1473511	*SERPINI1* (intron)	T	3.59	7.26E-06
12	rs12824981	*TMEM132D* (intron)	T	4.56	8.90E-06
9	rs35061590	*DAB2IP* (intron)	T	16.03	9.45E-06
9	rs13290547	*DAB2IP* (intron)	T	16.03	9.45E-06
QT Interval					
CHR	SNP	LOCATION	MINOR ALLELE	BETA	P
5	rs6894385	Intergenic	C	-9.92	1.03E-06
6	rs9342616	Intergenic	A	-6.92	1.04E-06
7	rs12666280	*DPP6* (intron)	C	8.93	1.66E-06
20	rs237450	Intergenic	A	6.86	5.36E-06
4	rs6819013	Intergenic	A	6.06	5.50E-06
4	rs4698433	*FGFBP2* (near 5' region)	T	7.23	5.53E-06
4	rs1483012	*LDB2* (intron)	G	6.02	7.19E-06

16	rs8045405	Intergenic	G	-11.38	9.44E-06
15	rs17237606	*UNC13C* (intron)	G	-8.91	9.92E-06
QRS Duration					
CHR	SNP	LOCATION	MINOR ALLELE	BETA	P
14	rs1867082	Intergenic	A	2.82	1.46E-06
6	rs504008	*NKAIN2* (intron)	C	2.70	2.58E-06
5	rs6861497	Intergenic	A	2.40	4.73E-06
11	rs308309	Intergenic	C	4.39	4.95E-06
6	rs12194062	*C6orf190* (intron)	T	5.16	5.10E-06
4	rs6820368	*ADH6* (intron)	C	9.90	5.57E-06
12	rs10784762	Intergenic	T	-2.38	6.31E-06
16	rs17444745	Intergenic	A	-6.39	6.96E-06
1	rs13375391	*HMCN1* (intron)	A	5.82	9.37E-06
PR Interval					
CHR	SNP	LOCATION	MINOR ALLELE	BETA	P
11	rs1994318	*MICAL2* (intron)	A	-6.79	1.52E-06
6	rs10447419	Intergenic	A	-8.65	1.95E-06
10	rs16926523	*MYO3A* (intron)	A	9.55	4.23E-06
2	rs7604827	*VWC2L* (intron)	C	6.54	4.52E-06
3	rs3733017	*BCL6* (intron)	G	-10.07	5.55E-06
15	rs746265	Intergenic	C	-6.27	7.92E-06
3	rs1524976	*MAGI1* (intron)	A	7.99	7.96E-06
1	rs3103778	*MFSD2* (intron)	G	6.30	8.67E-06
15	rs12595668	Intergenic	G	-6.26	9.01E-06

Generalization

In European-descent populations, 118 GWAS-identified SNPs have been reported for ECG traits in the NHGRI GWAS catalog as of October 2011. We examined the reported European-descent SNP-trait association to determine if the association "generalized" to African Americans. We considered an association generalized if the observed association was significant at a liberal threshold ($p<0.05$) and had consistent direct of effect as the original study. Of the 118 SNPs identified in the NHGRI GWAS catalog, 92 were also directly genotyped in our study population: 26, 14, 19, and 33 SNPs were previously associated with heart rate, PR interval, QRS duration, and QT interval, respectively.

Overall, there were five (of 92) SNP-trait associations that generalized to African descent populations with respect to both level of significance ($p<0.05$) and direction of effect: for heart rate rs4352210 (intergenic, $\beta= -1.6$, $p= 0.03$), for QTc interval, rs1112795 (*SCN5A,* $\beta = - 4.77$; $p=0.05$) and rs4725982 (*KCNH2,* $\beta = 4.79$; $p=0.03$), and for QRS duration, rs6795970 (*SCN10A,* $\beta = 2.43$; $p=0.01$) and rs11710077 (*SCN5A,* $\beta= 2.42$; $p=0.004$) (Table 2.8). No SNPs generalized for PR interval. There were six additional SNP-trait associations from the GWAS catalog that were significant (p-value <0.05) but trended in the opposite direction in African Americans compared with European-descent populations (after accounting for the coded allele): four for PR interval, three for heart rate, and one for QT interval (Appendix H).

An additional 40 SNP-trait associations failed to achieve the liberal threshold of significance ($p<0.05$) in African Americans but did have consistent directions of effect compared with European-descent populations (Figure 2.2, Appendix I). Based solely on direction of effect, approximately 50% and 52% of the associations tested for PR and QT intervals, respectively, have a consistent direction of effect in African Americans compared with European descent populations. For QRS duration, 37% of the associations tested generalized in African Americans while 44% generalized for heart rate. Collectively, 44 out of 92 (47%) SNP-trait associations had a consistent direction effect regardless of significance in African Americans compared to previously-identified associations in European-descent populations. The remaining 48/92 (52%) SNP-trait associations identified in European-descent populations did not generalize to African American with respect to either level of significance or direction of effect.

There were ten SNPs reported in the NHGRI catalog that are associated with ECG traits in Asian and Indian descent populations. Of these, six SNPs were also genotyped in our study population. None of these SNPs generalized in African Americans with respect to significance and direction of effect, although two (out of six) of these associations had a consistent direction of effect compared with the original study population.

Power

SNP-trait associations that failed to achieve significance or failed to generalize to African Americans may represent lack of power or true differences in allelic architecture and/or linkage disequilibrium patterns between the two populations. To help distinguish between these two possibilities, we calculated the power to detect the European-reported

association in this African American sample. Overall, 17 SNP-trait tests of association were adequately powered (>80%). Among the adequately powered tests of association, almost all (16/17) SNP-trait tests of association failed to generalize in African Americans with respect to level of significance and/or direction of effect (Table 2.9).

Allele Frequencies

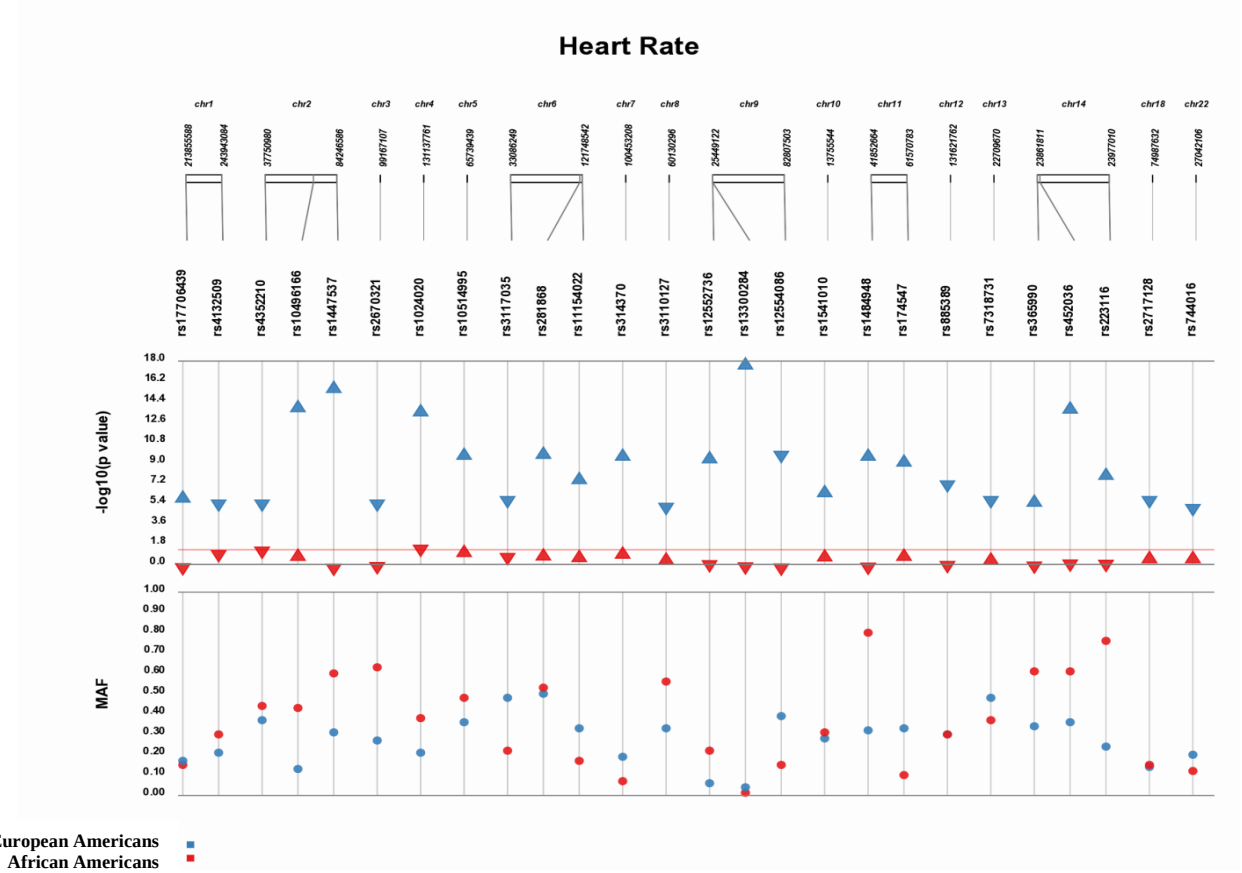
We calculated the coded allele frequencies for previously identified SNPs (from European descent populations) in African Americans (Appendix J). To measure population differences, we calculated Fst for all SNPs previously identified by GWAS in European descent populations that report sample sizes and minor allele frequencies (85 SNPs). There were 20 SNPs with an Fst value >0.15 (ranging from 0.15 to 0.85), which is indicative of significant population differentiation at these loci. We observed the largest Fst value (F= 0.85) for rs789852 located in the intron region of the *TMEMFF* gene located on chromosome 3. This SNP was previously associated with QT interval in Europeans (β= 0.25, p=7.0 x 10^{-7}) (Marroni et al., 2009). This association did not generalize in African Americans and trended in the opposite direction compared to Europeans (β = -0.95, p =0.60).

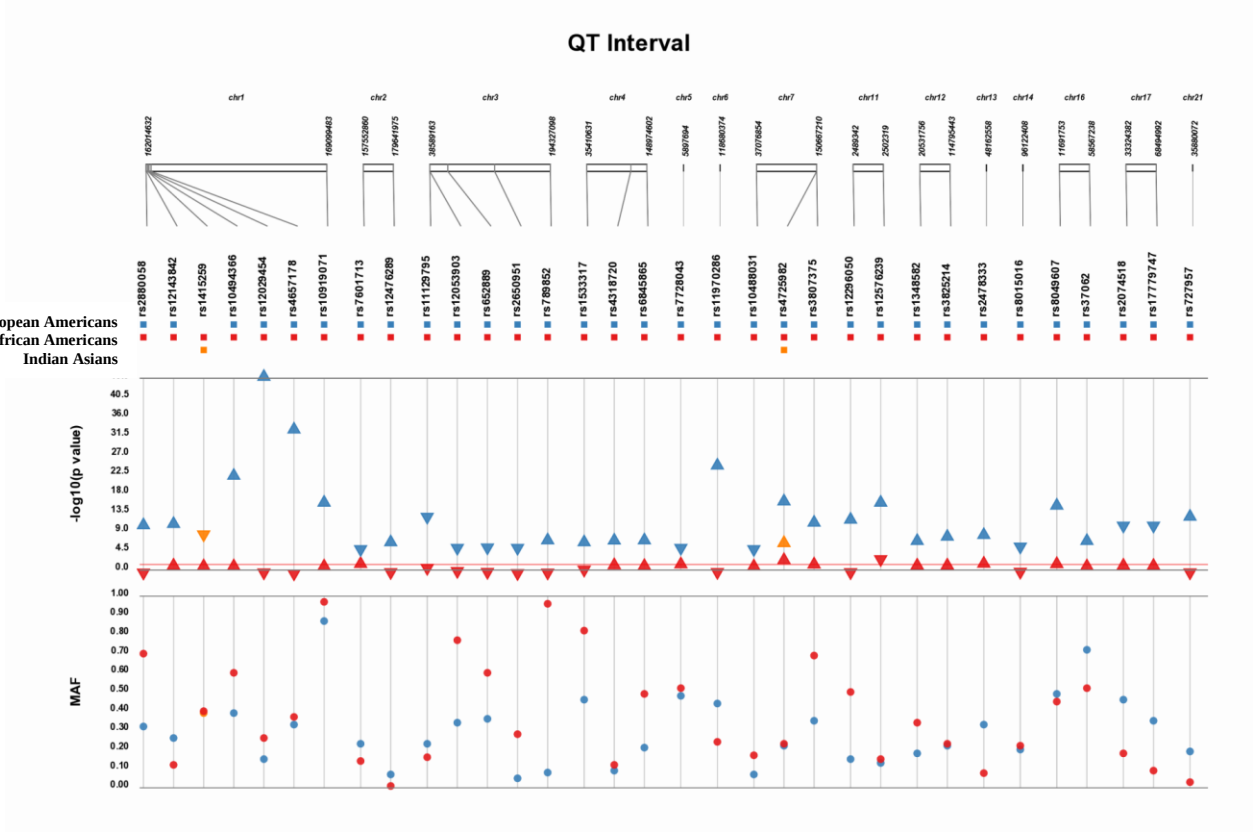
Table 2.8 GWAS-identified associations that generalize in African Americans. We compared GWAS-identified associations originally identified in European or Asian descent populations to tests of associations performed in 450 African Americans from BioVU and NUgene. We declared generalization as having a consistent direction of effect and a p-value ≤ 0.05. Abbreviations: CAF= coded allele frequency.

Trait	SNP	Coded Allele	Gene	Original Study				African Americans		
				Population	P-value	β	CAF	P-value	β	CAF
QT interval	rs11129795	A	*SCN5A*	European	5.0E-14	-1.27	0.23	0.05	-4.77	0.16
QT interval	rs4725982	T	*KCNH2*	European and Indian Asian	5.0E-16	1.58	0.22	0.03	4.79	0.23
QRS duration	rs6795970	A	*SCN10A*	European and Indian Asian	4.0E-09	5.17	0.36	0.01	2.43	0.08
QRS duration	rs11710077	T	*SCN5A*	European	1.0E-06	0.44	0.21	0.004	2.42	0.11
Heart rate	rs4352210	A	Intergenic	European	2.00E-06	-0.14	0.37	0.03	-1.6	0.44

Table 2.9 Adequately powered GWAS identified SNPs that do not generalize in African Americans. Using the effect size, mean and standard deviations reported in the original study, we calculated our power to detect the association given our sample size and minor allele frequency in African Americans. Abbreviations: CAF = coded allele frequency

SNP	Trait	Gene	Coded Allele	Original Study			African Americans (n=450)			
				CAF	P-value	Beta	CAF	P-value	Beta	Power
rs10494366	QT Interval	*NOS1AP*	G	0.39	5.00E-22	12.20	0.60	0.88	0.28	0.99
rs11129795	QRS duration	*SCN5A*	G	0.77	5.00E-10	-8.24	0.84	0.73	-0.25	0.99
rs12296050	QT Interval	*KCNQ1*	T	0.15	1.00E-11	13.16	0.50	0.52	-1.20	0.99
rs3807989	PR interval	*CAV1*	A	0.40	7.00E-13	6.40	0.62	0.46	0.94	0.95
rs3825214	QRS duration	*TBX5*	G	0.22	3.00E-13	7.35	0.23	0.52	-0.41	0.99
rs3825214	PR interval	*TBX5*	G	0.22	1.00E-07	5.88	0.22	0.79	-0.40	0.81
rs1321311	QRS duration	*CDKN1A*	T	0.21	3.00E-10	6.52	0.39	0.17	0.77	0.99
rs174547	Heart Rate	*FADS1*	C	0.33	2.00E-09	6.20	0.10	0.43	1.03	0.99
rs223116	Heart Rate	*MYH7,NDNG*	A	0.24	3.00E-08	7.40	0.76	0.44	-0.64	0.99
rs2461751	PR interval	Intergenic	G	0.44	8.00E-06	4.54	0.68	0.83	-0.30	0.80
rs281868	Heart Rate	*SLC35F1*	G	0.50	4.00E-10	6.30	0.53	0.41	0.64	0.99
rs314370	Heart Rate	*SLC12A9*	C	0.19	6.00E-10	7.60	0.07	0.28	1.58	0.99
rs365990	Heart Rate	*MYH6*	G	0.34	7.00E-06	5.25	0.61	0.61	-0.41	0.99
rs3807375	QT Interval	*KCNH2*	T	0.35	5.00E-11	11.95	0.69	0.32	1.95	0.99
rs452036	Heart Rate	*MYH6*	A	0.36	4.00E-14	7.80	0.61	0.41	-0.68	0.99
rs7660702	PR interval	*ARHGAP24*	T	0.74	3.00E-17	8.46	0.44	0.60	0.65	0.99


Heart Rate
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr18
chr22
213855588
243943084
37750980
84246586
99167107
131137761
65739439
33086249
121748542
100453208
60130296
25449122
82807503
13755544
41653664
61570783
131621762
22709870
23861811
23977010
74987632
27042106
rs17706439
rs4132509
rs4352210
rs10496166
rs1447537
rs2670321
rs1024020
rs10514995
rs3117035
rs281868
rs11154022
rs314370
rs3110127
rs12552736
rs13300284
rs12554086
rs1541010
rs1484848
rs174547
rs885389
rs7318731
rs365990
rs452036
rs223116
rs2717128
rs744016
-log10(p value)
18.0
16.2
14.4
12.6
10.8
9.0
7.2
5.4
3.6
1.8
0.0
MAF
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
European Americans
African Americans


QT Interval
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr11
chr12
chr13
chr14
chr16
chr17
chr21
rs2880058
rs12143842
rs1415259
rs10494366
rs12029454
rs4657178
rs10919071
rs7601713
rs12476289
rs11129795
rs12053903
rs652889
rs2650951
rs789852
rs1533317
rs4318720
rs6845865
rs7728043
rs11970286
rs10488031
rs4725982
rs3807375
rs12296050
rs12576239
rs1348582
rs3825214
rs2478333
rs8015016
rs8049607
rs37062
rs2074518
rs17779747
rs727957
European Americans
African Americans
Indian Asians
-log10(p value)
MAF

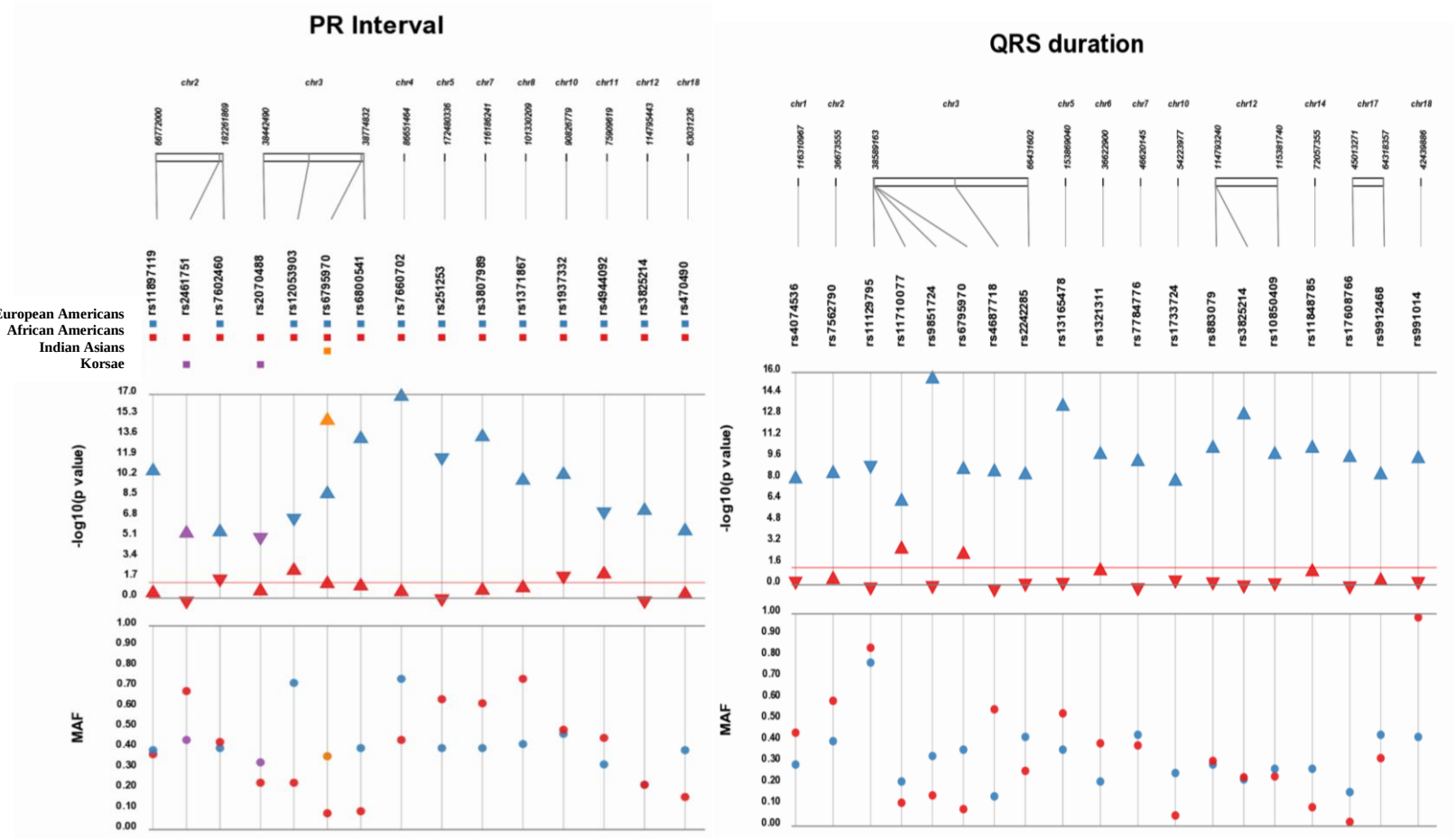


Figure 2.2 SNP-trait association comparison between European or Asian-descent and African American populations across ECG traits. Each SNP was tested for an association with each ECG trait assuming an additive genetic model adjusted for age and sex. P-values are –log10 transformed along the y-axis and corresponding location for each SNP is located on the x-axis. Each point represents a p-value for each population indicated by color (see legend). The direction of the arrows corresponds to the direction of the effect (measured by beta coefficient). The significance threshold is indicated by the red bar at p= 0.05. The bottom panel displays the minor allele frequency comparisons for both populations for each SNP.

Discussion

We conducted a genome-wide association study for ECG traits in African Americnas ascertained from biorepositories linked to electronic medical records. Although no test of association achieved genome-wide significance, here we identify novel associations with heart rate, QT interval, QRS duration, and PR interval at $p < 10^{-6}$. We also characterized previously reported associations in our African American study population derived in a clinical based setting. To our knowledge we are the first to perform a genome-wide association study for heart rate in African Americans and to characterize GWAS-identified variants from European and Asian-descent populations to African Americans for heart rate, QT interval, and QRS duration.

Discovery

We identified 36 novel associations with the four ECG traits at the liberal threshold of $p< 1.0 \times 10^{-6}$. While none of our associations met genome-wide significance, reporting these associations can provide insight to novel regions in the genome that impact ECG trait variability yet to be explored in African populations. Interestingly we were able to detect a novel region on chromosome 3 with three SNPs that are in high LD and were associated with increased heart rate in the *SERPINI1* (Appendix G.3). The *SERPINI1* gene encodes serine peptidase inhibitors secreted by axons in the brain and are associated with familial encephalopathy (Davis et al., 1999b; Davis et al., 1999a).

We also identified a region on chromosome 9 within the *DAB2IP* gene encompassing two SNPs that were associated with increased heart rate. This gene encodes a GTPase-activating protein and is an inhibitor of cell growth and survival**.** *DAB2IP* has been identified by a GWAS for an association with prostate cancer in European and African descent populations (Duggan et al., 2007). Recently, a GWAS of individuals from Iceland and the Netherlands identified an association with *DAB2IP* with abdominal aortic aneurysms and other vascular diseases including early onset myocardial infarction (Gretarsdottir et al., 2010).

Generalization

Four of the five generalizations we identified in African Americans are located in voltage-gated channels (Table 2.8). For QRS duration, two of these variants are within the *SCN5A/SCN10A* region and generalize in African Americans (Table 2.8). *SCN10A* variant rs975970, originally associated with PR interval, has been previously reported to have pleiotropic effects with QRS duration (Holm et al., 2010; Chambers et al., 2010). This association for QRS duration generalized to African Americans ($\beta = 2.43$, $p= 0.01$). There were a total of three associations that generalized in African Americans for QT interval, two of which are in voltage gated channels *KCNH2* and *SCN5A* for QT interval (Table 2.8). The third association, rs12143842, is located in *NOS1AP*, which has been consistently associated with prolonged QT interval (Arking et al., 2006; Post et al., 2007; Arking et al., 2009). The effect of this variant accounts for more than 1.5% of the trait variability in Europeans (Arking et al., 2006) and 1.3% in African Americans.

For PR interval, two previously identified associations in the *SCN10A* gene have consistent directions of effect with African Americans (Figure 2.2, Appendix I). Non-synonymous variant rs6795970 (in the *SCN10A* gene) has been recently reported to have an association with increased PR interval in Indians and Europeans (Holm et al., 2010; Pfeufer et al., 2010; Chambers et al., 2010; Denny et al., 2010b). In African Americans the magnitude this effect (β= 3.67) is consistent with Europeans (β= 5.17) (Figure 2.2, Appendix I).

There were 48 associations that did not generalize to African Americans with respect to direction of effect and statistical significance. GWAS associations identified in European descent populations may not generalize in African Americans for several reasons. More often than not, this is due to discordant minor allele frequencies and linkage disequilibrium (LD) patterns, both of which impact statistical power. On average for SNPs that do not generalize, the MAF in the original population was significantly different compared to the MAF in our study population ($p \leq 0.01$). The MAF was lower in African Americans compared to the original study population for 60% of the SNPs that did not generalize (Figure 2.2). Low minor allele frequency is directly correlated with power in genetic association studies. We also noticed this correlation with in our study; 13/79 inadequately powered tests of association had a minor allele frequency < 0.09.

The Effects of Linkage Disequilibrium

Linkage disequilibrium (LD) is often used in genetic association studies to select tagSNPs for genotyping to represent a region in the genome. While this

method is cost effective, identifying an association with a tagSNP does not necessarily identify the true functional variant given that the un-assayed functional variant is most likely in LD with the assayed variant. TagSNPs are also population-specific; therefore, associations identified in one population may not necessary generalize to another population if only the index variant is genotyped. Our observations in this African American population compared with European descent populations for ECG trait associations are consistent with known properties of tagSNPs. That is, there are differences in LD between African Americans and Europeans for known loci associated with PR interval, QRS duration, and QT interval. For the *SCN10A/SCN5* and *NOS1AP* regions we calculated pair-wise LD, measured by r^2 for each SNP pair. As expected, there is less linkage disequilibrium in African descent populations compared to European populations (Appendices K and L) (Barrett et al., 2005; Mueller et al., 2005; Reich et al., 2001b). This lack of LD could account for the non-generalization of associations in African Americans originally identified in European descent populations.

LD is often described as a "double-edged" sword: strong LD can be advantageous to identifying associations, but is it not useful when searching for the functional variants. Using populations with low amounts of LD can help fine map the functional variant (Rosenberg et al., 2010). Therefore, index associations that robustly associate with ECG traits in European descent populations but do not generalize to African descent populations could be prioritized for fine-mapping studies to identify the functional variant. Also, further large discovery and fine-

mapping studies should be performed in African descent populations to identify additional ECG-trait associated loci not found in other populations.

Limitations and Strengths

There were several factors that limit our ability to detect novel associations at the GWAS level ($p < 1.0 \times 10^{-8}$) with ECG traits. Our study population compared to recently published GWAS for ECG traits is relatively small and underpowered. Specifically, we were underpowered (<80%) to detect effects that explain less than 10% of the trait variability at $p<1.0 \times 10^{-8}$ even for common variants (MAF >0.05). Power also limited our ability to generalize European/Asian identified associations to African Americans. As previously mentioned power is directly correlated to allele frequencies. Most GWAS fixed-content products are biased to common variation and based on LD patterns European populations (Rosenberg et al., 2010). Therefore, the SNPs previously identified may only be common in Europeans and rare in other populations such as African Americans, limiting our power to generalize these variants. It is important to note the power calculations we report in our generalization analysis was limited to the effect sizes detected in the original study. These effect sizes could be subject to the "winner's curse" and over estimate the true effect size (Xiao and Boehnke, 2009; Zhong and Prentice, 2010). Using these potentially over estimated effect sizes in our power calculations could consequently over estimate our power to detect these effects.

Here we confirm the ability to use electronic medical records (EMRs) linked to DNA to identify genotype-phenotype relationships. We identify several

novel associations across the genome for heart rate, PR interval, QT interval, and QRS duration. Additionally, we generalize several biologically relevant associations identified by GWAS in European and Asian descent populations to African Americans. Most importantly, we extensively characterize European and Asian identified associations in African Americans by reporting population specific information such as allele frequency and linkage disequilibrium differences. Reporting population specific information on GWAS-identified variants can be useful to clinicians as it indirectly contributes to the field of personalized medicine.

Conclusions

We successfully assessed the generalizability of European identified genetic associations for ECG traits in African Americans derived from an electronic medical record. Additionally, we report novel associations for ECG traits in African Americans. Replication of these findings in an independent African American population is needed to confirm these associations. Most of the associations identified in European and Asian descent populations failed to generalize due to power (as a consequence of sample size and low minor allele frequencies). However, there were several adequately powered associations that did not generalize to African Americans. These data suggest that the function variant has yet to be identified for these traits by GWAS and demonstrate the need for future fine mapping studies in a large African-descent population.

C. Genetic Modifiers of ECG traits in African Americans

Introduction

There are several known risk factors of ECG traits such as age, sex, hyperlipidemia, hyperinsulinaemia, and hypertension (Lind et al., 1997). As previously mentioned, ECG traits also have a genetic component. Current heritability estimates report 30-60% of the variability in ECG traits is attributable to genetic factors (MATHERS et al., 1961; Perez and Keith, 1978; Havlik et al., 1980; Hanson et al., 1989; Kupper et al., 2004; Moller et al., 1982; Russell et al., 1998). It is important to note that original heritability estimates only account for additive genetic effects and could be underestimating the genetic component of common disease traits by not accounting for the possibility of gene-gene and gene-environment interactions(Manolio et al., 2009). To date there are few studies that investigate the role of genetic modifiers and ECG traits especially in African Americans.

One of the primary goals of the Population Architecture using Genomics and Epidemiology (PAGE) study is to identify genetic modifiers of common disease (Matise et al., 2011). As a member of the PAGE study, the Epidemiologic Architecture of Genes Linked to Environment (EAGLE), uses a diverse, cross-sectional survey, NHANES III, to identify genetic modifiers. Here we examine the effect of genetic modifiers (gene-gene and gene-environment) for three ECG traits: P duration, PR interval, and QRS interval in non-Hispanic blacks from

NHANES III (previously described). Using five previously identified SNPs from the Jackson Heart Study (See Section A of this Chapter) in *SCN5A*; we test for SNP-SNP interactions, as well as SNP-environment interactions with these SNPs and known environmental risk factors (age and sex).

Methods

Study Population and ECG measurements

Non-Hispanic blacks with normal ECGs from NHANES III were available for this study. For more detail on exclusion/inclusion criteria as well as population characteristics please refer to the methods sub section *"Study Population"* and Table 2.3 in part A of this chapter. Electrocardiograms (ECGs) were recorded on men and women in the mobile examination center (MEC) using a standard 12-lead resting ECG. ECGs were recorded using the Marquette MAC 12 (Marquette Medical Systems, Inc, Milwaukee, Wisconsin (U.S. DHHS, 1996). ECGs were only performed on individuals 40 years of age or greater, and NHANES III was the last NHANES survey that collected ECG data. More in-depth description on ECG measurements in NHANES III can be found in section A under *"Study Characteristics and ECG Measurements"* of this chapter (page 22).

SNP selection and Genotyping

DNA was extracted from crude cell lysates from lymphoblastoid cell lines established for NHANES III participants aged 12 and over as part of Genetic NHANES (Steinberg et al., 1997). We chose 12 SNPs that were significant at p

$1.0x\ 10^{-4}$ in Jackson Heart Study for genotyping in NHANES III. To date five SNPs have been successfully genotyped for this study for P duration, QRS interval and PR interval (7 SNPs have yet to be genotyped). All genotyping was performed in the Center for Human Genetics Research DNA Resources Core and the laboratory of Dr. Jonathan Haines using either Sequenom's iPLEX Gold assay on the MassARRAY platform (San Diego, CA) or Illumina's BeadXpress. All genotype data reported here passed CDC quality control metrics and are available for secondary analysis through CDC.

Statistical Analysis

Using linear regression, cross product terms (all pair-wise SNP-SNP or SNP-environment terms for five *SCN5A* variants) were added to the regression models for each ECG trait (PR interval, P duration, and QRS interval). All SNPs were coded additively, and all models were adjusted for the main effect of the SNP (both the SNP-SNP and SNP-environment models) and environmental factor (only the SNP-environment models). In addition to adjusting for main effects, all analyses were adjusted for known covariates age and sex. Analyses were limited to self-identified non-Hispanic blacks from NHANES III. All analyses were performed using the Statistical Analysis Software (SAS v.9.2; SAS Institute, Cary, NC) via the Analytic Data Research by Email (ANDRE) portal of the CDC Research Data Center (RDC) in Hyattsville, MD. We consider an interaction significant at $p<0.05$.

Results

We first sought to identify SNP-SNP interactions between the five SNPs in the *SCN5A* gene. We performed a total of 30 tests in non-Hispanic blacks with ECG measurements in NHANES. We did not detect any significant associations at $p<0.05$. In addition to SNP-SNP interactions we also tested for SNP-environment interactions between age and sex and each *SCN5A* SNP for three ECG traits. We performed a total of 30 tests and identified one interaction term (rs6768664* age; $\beta= 0.25$) that trended towards significance ($p= 0.06$).

Discussion

Here we tested for geneti or environmental modifiers of candidate genes identified association for ECG traits. None of the tests of GxG or GxE were significant at the $p<0.05$ level; however one trends towards significance. Several limitation imoacted our ability to identify modifiers as described below.

A major study limitation was sample size. We were severely under powered (<20%) to detect SNP-SNP and SNP-environment interactions, which typically have very small genetic effects and thus require larger sample sizes compared with studies designed to detect the main effects. Given our allele frequencies for *SCN5A* SNPs (Table 2.4), >4,000 participants would be required to detect gene-gene or gene-environment effects that explain less than 2% of the trait variability for any ECG traits tested.

In addition to sample size, another reason we failed to detect interactions in the present analysis maybe due to our limited SNP set. This study was limited to a small SNP set of five SNPs that were previously associated in JHS participants. Therefore all SNPs had to have a significant main effect with an ECG trait in order to be included in the present study. It is possible that GxG and/or GxE interactions can exsists in the absence of main effects.

Preliminary studies such as this provide the framework for future analysis. The results from this study give information about the sample size needed to detect gene-gene and gene-environment interactions in African Americans for *SCN5A* variants. One way to increase the sample size is to perform a meta-analysis, which combine association results across several study populations. Since NHANES participants are a part of EAGLE, a study site within PAGE, we can increase our sample size by including additional study sites within PAGE that also have ECG measurements in African Americans.

Although the results from this study did not successfully identify genetic modifiers ECG traits, it gives insight to the possibility of detecting interactions in future studies.

CHAPTER III

GENETIC VARIANTS ASSOCIATED WITH FIBRINOGEN AND HEMATOLOGICAL TRAITS

A. Identification and Replication of Fibrinogen Cluster Variants with Plasma Fibrinogen and Hematological Traits in African Americans[3]

Introduction

Circulating levels of hemostatic factors, such as plasma fibrinogen, and cellular components of the blood such as the concentration of hemoglobin (Hb), the numbers of white blood cells (WBC), red blood cells (RBC) and platelets (PLT), and the volumes of red blood cells (mean corpuscular volume, MCV) and platelets (mean platelet volume, MPV) are tightly regulated (Pennington et al., 2001; Evans et al., 1999). Abnormal levels of Hb, MPV, PLT, plasma fibrinogen, serum triglycerides, and WBC are associated with myocardial infarction (MI), coronary artery disease (CAD), and stroke (Meade et al., 1986; Tosetto et al., 2011; Boos and Lip, 2006; Danesh and Lewington, 1998; Ensrud and Grimm, Jr., 1992; Kannel et al., 1987). Plasma fibrinogen is a key player in the pathogenesis of cardiovascular disease, specifically arterial thrombosis. In the coagulation pathway, fibrinogen is converted to fibrin, the main protein component of an arterial clot (Collet et al., 1993; Voetsch and Loscalzo.J, 2003). The roles of genetic factors for

[3] Partially adapted from: Jeff JM, Brown-Gentry K, Crawford DC. Replication and characterization of genetic variants in the fibrinogen gene cluster with fibrinogen levels and hematological traits in the Third National Health and Nutrition Examination Survey.*Thrombosis and Haemostasis.* January 2012. 107 (3)

plasma fibrinogen and other coagulation factors have been well studied over the past 30 years; however, little is known about the genetics of other common hematological measurements and their relationship with cardiovascular related diseases.

Genetic factors are reported to account for up to 50% of the variability observed in plasma fibrinogen levels (Voetsch and Loscalzo.J, 2003). Heritability estimates for hematological measurements such as Hb, WBC, and PLT range from 0.37-0.89 (Pennington et al., 2001; Evans et al., 1999). Several genetic association studies have been conducted to assess the relationship between candidate genes of the coagulation pathway and fibrinogen, coagulation factors, and other common hemostatic measurements. Candidate gene studies have demonstrated that variation in fibrinogen genes (*FBA, FGB* and *FGG*) are associated with plasma fibrinogen levels (Siegerink et al., 2009; Boekholdt et al., 2001; Folsom et al., 2001; Koch et al., 2008; Hamsten et al., 1987).

While these studies have identified genetic variation associated with these quantitative traits, the discovery effort has been limited to European-descent populations. It is well known that genetic backgrounds vary among populations, both in allele frequency and in linkage disequilibrium patterns for many diseases including venous thrombosis (Margaglione and Grandone, 2011). As an example of allele frequency differences, one study performed in diverse populations demonstrated that Factor V Leiden (rs6025) and the prothrombin G20210A (rs1799963) variants differ across populations (Rosendaal et al., 1998). The prevalence of these variants in African Americans and Hispanics is 1-2% and

<0.04%, respectively compared to 5% and 2-4% observed in Europeans (Grody et al., 2001; Grody, 2003; McGlennen and Key, 2002). Since these well-known variants are less common in African Americans and Hispanics, it is possible that other variants have yet to be identified in these populations that explain more of the variance in hemostatic and/or hematological traits.

Using a candidate gene approach we investigated the role of several genetic variants in the fibrinogen gene cluster with plasma fibrinogen and several hematological parameters. We characterized the association of 25 single nucleotide polymorphisms (SNPs) and plasma fibrinogen levels in African Americans (referred to non-Hispanic blacks here) from the Third National Health and Nutrition Examination Survey (NHANES). We also tested for an association with these genetic variants and mean platelet volume, white blood cell count, platelet count, platelet distribution width, and serum triglycerides.

Methods

Study Population

All procedures were approved by the CDC's Ethics Review Board and written informed consent was obtained from all participants. This candidate gene association study was approved by the CDC's Ethics Review Board (protocols #2003-08 and #2006-11) and the University of Washington's Institutional Review Board (IRB #23667; HSRC D committee). Because no identifying information was accessed by the investigators, this study was considered exempt from Human

Subjects by Vanderbilt University's Institutional Review Board (IRB #061062; HS2 committee).

Ascertainment of NHANES III and method of DNA collection have been previously described (Crawford et al., 2006; Chang et al., 2009; Steinberg et al., 1997). The National Health and Nutritional Examination Surveys are cross-sectional surveys conducted across the United States by the National Center for Health Statistics (NCHS) at the Centers for Disease Control and Prevention (CDC). NHANES III was conducted between 1988-1990 (phase 1) and 1991-1994 (phase 2) (1996; 2004) as a complex survey design that over-sampled minorities (non-Hispanic blacks and Mexican Americans), the young, and the elderly. All NHANES have interviews that collect demographic, socioeconomic, dietary, and health-related data. Also, all NHANES study participants undergo a detailed medical examination at a central location known as the Mobile Examination Center (MEC). The medical examination includes the collection of physiological measurements by CDC medical personnel and blood and urine samples for laboratory tests. Beginning with phase 2 of NHANES III, DNA samples were collected from study participants aged 12 years and older.

Hematological Measurements

Blood was collected from all participants with the exception of those who reported having hemophilia or chemotherapy in the past four weeks (Gunter E. et al., 1996). Complete blood counts (CBCs) were performed on qualifying NHANES participants using the Beckman Coulter method, a method that sizes and counts particles by using measurable changes in electrical resistance produced by

nonconductive particles suspended in an electrolyte (Daum et al., 2000). A quantitative, automated, differential cell counter was used to calculate exact values for hematological measurements (Gunter E. et al., 1996). Plasma fibrinogen was measured from blood plasma on participants 40 years and older using the Clauss clotting method (Daum et al., 2000). Serum triglycerides were measured using standard enzymatic methods.

SNP Selection and Genotyping

DNA was extracted from crude cell lysates from lymphoblastoid cell lines established for NHANES III participants aged 12 and over as part of Genetic NHANES (Steinberg et al., 1997). TagSNPs were selected using LDSelect (Carlson et al., 2004) for multiple populations at r^2 >0.80 for common variants (minor allele frequency (MAF) >5%) in three candidate genes (*FGA*, *FGB*, and *FGG*) based on data available for European Americans and African Americans in SeattleSNPs (Crawford et al., 2005). A total of 25 SNPs were genotyped using the Illumina GoldenGate assay (as part of a custom 384 oligonucleotide pool assay (OPA) by the Center for Inherited Disease Research (CIDR) through the National Heart Lung and Blood Institute's Resequencing and Genotyping Service. A total of 7,159 samples were genotyped, including 2,631 non-Hispanic whites, 2,108 non-Hispanic blacks and 2,073 Mexican Americans. The data presented in this analysis were limited to non-Hispanic blacks. Quality control measurements were calculated locally using the Platform for the Analysis, Translation, and Organization of large-scale data (Grady et al., 2010). The average genotyping call rate was 96%. We flagged SNPs that deviated from Hardy Weinberg Equilibrium expectations (p-

value <0.001), MAF<0.05, and SNP call rates <95% for each subpopulation. In addition to these quality control metrics, we genotyped blinded duplicates as required by CDC, and all SNPs reported here passed quality control metrics required by CDC. All genotype data reported here were deposited into the NHANES III Genetic database and are available for secondary analysis through CDC.

Statistical Methods

All analyses were performed using the Statistical Analysis Software (SAS v.9.2; SAS Institute, Cary, NC) either locally or via the Analytic Data Research by Email (ANDRE) portal of the CDC Research Data Center (RDC) in Hyattsville, MD. Analyses were limited to self-identified non-Hispanic blacks and were at least 18 years of age. Participants with measurements outside of the normal range for any trait were excluded from the analysis, as extreme measurements for these traits can indicate inflammation or recent trauma. Pair-wise correlations for each trait were calculated using Pearson's correlation coefficient. None of the traits were correlated (defined as r>0.50; Appendix M). Descriptive statistics for all traits are reported in Table 3.1. Coded allele frequencies for all SNPs are reported in Appendix N. All traits followed a normal distribution with the exception of triglycerides, which was uniformly transformed by using the natural logarithm.

Pair-wise linkage disequilibrium was calculated for all 25 SNPs using Haploview (Appendix O) (Barrett et al., 2005). Single-locus tests of association were performed using linear regression for each fibrinogen SNP and each hematological trait: plasma fibrinogen, mean platelet volume, white blood cell

count, platelet distribution width, log-transformed serum triglycerides, and platelet count. For each test of association, we assumed an additive genetic model. Analyses were performed un-weighted and unadjusted and adjusted for age, sex, body mass index (BMI) and current smoking status (yes/no), and results were plotted using Synthesis View (Pendergrass et al., 2010b). Analyses were repeated to include time to last meal (in hours) as a covariate (data not shown). Current smoking status was determined by the question "Do you smoke cigarettes now?" and cotinine levels (>15 ng/ml).

Results

Plasma Fibrinogen and Replicated Associations

We replicated several single SNP associations from previous studies (Figure 3.1; Table 3.2). Replication was defined by statistical significance ($p < 0.05$) and consistent direction of effect with the same SNP in the same ancestral population across studies.

We were able to replicate reported associations in non-Hispanic blacks (Table 3.2)(Reiner et al., 2006a; Wassel et al., 2010h). Intronic *FGG* rs2066874 (MAF = 0.02; Appendix N) was associated with decreased plasma fibrinogen in non-Hispanic blacks ($\beta = -0.28$; $p=0.0007$; Table 3.2; Figure 3.1). Although not significant, intronic *FGA* rs2070017, which is rare in non-Hispanic whites (MAF=0.0012; Appendix N), trended towards an association with increased plasma fibrinogen in non-Hispanic blacks ($p=0.07$; $\beta = 0.07$). Overall, regardless of significance, all but one test in non-Hispanic blacks (*FGA* rs6050) trended in the same direction as the previous report after accounting for the coded allele (Figure

3.1; Table 3.2). Also, rs1049636, which was originally reported in European-descent populations (Reiner et al., 2006a), trended toward significance in non-Hispanic blacks (β = 0.06, p-value = 0.07; Figure 3.1).

We did not detect novel associations between genetic variants and plasma fibrinogen in non-Hispanic blacks (Appendix P). However, *FGA* rs2070022 located in the 3′ untranslated region of the gene, trended towards significance in non-Hispanic blacks (p = 0.07; β = -0.07; Figure 3.1). The two SNPs (replicated and novel) associated with decreased plasma fibrinogen in non-Hispanic blacks collectively accounted for 1.7% of the trait variability. With the inclusion of SNP rs2070017, which trended towards significance, these three SNPs accounted for 2.0% of the trait variability of plasma fibrinogen in non-Hispanic blacks.

Associations with Hematological Traits

In addition to testing for an association with plasma fibrinogen levels, we tested for associations between the fibrinogen gene cluster variation and traits measured in the CBC: platelet count, mean platelet volume, platelet distribution width, and white blood cell count (Figure 3.2). We identified three novel associations with fibrinogen variants and other CBC traits in non-Hispanic blacks (Figure 3.2, Table 3.3). Intronic *FGA* variant rs2070017 is associated with decreased platelet count in non-Hispanic blacks (β= -8.12, p-value = 0.01). A rare missense mutation in *FGG*, rs6063 (Gly712Arg, MAF < 0.001), is associated with increased platelet distribution width (β = 0.60, p-value =0.03). For mean platelet volume we identified an association with rs227409 in non-Hispanic blacks (β= -

1.11, p-value = 0.05). There were no significant associations observed the other traits and fibrinogen SNPs in non-Hispanic blacks (Figure 3.2; Appendix P).

Associations with Triglycerides

There were no significant associations between SNPs in the fibrinogen gene cluster and triglyceride levels in non-Hispanic blacks (Figure 3.2; Appendix P). Our primary analysis included participants regardless of fasting status, which can affect triglyceride levels in individual participants. To account for variability due to fasting status, we performed an additional analysis with inclusion of time to last meal as a covariate in the models. Inclusion of fasting status did not yield different association results compared to the original analysis (data not shown).

Table 3.1. Study population characteristics and hematological trait descriptive statistics for NHANES III non-Hispanic blacks participants. Un-weighted means (± standard deviations) or percentages are given for demographic, plasma fibrinogen, and hematologic traits, by subpopulation for adults in phase 2 of NHANES III. Abbreviations: non-Hispanic black (NHB)

Variable	Mean or %	Standard Deviation
Age (yrs)	40.7	±16
Female (%)	58	-
Current Smokers (%)	37	-
Body Mass Index (kg/m^2)	28.2	±6.67
Plasma Fibrinogen (g/L) (n=621)	2.95	±0.51
Platelet Count (n= 1,688)	271	±64
Platelet Distribution Width (mean %) (n= 1,689)	16.3	± 0.57
Mean Platelet Volume (fL) (n=1,720)	8.56	±0.98
Ln Triglycerides (mg/dL) (n= 1,708)	4.38	±0.35
White Blood Cell Count (n=1,722)	6.69	±2.11

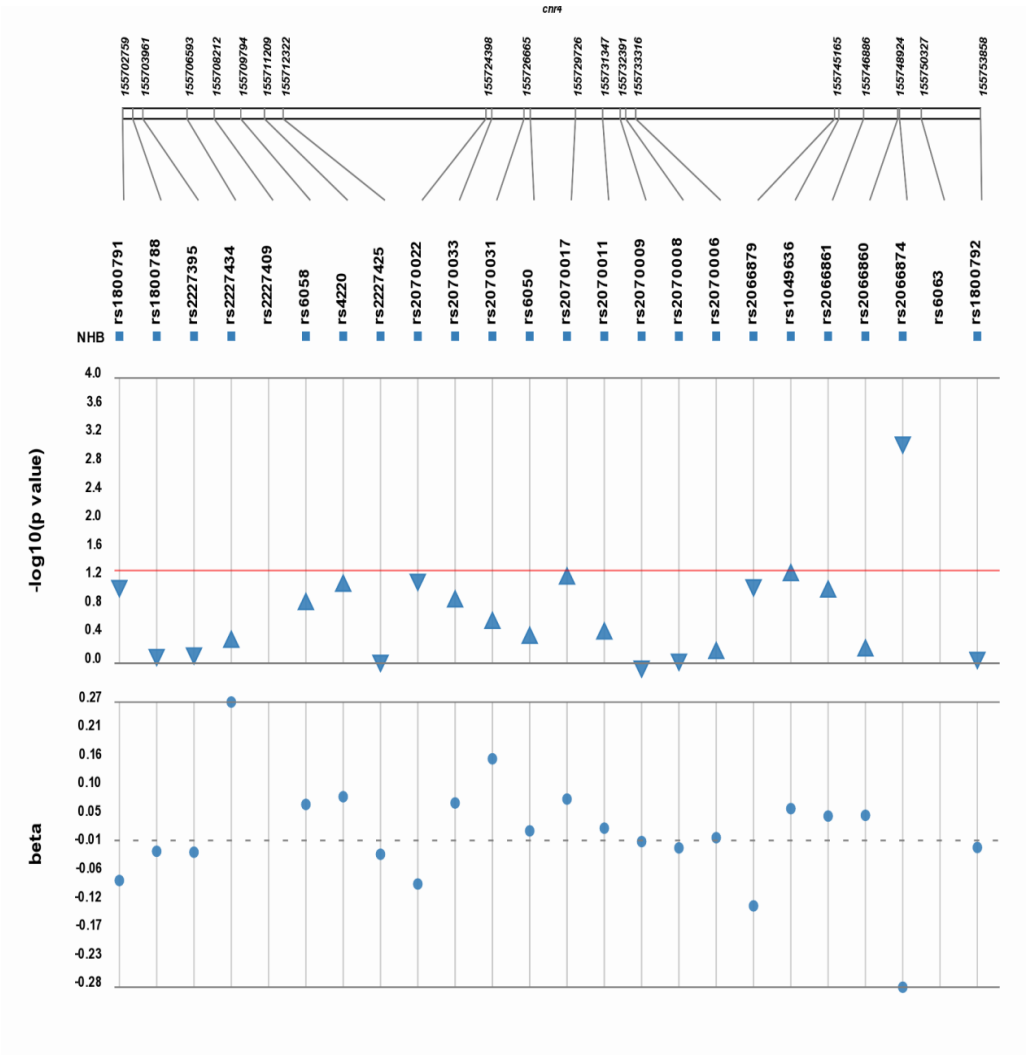


Figure 3.1. Synthesis view plot of association results for plasma fibrinogen in non-Hispanic blacks. The –log10 of the p-value and the direction of the effect (indicated by arrow direction) are on the y-axis and SNPs are located on the x-axis. Abbreviation: NHB = non-Hispanic black. Arrows above the red line represent a significant association ($p<0.05$).

Table 3.2. Replication of results from previous studies in non-Hispanic blacks for plasma fibrinogen levels. We compared results from our tests of association with plasma fibrinogen and all SNPs tested with the results from previous studies. Note that the coded allele may be different across studies. Abbreviations: Candidate gene Association Resource (CARe), Coronary Artery Risk Development in Young Adults (CARDIA), Rotterdam Study (RS), Framingham Heart Study (FHS), Cardiovascular Health Study (CHS), Atherosclerosis Rick in Communities Study (ARIC), Monitoring of Trends and Determinants of Cardiovascular Disease/Cooperative Health Research in the Region Augsburg (MONICA/KORA), genome-wide association study (GWAS).

Abbreviations: NS = not significant

	NHANES III non-Hispanic blacks				**CARe Wassel et al (GWAS)**			**CARDIA Reiner et al (candidate gene)**		
SNP	**Alleles**	**Coded Allele**	**β**	**P-value**	**Coded Allele**	**β**	**P-value**	**Coded Allele**	**β**	**P-value**
rs2070017	A/G	G	0.07	0.07	T	-0.16	4.82E-9	A	decreased	NS
rs2066874	A/G	G	-0.28	7.00 E-4	C	-0.33	2.86E-11	G	decreased	NS
rs6050	A/G	G	0.01	0.52	--	--	--	G	decreased	NS
rs6058	G/T	G	0.06	0.18	T	-0.17	9.52E-7	--	--	--

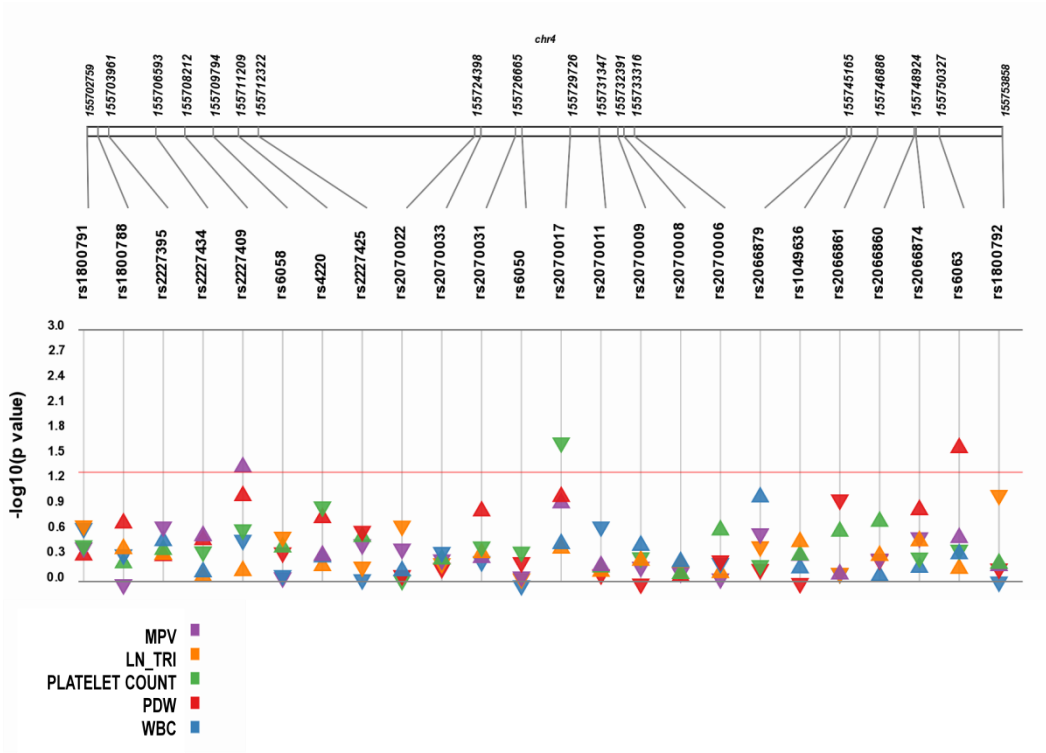


Figure 3.2. Synthesis view plot comparing significant results from tests of association for five hematological traits in NHANES non-Hispanic blacks. The –log10 of the p-value and the direction of the effect (indicated by arrow direction) is on the y-axis and SNPs are located on the x-axis. Each color represents the hematological traits tested for each SNP. Points above the red line represent a significant association ($p<0.05$). Abbreviations: non-Hispanic black (NHB, mean platelet volume (MPV), transformed triglycerides (ln_TRI), white blood cell count (WBC), and platelet distribution width (PDW).

Discussion

Using a large, diverse study population, we were able to replicate associations as well as identify novel associations between SNPs located in the fibrinogen gene cluster and levels of plasma fibrinogen in non-Hispanic blacks. Additionally, we were able to detect associations with fibrinogen SNPs and other CBC traits, indicative of pleiotropic effects.

While we successfully replicated genetic associations reported for plasma fibrinogen in non-Hispanic blacks (Table 3.2), three variants did not replicate the findings of previous studies (Table 3.2; Appendix P) (Reiner et al., 2006a; Wassel et al., 2010g; Dehghan et al., 2009). We were also unable to replicate the association with rs6058 and plasma fibrinogen in non-Hispanic blacks, most likely due to power (Table 3.2; Appendix P).

Several genetic variants have been identified to have an association with plasma fibrinogen in all three genes in the fibrinogen cluster (Reiner et al., 2006a; Wassel et al., 2010c). Most of these genetic variants are in non-coding regions of *FGB*, the rate limiting gene for the production of the fibrinogen polypeptide (Kottke-Marchant K, 2011). Little is known about function of genetic variation in the other genes, *FGA* and *FGG*, in the fibrinogen gene cluster. Of the SNPs targeted for replication for plasma fibrinogen levels, three were putatively functional SNPs based on gene location. Unfortunately, none of the novel variants we identified are in the coding regions of the fibrinogen cluster.

This study has several strengths; one is the breadth of laboratory measurements available for tests of association. Until recently, most genetic

association studies have ignored pleiotropy and have limited tests of association to a single trait or pathway. Pleiotropy is defined here and elsewhere as a locus affecting multiple traits (Wagner and Zhang, 2011). In a statistical setting, pleiotropy is detected as a single genetic variant associated with multiple traits, but this setting does not establish causality without further experiments. Epidemiologic studies such as NHANES make statistical tests for pleiotropy possible, which may ultimately identify novel genotype-phenotype relationships and pathways (Pendergrass et al., 2011). Our study is unique in that we also investigated the pleiotropic effects of the fibrinogen cluster with hematological traits. This is important given the complex pathophysiology of cardiovascular disease and the role hematological traits play in the development of disease. Several studies have been published on genetic variation with levels of plasma fibrinogen but little is known about how these same variants impact hematological traits. It is known that elevated levels of plasma fibrinogen are observed for several disease states such as inflammation, deep venous thrombosis, and arterial thrombosis (David Ginsburg, 2005). However, it is still unclear as to whether elevated levels of plasma fibrinogen are a result of fibrinogen itself or other intermediate phenotypes underlying the pathophysiology. For example, arterial thrombosis, a common characteristic and intermediate phenotype for cardiovascular disease, is very complex and is likely to involve the fibrinogen gene cluster. However, unlike venous thrombosis, which is mainly due to the accumulation fibrin clots, arterial thrombosis is primarily due to platelet adhesion and it is likely involve hematological traits tested in this study (Feinbloom and Bauer, 2005; Lane

and Grant, 2000; Lowe, 2008; Mackman, 2008; Lowe, 2008). We identified genetic associations with fibrinogen SNPs and other hematological traits suggesting possible pleiotropic effects. These findings support the idea that variants in the fibrinogen gene cluster may also affect the regulation of hematological traits.

A weakness of the present study is power and sample size. While NHANES III contains over 33,000 study participants, our study population sample size was limited for several reasons. The Centers for Disease Control and Prevention (CDC) only collected DNA samples from a subset of the entire dataset, which includes 7,159 participants from phase 2 of NHANES III (from three racial/ethnic groups). DNA was not collected from participants who reported having hemophilia or chemotherapy within four weeks of collection (Gunter E. et al., 1996). Of these samples, plasma fibrinogen levels in NHANES III was only measured for participants > 40 years of age (Gunter E. et al., 1996), which drastically reduces our sample size with plasma fibrinogen (n~ 620) compared to the hematological traits (n ~ 1,650) . Despite the small sample sizes for plasma fibrinogen, we were able to successfully replicate known associations at a liberal significance threshold of $p<0.05$.

In summary, we have replicated previous associations within the fibrinogen gene cluster with plasma fibrinogen levels in an African descent population (non-Hispanic blacks). Additionally, we show that these variants are also associated with other hematological traits. Identifying these associations will provide important information about intermediate phenotypes with the ultimate goal of predicting cardiovascular disease outcome.

B. Replication and Generalization of Genome-Wide Association Study (GWAS)-Identified Variants in African Americans as Part of the Population Architecture using Genomics and Epidemiology (PAGE) Study

Introduction

Genome wide association studies (GWAS) provide a relatively hypothesis free approach for detecting candidate disease regions in the genome. This approach tests the entire genome for an association between common variants and disease status or inter-individual trait variability. To date there are 31 GWAS in the NHGRI catalog of genome wide association studies for plasma fibrinogen and/or hematological traits (http://www.genome.gov/gwastudies/). Despite the success of these studies, only 6 % (5/31 GWAS) have been performed in African descent populations for these same traits. Most of these studies validate an association with alleles in the Duffy antigen/chemokine receptor *(DARC)* gene and white blood cell count (Crosslin et al., 2011; Reiner et al., 2011; Ramsuran et al., 2011). The presence of these "Duffy" alleles are common in African descent populations and are protective against malaria (Hadley and Peiper, 1997). Aside from these GWAS, there have been only two GWAS performed in African Americans: one for fibrinogen (Wassel et al., 2010f) and one for platelet aggregation-related traits (Johnson et al., 2010). No GWAS has yet been performed in African Americans for the other hematological traits such as mean platelet volume or platelet count.

The remaining 94% (26/31) of GWAS for fibrinogen or hematological traits have been limited to European or Asian descent populations, and it is unclear the

extent to which these associations generalize to African Americans. There is considerable trait variability between African and European descent populations for fibrinogen, platelet count, and white blood cell count (Cook et al., 2001; Hsieh et al., 2007; Segal and Moliterno, 2006). African Americans on average have a lower white blood cell count compared to Europeans, which is explained partly by the *DARC* locus; however recent GWAS have identified additional loci that explain this difference. Genetic association studies in African Americans for fibrinogen have yet to identify genetic risk factors that account for the differences in fibrinogen across populations.

In addition to trait variability between populations, there are also significant differences in allele frequencies and linkage disequilibrium patterns between Africans and Europeans at loci reported to effect fibrinogen and other coagulation factors. As previously mentioned, one study performed in diverse populations demonstrated that Factor V Leiden (rs6025) and the prothrombin G20210A (rs1799963) variants differ across populations (Rosendaal et al., 1998). The prevalence of these variants in African Americans and Hispanics is 1-2% and <0.04%, respectively compared to 5% and 2-4% observed in Europeans (Grody et al., 2001; Grody, 2003; McGlennen and Key, 2002) .

Replication and generalization of previously identified associations in African-descent populations has become an essential step for identifying the true "casual variant(s)" for common/complex diseases. The Population Architecture using Genomics and Epidemiology (PAGE) study is a large collaborative effort that

aims to examine whether GWAS-identified variants generalize to ethnically diverse population-based studies (Matise et al., 2011). The Epidemiologic Architecture for Genes Linked to the Environment (EAGLE) is a part of the PAGE study. Parallel with the goals of PAGE, EAGLE aims to characterize candidate gene and GWAS-identified variants using data from the cross-sectional survey, the National Health and Nutrition Examination Survey (NHANES). In the present study, we replicate previous GWAS associations in African Americans for fibrinogen and generalize GWAS-identified in Europeans to non-Hispanic blacks from NHANES. Additionally, we tested these variants for possible pleiotropic effects with triglycerides, mean platelet volume, and platelet count.

Methods

Study Population

All NHANES procedures were approved by the CDC Ethics Review Board and written informed consent was obtained from all participants. The present study was approved by the CDC Ethics Review Board. Because the study investigators did not have access to personal identifiers, this study was considered non-human subjects research by the Vanderbilt University Institutional Review Board.

As previously mentioned in section A of this chapter, NHANES are cross-sectional surveys that are conducted across the United States by the National Center for Health Statistics (NCHS) at the Centers for Disease Control and Prevention (CDC). One of the main goals of NHANES is to describe the health and

nutritional status of children and adults in the United States. To do so, they oversampled non-Hispanic blacks, Mexican-Americans, the young, and the elderly. NHANES III was conducted between 1988 and 1994, and DNA was extracted from crude cell lysates (Chang et al., 2009). In 1999 NHANES became a yearly survey and ~8000 new DNA samples were extracted from blood between 1999 and 2002.

For this study, ~ 2,108 and 1,350 samples from self-described non-Hispanics blacks from NHANES III and NHANES 1999-2002, respectively, were available for genotyping. We refer to this dataset as combined NHANES (n=3,458 non-Hispanic blacks). All participants were at least 18 years of age, and participants with measurements outside of the normal range for any trait were excluded from the analysis, as extreme measurements for these traits can indicate inflammation or recent trauma.

Hematological Measurements

Complete blood counts (CBCs) were performed on qualifying NHANES III participants using the Beckman Coulter method, a method that sizes and counts particles by using measurable changes in electrical resistance produced by nonconductive particles suspended in an electrolyte (Daum et al., 2000). For NHANES 1999-2000 and 2001-2002, CBCs were performed by a certified phlebotomist in the MEC using venipuncture. A quantitative, automated, differential cell counter was used to calculate exact values for hematological measurements (Gunter E. et al., 1996). Plasma fibrinogen was measured from blood plasma on participants 40 years and older using the Clauss clotting method in

all NHANES participants (Daum et al., 2000). Serum triglycerides were measured using standard enzymatic methods. Blood was collected from all participants with the exception of those who reported having hemophilia or chemotherapy in the past four weeks (Gunter E. et al., 1996).

SNP Selection and Genotyping

Using the NHGRI GWAS catalog (http://www.genome.gov/gwastudies/), we selected SNPs that have been previously identified by GWAS for the following traits: fibrinogen, mean platelet volume, red blood cell volume, platelet aggregation, mean corpuscular hemoglobin content, white blood cell count, red blood cell count and red blood cell volume. A total of 32 SNPs that were previously identified from GWA studies or well-studied in candidate gene analysis were selected for genotyping (Appendix Q), and four of the nominated SNPs were available for this study (the other SNPs have yet to be genotyped). Genotyping was performed in the Center for Human Genetics Research DNA Resources Core and the laboratory of Dr. Jonathan Haines. All SNPs were genotyped using either Sequenom's iPLEX Gold assay on the MassARRAY platform (San Diego, CA) or Illumina's BeadXpress. All genotype data reported here passed CDC quality control metrics and are secondary analysis through CDC. Aggregate data will be available through dbGaP under EAGLE/PAGE.

Statistical Methods

All analyses were performed using the Statistical Analysis Software (SAS v.9.2; SAS Institute, Cary, NC) either locally or via the Analytic Data Research by

Email (ANDRE) portal of the CDC Research Data Center (RDC) in Hyattsville, MD. Pair-wise correlations for each trait were calculated using Pearson's correlation coefficient. None of the traits were correlated (defined as r>0.50; Appendix M). Descriptive statistics for all traits are reported in Table 3.3. All traits followed a normal distribution with the exception of triglycerides, which was uniformly transformed by using the natural logarithm.

Using standard linear regression, assuming an additive model, we tested two GWAS-identified variants for an association with fibrinogen, one GWAS-identified variant with platelet count and one GWAS-identified variant with mean cell volume. Since we did not test mean cell volume in this analysis, this SNP was limited our analysis of pleiotropic effects with fibrinogen and the aforementioned hematological traits. To identify possible pleiotropic effects, we tested GWAS-identified variants for fibrinogen with an association with mean platelet volume, platelet count, and log-transformed triglycerides, and the GWAS-identified variant for platelet count was tested for an association with fibrinogen, triglycerides, and mean platelet volume. All tests were limited to non-Hispanic blacks and adjusted for age, sex, body mass index (BMI), and smoking status. Current smoking status was determined by the question "Do you smoke cigarettes now?" and cotinine levels (>15 ng/ml). Participants that answered "yes" and have cotinine levels >15ng/ml were classified as a current smoker.

Results

We were able to successfully replicate previous GWAS associations with fibrinogen cluster variants and plasma fibrinogen in our combined NHANES non-Hispanic blacks (Table 3.4). Replication is defined here as having a consistent direction of effect and statistical significance at the $p<0.05$ level in the same study population as the original study. We were able to replicate two associations: rs2070017 (β = -0.08, p= 0.02) located in the *FGA* gene and *FGG* intronic SNP rs2066874 (β = -0.28, p= 1.5 E-4) with decreased fibrinogen. Both associations were originally identified in African Americans(Wassel et al., 2010e) and are rare or monomorphic in European (MAF <0.001) and Asian (MAF <0.001) descent populations from HapMap (2003). The minor allele frequency for these variants in the Yoruban (YRI) or other African reference populations is ~0.20 and ~0.02 for rs2070017 and rs2066874, respectively (2003). Consistent with the YRI ancestral populations, the allele frequencies are comparable in non-Hispanic blacks: MAF = 0.10 and 0.02 for rs2070017 and rs2066874, respectively (Appendix N).

Currently, there are no GWAS published for hematological traits platelet count and mean platelet volume in African Americans. We tested a GWAS-identified variant from European Americans, rs210135, for an association in combined NHANES non-Hispanic blacks for platelet count. SNP rs210135 is located in the *BAK1* gene and is common in both European (CEU) and African (YRI) HapMap reference populations, 0.32 and 0.25 respectively (2003). The minor allele frequency in combined NHANES non-Hispanic blacks is 0.28 and is

consistent with reported values in HapMap reference populations (Appendix N). In Europeans rs210135 was associated with increased platelet count (β = 5.43, p-value = 3.7E-10 (Soranzo et al., 2009), an association that generalized in non-Hispanic blacks (β = 5.69, p-value 0.003; Table 3.4).

In addition to replication and generalization, we tested for pleiotropic effects with GWAS-identified variants and other related traits available in NHANES. For example rs1800562, identified by a GWAS in Europeans and was originally associated with mean cell volume (Soranzo et al., 2009), was tested here for an association with mean platelet volume, fibrinogen, triglycerides, and platelet count. Using this approach, a total of 13 tests of pleiotropy were performed in non-Hispanic blacks, and two were significant at a liberal threshold of $p<0.05$ (Table 3.5). More specifically, intronic variant rs1800562, located in the protein coding hemochromatosis gene *HFE*, was associated with plasma fibrinogen in combined NHANES non-Hispanic blacks (β = 0.25, p-value = 7.50 E-5; Table 3.5**)**. SNP rs210135, originally associated with platelet count in Europeans (Soranzo et al., 2009) and this non-Hispanic black sample, was also associated with mean platelet volume in non-Hispanic blacks (β = 0.05, p-value = 0.04; Table 3.5**).**

Table 3.3. Study Population Characteristics. All study participants were self-described non-Hispanic blacks ascertained for NHANES III or NHANES 1999-2002. Unweighted means (± standard deviations) or percentages are given for plasma fibrinogen, and hematologic traits in combined NHANES non-Hispanic blacks Abbreviations: Standard Deviation (SD), Nutrional Health or Nutrition Examination Survey (NHANES)

Variable	NHANES III		NHANES 1999-2002		Combined NHANES	
	N	Mean (SD)	N	Mean	N	Mean (SD)
Plasma Fibrinogen (g/L)	651	2.94 (± 0.52)	480	3.35 (± 0.46)	1,133	3.11 (± 0.53)
Platelet Count	1,688	271.44 (± 63.39)	1,328	258.00 (± 62.87)	3,071	272.01 (± 70.93)
Mean Platelet Volume (fL)	1,720	8.56 (± 0.99)	1,348	8.46 (±0.94)	3,068	8.52 (± 0.97)
ln(Triglycerides) (mg/dL)	1,708	4.59 (± 0.53)	543	4.39 (± 0.35)	2,385	4.60 (± 0.61)

Table 3.4. Replication/Generalization of GWAS identified variants in non-Hispanic blacks in combined NHANES dataset (NHANES III and NHANES 1999-2002) for fibrinogen and blood traits. We performed a standard linear regression adjusted for age, sex, body mass index (BMI) and smoking status for three GWAS-identified variants with noted phenotypes. Abbreviations NR = not reported, MAF = minor allele frequency.

SNP	Gene	Phenotype	Original Study					Combined NHANES Non-Hispanic blacks			
			Population Studied	n	β	P-value	MAF	n	β	P-value	MAF
rs2070017	*FGA*	Fibrinogen	African American	6,657	-0.16	4.8E-9	0.11	1,055	-0.08	0.02	0.10
rs2066874	*FGG*	Fibrinogen	African American	6,657	-0.33	2.9E-11	0.03	1,089	-0.28	1. 5 E-4	0.02
rs210135	*BAK1*	Platelet count	Europeans	13,943	5.43	3.7E-10	NR	2,963	5.98	0.003	0.28

Table 3.5. Significant pleiotropic effects detected with GWAS identified variants in non-Hispanic blacks in the combined NHANES dataset. We performed linear regression with GWAS-identified variants and hematological traits: platelet count, mean platelet volume, and log- transformed triglycerides. Significance was declared at $p<0.05$. Abbreviations: MPV = mean platelet volume

SNP	Gene	Original Study					Combined NHANES Non-Hispanic blacks			
		Original Phenotype	Population	n	β	P-value	Associated Phenotype	n	β	P-value
rs1800562	*HFE*	Mean cell volume	Europeans	13,943	0.37	1.4E-23	Fibrinogen	1,102	0.25	7.5 E-5
rs210135	*BAK1*	Platelet Count	Europeans	13,943	5.43	3.7E-10	MPV	2,960	0.05	0.04

Discussion

Replication of genetic associations is imperative, as initial studies are subject to the overestimated effect sizes referred to as the "winner's curse" (Xiao and Boehnke, 2009; Zhong and Prentice, 2010). We sought to first replicate previously identified fibrinogen associations in an independent African-descent population (non-Hispanic blacks from NHANES). Notably, we replicated two associations, rs2070017 and rs2066784 originally identified by GWAS for fibrinogen, in our study population (Table 3.4). These associations have been reported to be exclusive to individuals of African descent, as the effect allele is rare or not observed in European-descent populations. Our data support the latter observation as we also did not detect these associations in non-Hispanic whites from NHANES (p-value >0.2; see Chapter IV part B).

Variants targeted for replication here were originally identified using a specialized fixed-content 50K BeadChip known as the ITMAT-Broad-CARe (IBC) (Wassel et al., 2010b). Unlike traditional GWAS chips, the IBC BeadChip is hypothesis-driven and only includes markers in genes biologically relevant to cardiovascular disease or related phenotypes (Keating et al., 2008). A major objective of this custom-content chip was to target population-specific tagSNPs in these biologically-relevant genes so that all known common proxies could be captured in the experiment either directly or via linkage disequilibrium (Keating et al., 2008). The data we report in NHANES mirror the design of this BeadChip as

both variants have low frequencies in non-Hispanic whites and are population specific to African-descent samples.

The majority of the literature on the synthesis of fibrinogen has been limited to *FGB*, the rate limiting gene for the production of the fibrinogen poly peptide (Kottke-Marchant K, 2011). Little is known about the impact of genetic variants in the other genes, *FGA* and *FGG*, and the effect they may or may not have on fibrinogen production. The SNPs targeted for replication for plasma fibrinogen levels are located in the *FGA* or *FGB* genes. Although both SNPs are intronic, replication of these results can still be useful as intronic variants have been reported to effect protein synthesis(Law et al., 2007; Tokuhiro et al., 2003). Another possibility is that these intronic SNPs are correlated with the true functional SNP and tag this region in the genome. However, functional studies are needed to confirm role if, at all, these variants have on fibrinogen production.

GWAS fixed-content products are based on a subset of common SNPs in the genome that serve as proxies for SNPs not directly assayed (also known as tagSNPs) (Carlson et al., 2004). The use of tagSNPs is a cost effective alternative to genotyping all known common variants based on linkage disequilibrium patterns throughout the genome. Until recently tagSNPs targeted by GWAS products have been primarily based LD patterns and common variation in European populations, limiting the generalization of GWAS results to other populations.

Despite these challenges, we were able to generalize GWAS-identified variant rs210135 in non-Hispanic blacks for platelet count (Table 3.4). Rs210135 was previously genotyped on GWAS-fixed content platform in Europeans (Soranzo

et al., 2009). We were able to detect this association due to our increase in sample size (n= 2,963) and lack of minor allele frequency population differences between Europeans and non-Hispanic blacks (Fst < 0.01). Notably, this signal contains the gene *BAK1* which encodes a protein that has proapoptotic effects that control the lifespan of platelets (Mason et al., 2007) .

Recent searches for pleiotropy using methods such as the Phenome-wide associations studies (PheWAS) design (Pendergrass et al., 2011) has proven successful (Denny et al., 2010a; Denny et al., 2011). Using a similar approach, we identified two pleiotropic associations with GWAS-identified variants (Table 3.5). Our most significant finding rs1800562, originally reported to have an association with mean cell volume, is associated with increased levels of fibrinogen. This variant is located in the High Iron Fe (*HFE*) gene, which encodes the human hemochromatosis protein. This protein regulates iron absorption, and mutations in this gene lead to high levels iron in the blood resulting in hemochromatosis (Distante, 2006; Distante et al., 2000). The pleiotropic effect we detected with rs1800562 and increased fibrinogen levels is biologically relevant, as increased iron levels are also associated elevated fibrinogen levels in westernized populations(Miura et al., 2006).

An important aspect of replication and generalization of genetic variants is power. In the new era of GWAS most studies have thousands to hundred thousands of samples thereby increasing power to detect small genetic effects throughout the genome. Our study is small compared to the original studies we sought to replicate/generalize (Soranzo et al., 2009; Wassel et al., 2010l). In fact

our sample size is only about 16% of the size used in the Wassel et al GWAS (6,657 African Americans) and 21% of the samples from the Soranzo GWAS (13,943 Europeans). Given our sample size we had >99% power to detect genetic effects comparable with the Wassel et al GWAS with a minor allele frequency of ≥ 0.10 (in our sample) at $\alpha = 0.05$. We similarliy had 80% power to detect the effect originally reported by Soranzo et al in our non-Hispanic blacks given our sample size, minor allele frequency, and $\alpha = 0.05$.

In addition to power, it is important to note that we did not correct for multiple testing. We declared significance at a liberal p-value of > 0.05. Another limitation to this study is that we did not adjust for population stratification. While our study was limited to non-Hispanic blacks we did not perform formally adjust for possible population stratification that could have confounded our results. Ancestry informative markers (AIMs) are currently not available for NHANES; therefore we are limted to stratifying samples by race/ethnicity to avoid population stratification.

To our knowledge we are the first to report the replication of GWAS-identified variants in an African-descent population for fibrinogen levels. Not surprisingly, we were able to generalize GWAS-identified variants with comparable allele frequencies across populations. We are also the first to report pleiotropic effects with GWAS-identified variants for these traits in African Americans. Increased fibrinogen and other hematological traits have been consistently associated with cardiovascular disease, which disproportionately

affects African Americans at a higher rate compared with European descent populations (Segal and Moliterno, 2006; Mensah et al., 2005). Results from this study add to the limited genetic associations studies performed in African-descent populations for fibrinogen and hematological traits with the ultimate goal of finding genetic risk factors of cardiovascular disease.

C. Genetic Modifiers of Fibrinogen Levels in African Americans

Introduction

Fibrinogen levels are highly heritable with reports of up to 50% of the trait variability attributable to genetic factors (Voetsch and Loscalzo.J, 2003). These estimates include genes within the fibrinogen gene cluster (*FGA, FGB,* and *FGG)* as well as other genes that regulate the production of fibrinogen (de Maat, 2005). However, collectively these loci do not account for the expected genetic component for fibrinogen levels. For complex traits such as fibrinogen that are highly variable, the search for novel disease loci that are highly penetrant or accurately predict fibrinogen levels has not been as successful compared to simple Mendelian disorders. Like many traits associated with cardiovascular disease, the role of the environment, several loci in the genome, or the interaction amongst genes and/or with environment can account for some of the heritability yet to be described for fibrinogen levels.

Fibrinogen has three pairs of polypeptides Aα, Bβ, γ that are encoded by three genes: *FGA, FGB,* and *FGG,* respectively(Mosesson, 1998). These polypeptides form disulfide bonds at the near their N-terminus and once connected make up fibrinogen (Zhang and Redman, 1996; Zhang and Redman, 1994). Fibrinogen synthesis is highly dependent on the interactions between fibrinogen polypeptide pairs. Therefore it is highly plausible that interactions between genes within the fibrinogen gene cluster may affect fibrinogen levels. In addition to possible gene-gene interactions, the interplay between genes and

environmental risk factors can explain the inter-individual variability of fibrinogen levels. Risk factors such as smoking, age, lack of exercise, sex, use of contraceptives, and body mass index (BMI) significantly contribute to the variability of fibrinogen levels amongst individuals (de Maat, 2001). However, after accounting for these risk factors, a significant amount of the variance in fibrinogen levels remains unknown.

One of the primary goals of the Population Architecture using Genomics and Epidemiology (PAGE) consortium is to identify genetic modifiers of common disease (Matise et al., 2011). Likewise, the Epidemiologic Architecture of Genes Linked to Environment (EAGLE), which is a part of the PAGE consortium, uses a diverse, cross-sectional survey that is rich in environmental information to identify genetic modifiers. Here we examine the effect of genetic modifiers (gene-gene and gene-environment) with plasma fibrinogen levels in non-Hispanic blacks from NHANES III (previously described). Using 25 SNPs in the fibrinogen cluster, we test for and identify gene-gene interactions within this cluster, as well as gene-environment interactions with these SNPs and known environmental risk factors.

Methods

Study Population

Non-Hispanic blacks with normal fibrinogen levels from NHANES III were available for this study. For more detail on exclusion/inclusion criteria as well as population characteristics please refer to the methods sub section *"Study Population"* and Table 3.1 in part A of this chapter.

SNP selection and Genotyping

A total of 25 tagSNPs in *FGA, FGB,* or *FGG* were genotyped using the Illumina GoldenGate assay. TagSNPs were selected using LDSelect (Carlson et al., 2004) for multiple populations at r^2 >0.80 for common variants (minor allele frequency (MAF) >5%) in three candidate genes (*FGA*, *FGB,* and *FGG*) based on data available for European Americans and African Americans in SeattleSNPs (Crawford et al., 2005). For more information on genotyping and SNP selection please refer to the methods sub section *"SNP selection and genotyping"* in part A of this chapter (page 75).

Fibrinogen measurements

Plasma fibrinogen was measured from blood plasma on participants 40 years and older using the Clauss clotting method (Daum et al., 2000). Descriptive statistics for fibrinogen levels are described in the results sub-section "*Population Characteristics"* and Table 3.1 in part A of this chapter (page 81). Participants with fibrinogen levels greater than >4.0 g/L were excluded from the analysis since extreme measurements can indicate inflammation or recent trauma.

Statistical Methods

Using linear regression, cross product terms (all pair-wise gene-gene or gene-environment terms for 25 fibrinogen variants) were added to the regression models for plasma fibrinogen. All SNPs were coded additively and all models were adjusted for the main effect of the SNP (both the GxG and GxE models) and environmental factor (only the GxE models). In addition to adjusting for

main effects, all analyses were adjusted for known covariates age, BMI, sex, and smoking status. Analyses were limited to self-identified non-Hispanic blacks from NHANES III. All analyses were performed using the Statistical Analysis Software (SAS v.9.2; SAS Institute, Cary, NC) via the Analytic Data Research by Email (ANDRE) portal of the CDC Research Data Center (RDC) in Hyattsville, MD. We consider an interaction term significant at $p<0.05$.

Results

Gene-Gene Interactions

We identified several gene-gene (GxG) interactions within the fibrinogen gene cluster that effect fibrinogen levels in non-Hispanic blacks (Table 3.6). We define a significant GxG interaction here as having significant interaction term ($p<0.05$) from two different genes in the fibrinogen gene cluster. A total of six unique GxG interactions were significant at the $p< 0.05$ level out of 276 total tests performed, which is less than expected by chance alone. Three GxG interactions were associated with decreased fibrinogen levels and three were associated with increased fibrinogen levels (Table 3.6).

Missense variant rs6050 (Thr312Ala) in the *FGA* gene is a variant known to be associated with fibrinogen levels in Europeans. We were unable to detect this association with non-Hispanic blacks in our original single SNP analysis (see results section of part A of this chapter). However, the interaction term rs6050 x rs2227395 (*FGG*) was significantly associated with decreased fibrinogen levels in non-Hispanic blacks ($\beta = 0.15$, p-value = 0.017). Significant interactions

terms with *FGB* variant rs2227395 were consistently observed with SNPs in the *FGA* gene (3/4 tests of interaction; Table 3.6). None of the SNPs with significant GxG interaction terms were significantly associated with fibrinogen levels in the single SNP tests of association (Table 3.6; Appendix P).

Gene-Environment Interactions

We tested for gene-environment (GxE) interactions with fibrinogen cluster variants and sex, age, smoking status, and BMI separately for an association with plasma fibrinogen. All three environmental factors we tested have previously described as significant risk factors for elevated fibrinogen (de Maat, 2001)**.** We were able to detect 11 significant SNP environment interactions at $p < 0.05$ level out of 99 total test performed. Interestingly interaction terms with two SNPs, rs2227434 (*FGB*, Pro337Ser) and rs6063 (*FGG*, Gly712Arg) were consistently associated fibrinogen levels and represent more than half (6/11) of the significant interactions (Table 3.7). These SNPs were not significant in the single SNP test of association and are rare in all NHANES III populations MAF<0.001, which could a result of small sample size (Table 3.7; Appendices N and P).

Two SNP x BMI interactions were associated with fibrinogen levels at $p<0.05$ (Table 3.7). The interaction term rs6050 x BMI was significantly associated with increased fibrinogen levels (Table 3.7). Non-synonymous SNP rs6050 (Thr331Ala) is located in the *FGA* and was associated with decreased levels of serum fibrinogen in non-Hispanic whites but not non-Hispanic blacks in a single SNP test of association (Appendix P). Another interaction term with

BMI, rs2070006 x BMI, was significantly associated with decreased fibrinogen levels (Table 3.7). Rs2070006 is also located in the *FGA* gene; however this SNP was not significant in any NHANES III population in a single SNP model (Appendix P). Excluding rs2227434 and rs6063, there was only one SNP x sex interaction that was significantly associated with increased fibrinogen levels in non-Hispanic blacks, rs6058 x sex (Table 3.7). This SNP was not associated with fibrinogen levels in non-Hispanic blacks in the single SNP test of association (Table 3.6; Appendix P).

There were two SNP x smoking interaction terms that were significantly associated with decreased serum fibrinogen, rs2070033 x smoking and rs2070008 x smoking in non-Hispanic blacks (Table 3.7). Both SNPs are located in the *FGA* gene and were not associated with fibrinogen levels in the single SNP test of association (Appendix P).

Table 3.6. Significant fibrinogen cluster gene-gene interactions in NHANES III non-Hispanic blacks for plasma fibrinogen (n = 621). Using multivariate linear regression, we added cross product terms (all pair-wise SNP-SNP) to the regression models for serum fibrinogen. Abbreviations: SE = standard error.

Gene-Gene Interactions	SNP-SNP Interactions	SNP 1 Main Effect		SNP 2 Main Effect		Interaction Term	
		Beta (SE)	p-value	Beta (SE)	p-value	Beta (SE)	p-value
FGB x *FGG*	rs1800791 x rs1800792	-0.08 (0.04)	0.07	-0.01 (0.04)	0.71	0.20 (0.08)	0.02
FGA x *FGB*	rs2070006 x rs2227395	0.01 (0.03)	0.85	<0.001 (0.03)	0.94	0.14 (0.07)	0.04
FGA x *FGB*	rs2070009 x rs2227395	<0.001 (0.03)	0.94	<0.001 (0.03)	0.94	-0.16 (0.07)	0.02
FGA x *FGG*	rs2070022 x rs1049636	-0.08 (0.04)	0.06	0.06 (0.03)	0.07	-0.18 (0.08)	0.03
FGB x *FGG*	rs2227395 x rs2066861	-0.02 (0.05)	0.61	0.05 (0.03)	0.12	0.17 (0.07)	0.01
FGA x *FGB*	rs6050 x rs2227395	0.02 (0.03)	0.52	<0.001 (0.03)	0.94	-0.15 (0.06)	0.02

Table 3.7. Gene-environment interactions with fibrinogen variants and age, sex, BMI, and smoking status in non-Hispanic blacks (n=621) from NHANES III. Using linear regression we added the multiplicative terms (SNP x environmental factor) to the regression model while simultaneously adjusting for covariates: age, sex, BMI, and smoking status when necessary. Below are the results for the single SNP or environmental factor tests of association as well as the results for the interaction term. Note: there were several values that were suppressed for rs6063 by the the CDC's Research Data Center (RDC) due to low counts (<5).

Interaction	Gene	SNP Main Effect		Environment Main Effect		SNP-Environment Interaction Effect	
		Beta (SE)	p-value	Beta (SE)	p-value	Beta (SE)	p-value
rs6063 *age	*FGG*	Suppressed		0.008 (0.002)	3.69E-07	0.004 (<0.01)	3.69E-07
rs2227434*age	*FGB*	0.26 (0.49)	0.59	0.008 (0.002)	3.76E-07	0.004 (<0.01)	3.76E-07
rs6063*BMI	*FGG*	Suppressed		0.02 (0.003)	4.21E-06	0.01 (<0.01)	4.21E-06
rs2227434*BMI	*FGB*	0.26 (0.49)	0.59	0.02 (0.003)	4.32E-06	0.01 (<0.01)	4.32E-06
rs6050*BMI	*FGA*	0.01 (0.03)	0.52	0.01 (0.003)	5.87E-06	0.01 (0.05)	8.10E-03
rs6063*sex	*FGG*	Suppressed		0.10 (0.04)	0.02	0.05 (0.02)	0.02
rs2227434*sex	*FGB*	0.26 (0.49)	0.59	0.09 (0.04)	0.02	0.05 (0.02)	0.02
rs6058*sex	*FGB*	0.06 (0.05)	0.18	0.10 (0.04)	0.01	0.22 (0.10)	0.03
rs2070033*smoking	*FGA*	0.07 (0.05)	0.16	-0.06 (0.04)	0.14	-0.22 (0.10)	0.03
rs2070006*BMI	*FGA*	0.005 (0.03)	0.85	0.02 (0.003)	6.94E-06	-0.01 (<0.01)	0.03
rs2070008*smoking	*FGA*	-0.01 (0.04)	0.75	-0.05 (0.04)	-0.16	-0.19 (0.01)	0.04

Discussion

In the present study we performed an exhaustive analysis within the fibrinogen gene cluster to detect gene-gene and gene-environment interactions that may affect fibrinogen levels in non-Hispanic blacks from NHANES III. Identifying genetic modifiers of fibrinogen levels is important since the genetic contribution of variable fibrinogen levels at the population level is poorly understood. Additionally, current heritability estimates (h^2, or narrow sense heritability) do not include possible gene-environment interactions. Here using a multiplicative approach, we tested for statistical epistasis or the deviation from linearity in the presence and absence of main effects. We tested for interactions amongst 25 tagSNPs in the fibrinogen gene cluster (gene-gene) as well as tested each SNP with environmental risk factors: age, sex, BMI and smoking status (gene-environment). We observed six gene-gene interactions and eleven gene-environment interactions with an empirical p-value < 0.05.

One criticism of exhaustively testing for interactions is the biological interpretation of the results. The fibrinogen gene cluster is an ideal candidate for identifying gene-gene interactions since each gene is dependent on the other to produce active fibrin in the blood (Zhang and Redman, 1996; Zhang and Redman, 1994). In our gene-gene analysis we detected several interactions between different genes in the cluster (Table 3.6). More importantly we identify several interactions with rs2227395, which is in the *FGB* gene is the known as the rate-limiting gene in fibrinogen production. It is important to note that this SNP did not have a significant main effect in the single SNP test of association in non-Hispanic blacks (Appendix P). Interestingly, all gene-gene interactions we identified were observed in the absence of

main effects (Table 3.6). It is possible that SNPs such as rs2227395 (where we did not observe a significant main effect) are interacting with other SNPs and ultimately affecting fibrinogen levels, an effect that would not be detected at the single SNP level.

Environmental variables are well-known risk factors of elevated fibrinogen levels (de Maat, 2001). It is accepted by many in the field that genes interact with the environment (Manolio et al., 2009). We identified several gene-environment interactions that were associated with fibrinogen levels in non-Hispanic blacks (Table 3.7). Interestingly, we observed similar results for interaction terms containing rs2227434 or rs6063 and environmental factors: age, sex, and BMI (Table 3.7). More often than not identical association results for different SNPs is indicative of high correlation or linkage disequilibrium between SNPs. We attempted to calculate linkage disequilibrium between these SNPs in our study population but failed obtain accurate results due to low genotype accounts for both SNPs (Appendix O). We then attempted to use reference HapMap population YRI to determine LD between these SNPs but like our study population genotype frequencies are too low for an accurate measurement (2003). As previously mentioned, both SNPs are rare in non-Hispanic blacks with minor allele frequencies (MAF) less 0.001 and we only have 5% power to detect these effects given the low MAF in our study population. After evaluating the main effect of the environmental factors we noticed that the association results for these genetic variants and the interaction term with rs2227434 or rs6063 were identical as well (Table 3.7). The association results we observe for these interaction terms may be false positives and actually represent the environmental factors alone, given how

rare these SNPs are. For these reasons we caution the interpretation of these results. On the contrary, we were able to successfully identify two SNP x BMI interactions both of which have significant environmental main effects as well as two SNP x smoking interactions and one SNP x sex interaction.

Two major weaknesses to this study are power and multiple testing. While NHANES III contains over 33,000 study participants, our study population sample size was limited for several reasons. The Centers for Disease Control and Prevention (CDC) only collected DNA samples from a subset of the entire dataset, which includes 7,159 participants from phase 2 of NHANES III. DNA was not collected from participants who reported having hemophilia or chemotherapy within four weeks of collection (Gunter E. et al., 1996). Our outcome variable for this study was limited to plasma fibrinogen which was only collected on a subset of NHANES participants. Plasma fibrinogen levels in NHANES III was only measured for participants > 40 years of age (Gunter E. et al., 1996), which drastically reduces our sample size to 621 non-Hispanic blacks. Due to sample size, on average we were only 70% powered to detect effects that explain less < 0.5% of the trait variability at $\alpha = 0.05$ and MAF of at least 0.05.

One caveat of exhaustively testing all pair-wise interactions is the increase in type I error or false positives. It is important to note for this study that we did not correct for multiple testing, which increases the likelihood of reporting false positives, thereby reducing our confidence of our reported findings.

Despite the small sample sizes, we were able to successfully identify several gene-gene and gene-environment associations at a liberal significance threshold of

$p<0.05$. To date this is the first study that examines the role genetic modifiers have on plasma fibrinogen levels in an African American population. Therefore, all results reported are novel and yet to be reported in the literature. Despite the statistical evidence of epistasis, we cannot confirm the effect these interactions have on fibrinogen levels. Future analyses are needed with larger sample sizes for replication and to distinguish between a true interaction and a spurious false positive result.

CHAPTER IV

GWAS- AND CANDIDATE GENE-IDENTIFIED VARIANTS FOR ELECTROCARDIOGRAPHIC TRAITS AND FIBRINOGEN LEVELS: REPLICATION IN EUROPEAN AMERICANS AND CHARACTERIZATION IN MEXICAN AMERICANS

A. Electrocardiographic traits: Replication in European Americans and Characterization in Mexican Americans

Introduction

Cardiovascular disease (CVD) by definition is the pathogenesis related to the heart and its associated vasculature (Arking and Chakravarti, 2009). An electrocardiogram (ECG) measures the electrical activity of the heart and is a useful tool in assessing cardiac arrhythmias, myocardial infarction, pericarditis, and other cardiac abnormalities (Micheal A Chizner, 2004). The ECG records electrical depolarization and repolarization of the heart and provides information about the electrical activity of the cardiac cycle. Candidate gene studies have identified several genes throughout the genome that effect heart depolarization and conduction measured by ECG traits. Genes that encode voltage gated potassium and sodium channels play an important role in the pathophysiology of long QT syndrome (George, Jr. et al., 1995; Wang et al., 1995b; Wang et al., 1995a; Lai et al., 1994; Splawski et al., 2002; Splawski et al., 1997; Wang et al., 1996). In Chapter II, we identified 13 independent associations in the $NA_v1.5$ sodium channel α subunit gene (*SCN5A*) with ECG traits in African Americans from the Jackson Heart Study (Table 2.2;(Jeff et al., 2011b). Many of the associations identified in this

African American population were novel, and it is unclear if these associations generalize to other populations such as Mexican Americans.

There are over 120 unique trait-SNP associations reported in the National Human Genome in Research Institute's (NHGRI) GWAS catalog for ECG traits in European or Asian descent populations (http://www.genome.gov/gwastudies/). The successes of these GWAS have identified several variants throughout the genome such as *NOS1AP* and *SCN10A* that replicate across several studies (Newton-Cheh et al., 2009; Smith et al., 2009b; Chambers et al., 2010; Arking et al., 2006; Post et al., 2007); (Holm et al., 2010); (Newton-Cheh et al., 2007b) (Marroni et al., 2009) (Pfeufer et al., 2010). Most of these studies, discovery and replication, have been limited to European and Asian descent populations. To date, there have been no studies confirming whether these associations generalize in Mexican Americans.

One of the primary goals of Epidemiologic Architecture for Genes Linked to the Environment (EAGLE) as part of the Population Architecture using Genomics and Epidemiology (PAGE) study is to characterize GWAS-identified variants in a diverse cross-sectional survey, the National Health and Nutrition Examination Surveys (NHANES) (Matise et al., 2010). Using the Third NHANES (Elaine W et al., 1996), we first generalize *SCN5A* candidate gene associations identified in African Americans to non-Hispanic whites (a European descent population) and Mexican Americans. Additionally, we determine the extent to which GWAS-identified variants replicate in non-Hispanic whites and generalize to Mexican Americans.

Methods

Study Population

All analyses were performed in non-Hispanic whites and Mexican Americans from NHANES III. Ascertainment of NHANES III and method of DNA collection have been previously described (Crawford et al., 2006; Chang et al., 2009; Steinberg et al., 1997). The National Health and Nutritional Examination Surveys are cross-sectional surveys conducted across the United States by the National Center for Health Statistics (NCHS) at the Centers for Disease Control and Prevention (CDC). NHANES III was conducted between 1988-1990 (phase 1) and 1991-1994 (phase 2) (1996; 2004) as a complex survey design that over-sampled minorities, the young, and the elderly. All NHANES have interviews that collect demographic, socioeconomic, dietary, and health-related data. Also, all NHANES study participants undergo a detailed medical examination at a central location known as the Mobile Examination Center (MEC). Beginning with phase 2 of NHANES III, DNA samples were collected from study participants aged 12 years and older.

ECG traits

Electrocardiograms (ECGs) were recorded on men and women in the mobile examination center (MEC) using a standard 12-lead resting ECG. ECGs were recorded using the Marquette MAC 12 (Marquette Medical Systems, Inc, Milwaukee, Wisconsin (U.S. DHHS, 1996). NHANES III 12-lead ECG data were recorded with eight independent components of the 12 standard leads simultaneously. ECG data were also sampled at 250 samples per second per channel, giving the availability of multiple

simultaneous ECG leads for analysis. ECGs were only performed on individuals 40 years of age and older. NHANES III is the only contemporary NHANES that collected ECG data.

SNP Selection and Genotyping

DNA was extracted from crude cell lysates from lymphoblastoid cell lines established for NHANES III participants aged 12 and over as part of Genetic NHANES (Steinberg et al., 1997). *SCN5A* SNPs were previously identified in African Americans from the Jackson Heart Study (Jeff et al., 2011b). In this analysis, SNPs were selected based on the linkage disequilibrium patterns in African Americans (Bush et al., 2009). Twelve SNPs were associated with ECG traits at $p<1.0x\ 10^{-4}$ in Jackson Heart Study, and to date, five of these SNPs have been successfully genotyped in NHANES III.

We also selected SNPs that have been previously identified by GWAS for all ECG traits using the NHGRI genome catalog (http://www.genome.gov/gwastudies/). A total of 46 SNPs that were previously identified from GWA studies or well-studied in candidate gene analysis were selected for genotyping, and 15 of the nominated SNPs were available for this study. We did not consider SNPsfrom publicatons that did not report effect sizes and coded alleles from the original analysis (7 SNPs total; (Newton-Cheh et al., 2007b; Vasan et al., 2007; Smith et al., 2009b)).

All genotyping was performed in the Center for Human Genetics Research DNA Resources Core and the laboratory of Dr. Jonathan Haines. All SNPs were genotyped using either Sequenom's iPLEX Gold assay on the MassARRAY platform (San Diego, CA) or Illumina's BeadXpress. All genotype data reported here passed CDC quality

control metrics and are available for secondary analysis through CDC. Aggregate data will be available through dbGaP under EAGLE/PAGE.

Statistical Methods

All analyses were performed using the Statistical Analysis Software (SAS v.9.2; SAS Institute, Cary, NC) either locally or via the Analytic Data Research by Email (ANDRE) portal of the CDC Research Data Center (RDC) in Hyattsville, MD. All traits followed a normal distribution and descriptive statistics are reported in Table 4.1. Participants with QRS duration >120 m/sec were excluded from all analysis. Using standard linear regression and assuming an additive genetic model, we tested each SNP for an association with ECG traits: P duration, QRS duration, QT interval and heart rate. We did not test *SCN5A* variants for an association with heart rate and QT interval since these SNPs were not associated with these traits in the original analysis. All tests were limited to non-Hispanic whites and Mexican Americans and adjusted for age and sex. All analysis were stratified by self-reported race/ethnicity. Significance was declared at a liberal threshold of $p<0.05$ unadjusted for multiple testing.

Results

Population Characteristics

Our study population characteristics in NHANES III non-Hispanic whites and Mexican Americans were comparable to JHS African Americans (Table 4.1). Non-Hispanic whites on average were younger (average age = 61.5 yrs) compared to JHS African Americans and Mexican Americans (Table 4.1). However, age was not statistically different between Mexican Americans and JHS African Americans (Table

4.1). There were differences between NHANES III populations and Jackson Heart Study (JHS) African Americans (Table 4.1). In addition to age, P duration, PR interval, and QRS duration all varied between non-Hispanic whites and JHS African Americans (Table 4.1).

Generalization of Candidate Gene-identified Variants

We genotyped five candidate SNPs in the *SCN5A* gene in non-Hispanic whites and Mexican Americans from NHANES III that were previously associated in JHS African Americans (see Chapter II(Jeff et al., 2011b). Using the same approach as the original study, we performed tests of association to generalize JHS associations in non-Hispanic whites and Mexican Americans for ECG traits. We considered a SNP generalized if it had consistent direction of effect and was statistically significant at $p<0.05$. We performed a total of 15 tests of association for three ECG traits: P duration, PR interval, and QRS duration (five SNP tests of association per trait).

We were able to generalize two ECG trait-SNP associations in non-Hispanic whites and Mexican Americans (Table 4.2). In JHS African Americans, rs7374138 was associated with decreased P duration (β = -1.5, p=3.6R-03), PR interval (β = -4.0, p=2.4E-05) and QRS duration (β = -1.3, p=6.2E-06) (Chapter II, Table 2.2). For P duration, the association generalized in non-Hispanic whites (β = -2.50, p=0.006). For PR interval, the association generalized to Mexican Americans (β = -5.57, p< 0.005; Table 4.2). Interestingly, the magnitude of the effect is larger compared to JHS results for both NHANES III populations. Though not statistically significant, rs7374138 had a

consistent direction of effect in non-Hispanic whites and Mexican Americans for the other ECG traits (Table 4.2).

Generalization of GWAS-identified variants

A total of 15 GWAS identified SNPs were genotyped and tested for an association with ECG traits in Mexican Americans. We declared a generalization as being statistically significant and having the same direction of effect with the previous study. Notably, rs12143842 located in *NOS1AP*, which has been replicated several times in Europeans, generalizes to Mexican Americans (β= 4.03, p= 0.02; Table 4.3). Regardless of significance, half of the SNPs tested also have a consistent direction of effect compared to the original study (Table 4.3).

Replication

We genotyped 15 GWAS identified variants in non-Hispanic whites for replication and generalization. As previously mentioned, we did not consider SNPs from publications that did not report effect sizes and coded alleles from the original analysis (seven SNPs), resulting in total of eight SNPs considered here. Replication was defined by statistical significance ($p < 0.05$) and consistent direction of effect with the same SNP in the same ancestral population. Seven SNPs were tested for replication with QT interval, and one SNP was test for replication with PR interval in NHANES III non-Hispanic whites. Previously identified variants rs12143842 and rs12029454, located in *NOS1AP*, replicated with QT interval in NHANES III non-Hispanic whites ($p < 0.001$; Table 4.3). These two associations are not in linkage disequilibrium in CEU reference population ($r^2=0.03$) and likely represent independent effects. Another SNP, rs11970286,

located in *PLN*, replicated in non-Hispanic whites and was associated with increased QT interval (β=2.97, p= 0.01; Table 4.3). Consistent with other findings in NHANES III, the effect size was larger compared to the original study. Regardless of significance, several associations (63%) had a consistent direction of effect and only three (3/8) trended in the opposite direction compared to the original study (Table 4.3).

Table 4.1 Study population characteristics. Means and standard deviations were calculated for ECG traits and age, by NHANES III population. Abbreviations: non-Hispanic white (NHW), Mexican Americans (MA), standard deviation (SD).

Trait	JHS African Americans		NHANES III non-Hispanic whites		NHANES III Mexican Americans	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**
Age (yrs)	56.5	±11.73	61.5	±13.51	55.66	±12.10
P Duration (msec)	118.5	±13.08	109.8	±18.24	109.8	±12.47
PR Interval (msec)	171.6	±33.02	162.6	±26.83	157.2	±23.01
QRS duration (msec)	92.3	±10.12	95.04	±10.25	95.85	±10.90

Table 4.2. Generalization of *SCN5A* variants in non-Hispanic whites and Mexican Americans from NHANES III. Linear regression adjusted for age and sex was performed on three previously identified associations in non-Hispanic whites and Mexican Americans.

SNP	Phenotype	JHS African Americans (n = 3,054)		NHANES III non-Hispanic whites (n = 1,051)		NHANESIII Mexican Americans (n = 580)	
		β	P-value	β	P-value	β	P-value
rs7637849	P Duration	-1.70	1.1E-3	2.93	0.03	-1.35	0.40
rs7374138	P Duration	-1.50	3.6E-3	-2.50	0.006	-1.10	0.20
rs11129796	P Duration	-0.81	0.18	8.41	0.14	0.07	0.96
rs7629265	P Duration	-3.20	7.5E-5	10.89	0.34	-2.53	0.65
rs6768664	P Duration	0.60	0.18	-1.06	0.19	-0.57	0.40
rs7637849	PR Interval	-4.20	2.3E-5	1.45	0.46	-2.92	0.31
rs7374138	PR Interval	-4.00	2.4E-5	-2.24	0.09	-5.57	3.4E-4
rs11129796	PR Interval	-3.50	2.4E-4	-3.01	0.71	-0.93	0.71
rs7629265	PR Interval	-7.80	2.4E-7	2.93	0.84	17.67	0.08
rs6768664	PR Interval	3.00	2.4E-4	-1.3	0.20	0.21	0.86
rs7637849	QRS Duration	-0.76	1.2 E-2	3.06	0.32	0.76	0.30
rs7374138	QRS Duration	-1.30	6.2E-6	-1.20	0.09	-1.90	0.16
rs11129796	QRS Duration	-1.10	2.5E-3	-0.41	0.402	0.95	0.42
rs7629265	QRS Duration	-1.50	1.8E-3	-1.00	0.85	5.99	0.20
rs6768664	QRS Duration	0.67	9.5E-3	-0.30	0.45	-0.35	0.55

Table 4.3 Replication and Generalization of European GWAS-identified variants in NHANES III Non-Hispanic whites and Mexican Americans, respectively. We performed a standard linear regression adjusted for age and sex for eight GWAS-identified variants with QT or PR Intervals. Abbreviations: NR= original study did not report (NR) data.

SNP	Gene(s)	Phenotype	Original Study			NHANES III Non-Hispanic whites (n=1050)		NHANES III Mexican Americans (n =580)	
			Original Population	β	P-value	β	P-value	β	P-value
rs37062	*CNOT1*	QT Interval	European	1.75	3.0E-25	-1.02	0.44	1.95	0.17
rs12143842	*NOS1AP*	QT Interval	European	3.15	2.0E-78	4.26	1.0E-3	4.03	0.02
rs7188697	*NDRG4*	QT Interval	European	1.66	7.0E-25	0.98	0.45	2.13	0.13
rs2461751	Intergenic	PR Interval	Pacific Islanders	4.54	8.0E-6	-0.15	0.92	-0.78	0.59
rs11970286	*PLN*	QT Interval	European	1.64	2.0E-24	2.97	0.01	2.00	0.23
rs12029454	*NOS1AP*	QT Interval	European	2.98	3.0E-45	5.25	9.0E-4	-0.80	0.64
rs2074238	*KCNQ1*	QT Interval	European	7.88	3.0E-17	-2.88	0.18	-0.66	0.87
rs4725982	*KCNH2*	QT Interval	European	1.58	5.0E-16	2.04	0.14	-0.19	0.89
rs882300	Intergenic	PR Interval	European	NR	3.2 E-7	0.78	0.50	0.52	0.74
rs10498091	Intergenic	Echo. Vent. Traits	European	NR	6.00 E-6	-0.87	0.57	-2.77	0.19
rs10504543	*KCNB2*	Echo. Vent. Traits	European	NR	5.00E-06	0.12	0.92	-0.04	0.97
rs10507380	*RPL21*	QT Interval	European	NR	8.00E-06	-1.01	0.54	4.28	0.13
rs1379659	*SLIT2*	Echo. Vent. Traits	European	NR	1.00E-7	Not tested due to small sample size			
rs2008242	*MSX1*	PR Interval	Pacific Islander	NR	3.00 E-6	-0.01	0.99	-0.15	0.96
rs7512898	NR	PR Interval	Pacific Islander	NR	5.00E-6	-0.002	0.99	-0.24	0.86

Discussion

To identify genetic risk factors that contribute to ECG trait variability, we performed a candidate gene study of *SNC5A* variation in the Jackson Heart Study to determine the extent to which African American genetic risk factors generalize in other populations (Chapter II). We then generalized two *SCN5A* associations in NHANES III with P duration and PR interval in European Americans and Mexican Americans, respectively. Additionally, we replicated and generalized *NOS1AP* variants originally identified by GWAS in non-Hispanic whites and Mexican Americans, respectively. The characterization of genetic variation in multiple populations adds to the understanding of the role of genetics in cardiovascular disease on a population level.

Generalization of SCN5A associations

The $NA_v1.5$ sodium channel α subunit is encoded by *SCN5A* which is the predominant α-subunit expressed in the heart (George, Jr. et al., 1993; Gellens et al., 1992). Mutations in *SCN5A* have been previously described for cardiac conduction disease, long QT syndrome, Brugada syndrome, and other life threatening arrhythmias (Wang et al., 1995c; Bezzina et al., 1999; Schott et al., 1999; Chen et al., 1998). In Chapter II, we describe *SCN5A* variation in greater than 3,000 African Americans from the Jackson Heart Study. We identified 13 associations with ECG traits in the *SCN5A* region. In the present study, we examined the generalizability of these SNPs in other populations and generalized two associations. One association, rs7374138 was associated with decreased P duration in non-Hispanic whites and decreased PR interval in Mexican

Americans. This SNP is located intron region of *SCN5A* and to date has not been described in non-Hispanic whites or Mexican Americans.

Over 90% of the SNPs tested (28 out of 30 associations) did not generalize in our study. Of these, six and seven associations in non-Hispanic whites and Mexican Americans, respectively, had consistent directions of effect regardless of significance compared to African Americans from the JHS. The lack of generalizability could be due to power and/or low minor allele frequencies. Indeed, the minor allele frequencies (MAF) for four *SCN5A* variants were rare (MAF <0.05) in non-Hispanic whites and Mexican Americans, effectively reducing the power for these tests of association (Appendix S). Overall, we were inadequately powered (<50%) to detect JHS effects in non-Hispanic whites and Mexican Americans for *SCN5A* variants with the exception of rs7374138 (MAF >0.10in either population, Appendix R).

The lack of generalizability could also be due to differing patterns of linkage disequilibrium, which can also affect the power to detect an association. In both GWAS and in candidate gene studies such as the one performed in JHS for *SCN5A*, the majority of associated SNPs are tagSNPs. That is, the associated SNPs are not the functional or true "risk" SNP but are in linkage disequilibrium with the un-genotyped functional variant. It is known that genome-wide linkage disequilibrium patterns vary across populations (Bush et al., 2009). Thus, the associated or tagSNPs identified in *SCN5A* in African Americans may not tag or be in linkage disequilibrium with the true functional or risk variants in non-Hispanic whites and Mexican Americans.

Replication and generalization of ECG trait GWAS

There have been several GWAS performed for ECG traits in recent years, many of which replicate known ion-channel polymorphisms and as of recently variants in the *NOS1AP* region. Here we tested eight of these previously identified variants with two ECG traits (PR and QT interval), in non-Hispanic whites and Mexican Americans (a total of 32 tests). We were able to replicate three associations in non-Hispanic whites and generalize one association in Mexican Americans.

Replication was defined in this analysis as having a consistent direction of effect and statistical significance in the same population as the original study. Using this definition we were able to replicate three GWAS identified in non-Hispanic whites. We replicate two independent associations in *NOS1AP* and one association in *PLN* for QT interval. Additionally, we were able to generalize one *NOS1AP* variant rs12143842 in Mexican Americans.

We were significantly underpowered to replicate/generalize GWAS identified variants in NHANES III non-Hispanic whites and Mexican Americans. We performed power calculations using the effect size from the original study to determine our power to detect the same effects given our sample size and minor allele frequency. For the SNPs that did not generalize we had <60% power to detect these effects in non-Hispanic whites and <45% power to detect these effects in Mexican American for any ECG trait. Despite this limitation, we were able to successfully generalize associations with large effects that we were adequately powered (>80%) to detect in non-Hispanic whites.

Strengths and Limitations

One limitation to the present study is sample size. The sample sizes in the NHANES III populations are less than a third the size compared to original studies. As previously mentioned, our small sample size limited our power to replicate and generalize previous associations, specifically those with small effect sizes. The overestimation of effect sizes is often observed in the discovery analyses and referred to as the "winner's curse" (Xiao and Boehnke, 2009; Zhong and Prentice, 2010). In our analysis we observed the opposite of the winner's curse, with effect sizes larger compared to the original study despite small sample size. This could be evidence of spurious associations and further analysis in a study population with larger sample sizes is needed to confirm these results.

The need for future studies with larger sample sizes is evident from this work. To date, NHANES III is the only NHANES dataset that collected ECG traits. The potential findings or lack thereof from this work emphasizes the importance of collecting these measurements. Since NHANES collection is continual, measuring ECG traits would be very useful for future analysis and possible meta-analysis. In addition to the NHANES datasets, there are other study sites within the PAGE consortium with ECG measurements in diverse populations such as African Americans (ex. Coronary Artery Risk Development in Young Adults (CARDIA) and Atherosclerosis Risk in Communities (ARIC)).

A major limitation related to these efforts to generalize across populations and across studies is the lack of standard publication or reporting guidelines. That is, seven

associations reported in the literature did not contain data related to the direction of the effect. Our definition of generalization requires statistical significance and consistent direction of effect compared with the original report. For these seven variants, we are limited to reporting the NHANES III results of tests of association, but we cannot make definitive statements concerning generalization across populations.

Using a diverse cohort, NHANES III, we successfully replicated and generalized previous associations for ECG traits. More importantly we replicated and generalized rs12143842, which is located in the *NOS1AP* gene in non-Hispanic whites and Mexican Americans. To date the *NOS1AP* SNPs have no known function; however these associations have successfully replicated in several European descent populations. This is the first time these associations have been described in Mexican Americans. Despite the identification of the *NOS1AP* variant by multiple studies, the extent to which this gene affects QT interval is still poor understood.

Conclusions

Although we were underpowered to detect most associations, we were successfully able to generalize *SCN5A* variant rs7374138 in our study populations. We were also able to confirm *NOS1AP* variants in NHANES III. The two variants we identify have no obvious function; therefore further fine-mapping of this region is warranted to determine the role of these variants with QT interval. This study also suggests that genetic variants that replicate/generalize can be prioritized for clinical use given the generalizability of these associations in multiple populations. Overall, these

data server as a starting point for future genetic association studies of ECG traits in Mexican Americans and other diverse populations.

B. Plasma Fibrinogen and Hematological Traits: Replication in European Americans and Generalization in Mexican Americans[4]

Introduction

African Americans and Hispanics have more risk factors for cardiovascular disease (CVD) putting them at higher risk compared to European Americans (Mensah et al., 2005; Winkleby et al., 1998). Coagulation factors such as fibrinogen and other hematological traits are quantitatively measured from the blood and are associated with CVD risk (Thompson et al., 1995; Meade et al., 1986). Heritability estimates report between 0.37-0.89 of the trait variability for fibrinogen and other hematological factors attributable to genetic factors (Voetsch and Loscalzo.J, 2003; Pennington et al., 2001); (Evans et al., 1999). The accuracy of current heritability estimates studies is unclear, since most of the heritability studies were limited to familial data (mostly twin studies) in European descent populations and do not include possible gene-environment interactions that effect trait variability. The disparity of CVD risk disproportion could be due to genetic differences between populations for intermediate traits such as fibrinogen. It is well-known that there are significant differences in allele frequencies and linkage disequilibrium patterns across populations. Despite this disparity most genetic association studies, including heritability studies for fibrinogen and hematological phenotypes have been limited to Europeans descent populations.

[4] Partially adapted from: Jeff JM, Brown-Gentry K, Crawford DC. Replication and characterization of genetic variants in the fibrinogen gene cluster with fibrinogen levels and hematological traits in the Third National Health and Nutrition Examination Survey In Press: *Thrombosis and Haemostasis*

Genome wide association studies (GWAS) and candidate gene studies have successfully identified several variants with fibrinogen. To date there are 31 GWAS in the NHGRI catalog of genome wide association studies for plasma fibrinogen and/or hematological traits (http://www.genome.gov/gwastudies/). Despite the success of these studies, the majority of these GWAS were performed in European descent populations and none have been reported for Mexican Americans. As previously mentioned in Chapter III, the frequency of variants in the coagulation pathway vary across populations (Rosendaal et al., 1998; Grody et al., 2001; Grody, 2003; McGlennen and Key, 2002). The extent to which these variants generalize to other populations is important. In addition to generalization, replication of genetic associations is imperative, as initial studies are subject to the overestimated effect sizes referred to as the "winner's curse" (Xiao and Boehnke, 2009; Zhong and Prentice, 2010). With the exception of genes that encode coagulation factors such as Factor V Lieden, replication of associations with fibrinogen and hematological phenotypes has been inconsistent. This is mainly due to different study designs, genetic and/or trait heterogeneity, and differences in environmental exposures across studies.

The Population Architecture using Genomics and Epidemiology (PAGE) study is a large collaborative effort that aims to examine whether GWAS-identified variants generalize to ethnically diverse population-based studies (Matise et al., 2011). The Epidemiologic Architecture for Genes Linked to the Environment (EAGLE) is a member of the PAGE consortium and uses data from a cross-sectional survey, the National Health and Nutrition Examination Survey (NHANES), to identify and describe the genetic architecture of GWAS identified variants. In the present study, we first use NHANES III

to identify and replicate genetic associations with 25 fibrinogen cluster variants with fibrinogen and other hematological traits measured in non-Hispanic whites and Mexican Americans. Then we tested for and identify gene-gene interactions within this cluster, as well as gene-environment interactions with these SNPs and known environmental risk factors. Then we sought to replicate/generalize two fibrinogen cluster variants that were not a part our original study in NHANES IIII in non-Hispanic whites and Mexican Americans from NHANES 99-02. Lastly, using a combined NHANES dataset we replicate and/or generalize previous GWAS associations in non-Hispanic whites and Mexican Americans for fibrinogen and test for possible pleiotropic effects with triglycerides, mean platelet volume, and platelet count.

Methods

Study Population

The National Health and Nutritional Examination Surveys are cross-sectional surveys conducted across the United States by the National Center for Health Statistics (NCHS) at the Centers for Disease Control and Prevention (CDC). Ascertainment of NHANES III and method of DNA collection have been previously described (Crawford et al., 2006; Chang et al., 2009; Steinberg et al., 1997). NHANES III was conducted between 1988-1990 (phase 1) and 1991-1994 (phase 2) (1996; 2004) as a complex survey design that over-sampled minorities, the young, and the elderly. To do so, they oversampled non-Hispanic blacks, Mexican-Americans, the young, and the elderly. In 1999 NHANES became a yearly survey and ~8000 new DNA samples were extracted from blood between 1999 and 2002 (NHANES 99-02). All NHANES have interviews

that collect demographic, socioeconomic, dietary, and health-related data. Also, all NHANES study participants undergo a detailed medical examination at a central location known as the Mobile Examination Center (MEC). The medical examination includes the collection of physiological measurements by CDC medical personnel and blood and urine samples for laboratory tests.

All procedures were approved by the CDC's Ethics Review Board and written informed consent was obtained from all participants. This candidate gene association study was approved by the CDC's Ethics Review Board (protocols #2003-08 and #2006-11) and the University of Washington's Institutional Review Board (IRB #23667; HSRC D committee). Because no identifying information was accessed by the investigators, this study was considered exempt from Human Subjects by Vanderbilt University's Institutional Review Board (IRB #061062; HS2 committee).

For this analysis, 4,167 and 5,880 samples from self-described non-Hispanic whites and Mexican Americans from NHANES III and NHANES 1999-2002, respectively, were available for genotyping. We refer to this dataset as combined NHANES, which collectively contain 10,047 samples (n=6,414 non-Hispanic whites and 3,633 Mexican Americans). All participants were at least 18 years of age, and participants with measurements outside of the normal range for any trait were excluded from the analysis, as extreme measurements for these traits can indicate inflammation or recent trauma.

Fibrinogen Measurements

Plasma fibrinogen was measured from blood plasma on participants 40 years and older using the Clauss clotting method (Daum et al., 2000). Descriptive statistics for fibrinogen levels are described in Table 4.4. Participants with fibrinogen levels greater than >4.0 g/L were excluded from the analysis since extreme measurements can indicate inflammation or recent trauma.

Hematological Measurements

Blood was collected from all participants with the exception of those who reported having hemophilia or chemotherapy in the past four weeks (Gunter E. et al., 1996). Complete blood counts (CBCs) were performed on qualifying NHANES participants using the Beckman Coulter method, a method that sizes and counts particles by using measurable changes in electrical resistance produced by nonconductive particles suspended in an electrolyte (Daum et al., 2000). A quantitative, automated, differential cell counter was used to calculate exact values for hematological measurements (Gunter E. et al., 1996). Serum triglycerides were measured using standard enzymatic methods. Descriptive statistics for hematological traits and serum triglycerides are described in Table 4.4.

SNP Selection and Genotyping

DNA was extracted from crude cell lysates from lymphoblastoid cell lines established for NHANES III participants aged 12 and over as part of Genetic NHANES (Steinberg et al., 1997). TagSNPs were selected using LDSelect (Carlson et al., 2004) for multiple populations at r^2 >0.80 for common variants (minor allele frequency (MAF) >5%) in three candidate genes (*FGA*, *FGB*, and *FGG*) based on data available for European Americans and African Americans in SeattleSNPs (Crawford et al., 2005). A total of 25 fibrinogen cluster SNPs were genotyped using the Illumina GoldenGate assay (as part of a custom 384 oligonucleotide pool assay (OPA) by the Center for Inherited Disease Research (CIDR) through the National Heart Lung and Blood Institute's Resequencing and Genotyping Service. The data presented in this analysis were limited to non-Hispanic whites and Mexican Americans from NHANES III. All genotype data reported here were deposited into the NHANES III Genetic database and are available for secondary analysis through CDC.

Using the NHGRI genome catalog (http://www.genome.gov/gwastudies/), we selected SNPs that have been previously identified by GWAS for the following traits: fibrinogen, mean platelet volume, red blood cell volume, platelet aggregation, mean corpuscular hemoglobin content, white blood cell count, red blood cell count and red blood cell volume. A total of 32 SNPs that were previously identified from GWAS or well-established in candidate gene studies were selected for genotyping (Appendix Q), and four of the nominated SNPs were available for this study. Genotyping was performed in the Center for Human Genetics Research DNA Resources Core and the laboratory of

Dr. Jonathan Haines. All SNPs were genotyped using either Sequenom's iPLEX Gold assay on the MassARRAY platform (San Diego, CA) or Illumina's BeadXpress. All genotype data reported here passed CDC quality control metrics and are available for secondary analysis through CDC. Aggregate data will be available through dbGaP under EAGLE/PAGE.

Statistical Methods

All analyses were performed using the Statistical Analysis Software (SAS v.9.2; SAS Institute, Cary, NC) either locally or via the Analytic Data Research by Email (ANDRE) portal of the CDC Research Data Center (RDC) in Hyattsville, MD. Analyses were limited to self-identified non-Hispanic whites and Mexican Americans who were at least 18 years of age. Participants with measurements outside of the normal range for any trait were excluded from the analysis, as extreme measurements for these traits can indicate inflammation or recent trauma. Pair-wise correlations for each trait were calculated using Pearson's correlation coefficient. None of the traits were correlated (defined as $r>0.50$; Appendix M). Descriptive statistics for all traits are reported in Table 4.4. Coded allele frequencies for fibrinogen cluster SNPs are reported in Appendix N. All traits followed a normal distribution with the exception of triglycerides, which was uniformly transformed by using the natural logarithm.

Pair-wise linkage disequilibrium was calculated for all 25 fibrinogen cluster SNPs using Haploview (Appendix O) (Barrett et al., 2005). Single-locus tests of association were performed using linear regression for each fibrinogen SNP and each hematological trait in NHANES III: plasma fibrinogen, mean platelet volume, white blood cell count,

platelet distribution width, log-transformed serum triglycerides, and platelet count. For each test of association, we assumed an additive genetic model. Analyses were performed un-weighted and unadjusted and adjusted for age, sex, body mass index (BMI) and current smoking status (yes/no), and results were plotted using Synthesis View (Pendergrass et al., 2010b). Analyses were repeated to include time to last meal (in hours) as a covariate (data not shown). Current smoking status was determined by the question "Do you smoke cigarettes now?" and cotinine levels (>15 ng/ml). We also tested two additional SNPs identified by candidate gene studies in NHANES 99-02. Using the same approach described for NHANES III, we tested these two SNPs for an association with fibrinogen levels in non-Hispanic whites and Mexican Americans from NHANES 99-02.

Using standard linear regression, assuming an additive model, we tested two GWAS-identified variants for an association with fibrinogen, one GWAS-identified variant with platelet count, and one GWAS-identified variant with mean cell volume in the combined NHANES dataset (NHANES III and NHANES 99-02). Since we did not test mean cell volume in this analysis, this SNP was limited our analysis of pleiotropic effects with fibrinogen and the aforementioned hematological traits. To identify possible pleiotropic effects we tested GWAS-identified variants for fibrinogen with an association with mean platelet volume, platelet count, and log-transformed triglycerides, and a GWAS-identified variant for platelet count was tested for an association with fibrinogen, triglycerides, and mean platelet volume. All tests were limited to non-Hispanic whites or Mexican Americans and adjusted for age, sex, body mass index (BMI), and smoking status. Current smoking status was determined using the same criteria as NHANES III.

We used a standard linear regression to perform all gene-gene (GxG) and gene-environment (GxE) interaction analyses. Cross product terms (all pair-wise gene-gene or gene-environment terms for 25 fibrinogen variants) were added to the regression models for plasma fibrinogen in non-Hispanic whites and Mexican Americans from NHANES III. All SNPs were coded additively and all models were adjusted for the main effect of the SNP and environmental factor. In addition to adjusting for main effects, all analyses were adjusted for known covariates age, BMI, sex, and smoking status. Analyses were limited to self-identified non-Hispanic blacks from NHANES III. All analyses were performed using the Statistical Analysis Software (SAS v.9.2; SAS Institute, Cary, NC) via the Analytic Data Research by Email (ANDRE) portal of the CDC Research Data Center (RDC) in Hyattsville, MD. We consider an interaction term significant at $p<0.05$.

Results

Novel Associations

We identified two novel associations between genetic variants and plasma fibrinogen in non-Hispanic whites and Mexican Americans from NHANES III (Figure 4.1, Table 4.5). In non-Hispanic whites, intronic *FGB* rs2227395 ($\beta = -0.08$, $p=0.0007$; Table 4.5) was associated with decreased plasma fibrinogen levels. *FGB* rs2227395 is in strong linkage disequilibrium (LD) ($r^2= 0.86$; Appendix O) with rs4220 (previously identified), and thus likely represents the same effect. *FGA* rs2070022, located in the 3′ untranslated region of the gene, was associated with decreased plasma fibrinogen levels in Mexican Americans ($\beta = -0.07$; Table 4.5). This association accounted for 1.0% of the trait variability in Mexican Americans.

In addition to testing for an association with plasma fibrinogen levels, we tested for associations between the fibrinogen gene cluster variation and traits measured in the CBC: platelet count, mean platelet volume, platelet distribution width, and white blood cell count in non-Hispanic whites and Mexican Americans from NHANES III (Figure 4.2). We identified six significant associations with fibrinogen SNPs and other hematological traits (Figure 4.2; Table 4.5). *FGA* rs6050 (Thr312Ala) and *FGG* rs2066879 were associated with decreased (β = -0.03; p = 0.04) and increased (β = 0.77; p = 0.02) platelet distribution width in non-Hispanic whites, respectively. Increased platelet distribution width was associated with rs2227425 in Mexican Americans (Table 4.5). Non-synonymous *FGG* rs4220 (Arg478Lys) and *FGB* intronic rs2227395 were both associated with decreased white blood cell count in Mexican Americans (β = -0.28, p = 0.006 and β = -0.23, p = 0.02, respectively). We identified a significant association with rs1800792, located in the 5′ flanking region of FGG, and decreased platelet count in Mexican Americans (Figure 4.2; Table 4.5). There were no significant associations observed between mean platelet volume and fibrinogen SNPs in any subpopulation (Figure 4.2).

There were two significant associations with serum triglycerides in Mexican Americans. Intronic *FGG* rs2066860 was associated with increased serum triglycerides and 5′ flanking *FGG* rs1800792 was associated with decreased triglyceride levels (Figure 4.2; Table 4.5). Our primary analysis included participants regardless of fasting status, which can affect triglyceride levels in individual participants. To account for variability due to fasting status, we performed an additional analysis with inclusion of time to last

meal as a covariate in the models. Inclusion of fasting status did not yield different association results compared to the original analysis (data not shown).

Table 4.4 Study population characteristics and hematological trait descriptive statistics for non-Hispanic whites and Mexican Americans in NHANES. Unweighted means (± standard deviations) or percentages are given for demographic, plasma fibrinogen, and hematologic traits, by subpopulation for adults in phase 2 of NHANES III. Abbreviations: non-Hispanic white (NHW), and Mexican American (MA).

Variable	*non-Hispanic Whites*						*Mexican Americans*					
	NHANES III		**NHANES 99-02**		**Combined NHANES**		**NHANES III**		**NHANES 99-02**		**Combined NHANES**	
	n	**Mean (SD)**	**n**	**Mean (SD)**	**n**	**Mean (SD)**	**n**	**Mean (SD)**	**n**	**Mean (SD)**	**n**	**Mean (SD)**
Plasma Fibrinogen (g/L)	1,452	2.93 (±0.50)	1,809	3.31 (±0.45)	3,262	3.14 (±0.51)	696	2.97 (±0.50)	763	3.29 (±0.47)	1,459	3.14 (±0.51)
Platelet Count	2,384	252 (±58)	3,945	258 (±62)	6,410	260 (±67)	1,732	266 (±62)	1,848	267 (±61)	3,628	270 (±68)
Platelet Distribution Width (mean %)	2,378	16.5 (±0.47)	N/A	N/A	N/A	N/A	1,729	16.4 (±0.54)	N/A	N/A	N/A	N/A
Mean Platelet Volume (fL)	2,407	8.43 (±0.92)	4,000	8.22 (±0.87)	6,407	8.31 (±0.89)	1,748	8.57 (±0.98)	1,873	8.38 (±0.86)	3,621	8.48 (±0.87)
ln Triglycerides (mg/dL)	2,400	4.51 (±0.32)	1,215	4.53 (±0.33)	4,386	4.87 (±0.59)	1,050	4.52 (±0.31)	538	4.54 (±0.33)	2,701	4.93 (±0.61)
White Blood Cell Count	2,409	7.38 (±2.46)	4,001	7.42 (±2.19)	N/A	N/A	1,754	7.73 (±2.08)	1,875	7.44 (±2.01)	N/A	N/A

Table 4.5 Novel significant associations between the fibrinogen cluster and hematological traits in NHANES III. Linear regressions adjusting for age, sex, BMI and smoking status were performed for 25 SNPs, assuming an additive genetic model for five hematological traits. Significant associations ($p < 0.05$) between the fibrinogen cluster variants and any one hematologic traits in any NHANES III population are shown.

SNP	Phenotype	Location/Gene	Race/Ethnicity	N	BETA	SE	p-value
rs2227395	Plasma fibrinogen	Intron/ *FGB*	Non-Hispanic White	1,391	-0.08	0.02	7.00E-04
rs2070022	Plasma fibrinogen	3′ UTR/ *FGA*	Mexican American	659	-0.07	0.03	1.06E-02
rs2066860	LN Triglycerides	Intron/ *FGG*	Mexican American	1,005	0.13	0.06	2.27E-02
rs1800792	LN Triglycerides	5′ flanking/ *FGG*	Mexican American	1,005	-0.03	0.01	2.84E-02
rs6050	Platelet Distribution Width	Thr312Ala/*FGA*	Non-Hispanic White	2,270	-0.03	0.02	4.06E-02
rs1800792	Platelet Count	Promoter/ *FGG*	Mexican American	1,663	-5.32	2.30	2.07E-02
rs4220	White blood cell count	Arg478Lys/*FGB*	Mexican American	1,682	-0.28	0.10	6.20E-03
rs2227395	White blood cell count	Intron/ *FBG*	Mexican American	1,682	-0.23	0.10	2.00E-02
rs2227425	Platelet Distribution Width	3'UTR/ *FGB*	Mexican Americans	1,661	1.14	0.54	3.00E-02
rs2066879	Platelet Distribution Width	3' near *FGG*	Non-Hispanic White	2,268	0.77	0.32	2.00E-02

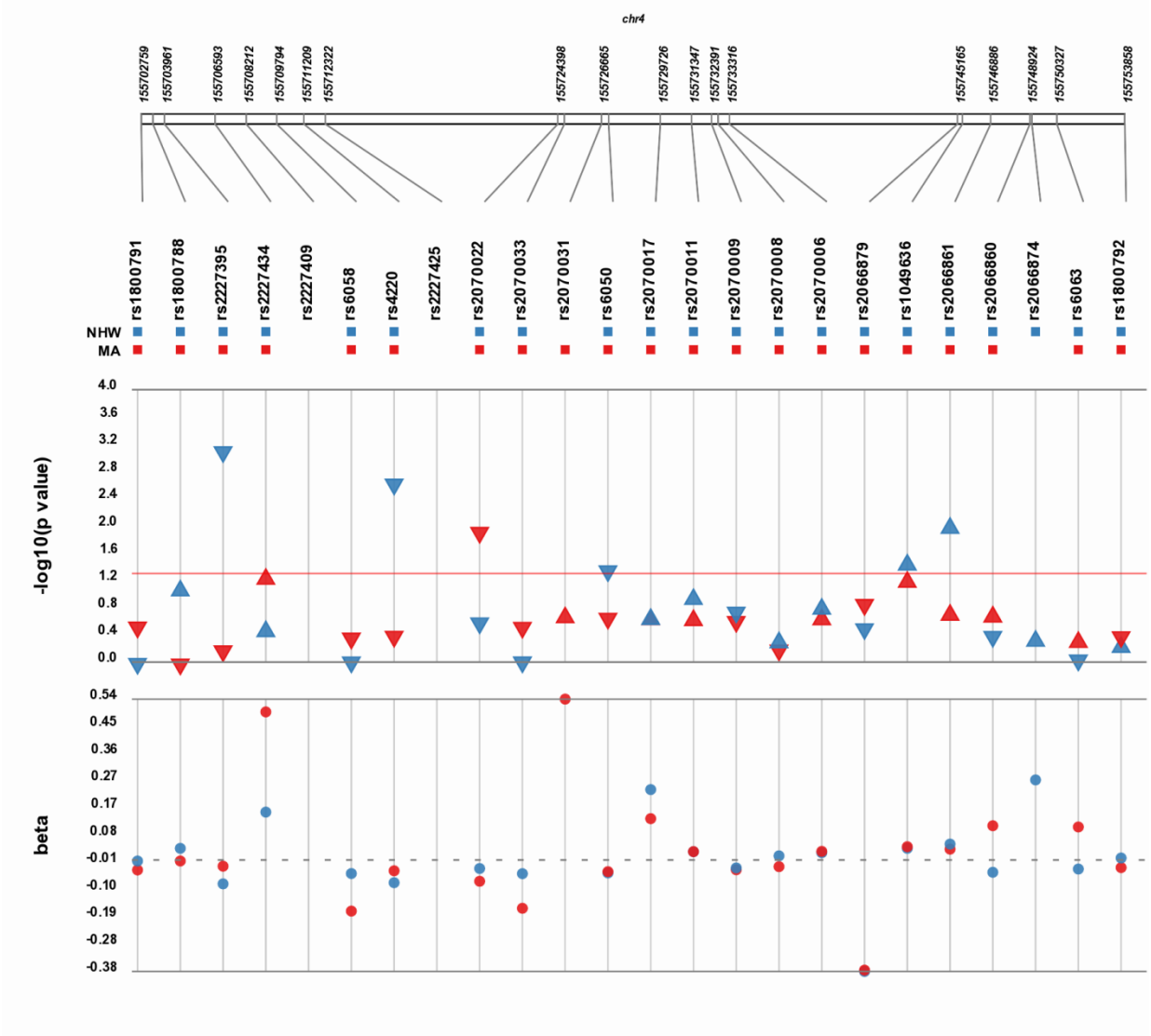


Figure 4.1 Synthesis view plot comparing results from tests of association across NHANES subpopulations for plasma fibrinogen. The –log10 of the p-value and the direction of the effect (indicated by arrow direction) are on the y-axis and SNPs are located on the x-axis. Each color represents a different NHANES III population: blue (non-Hispanic white; NHW), red (Mexican American; MA). Arrows above the red line represent a significant association ($p<0.05$).

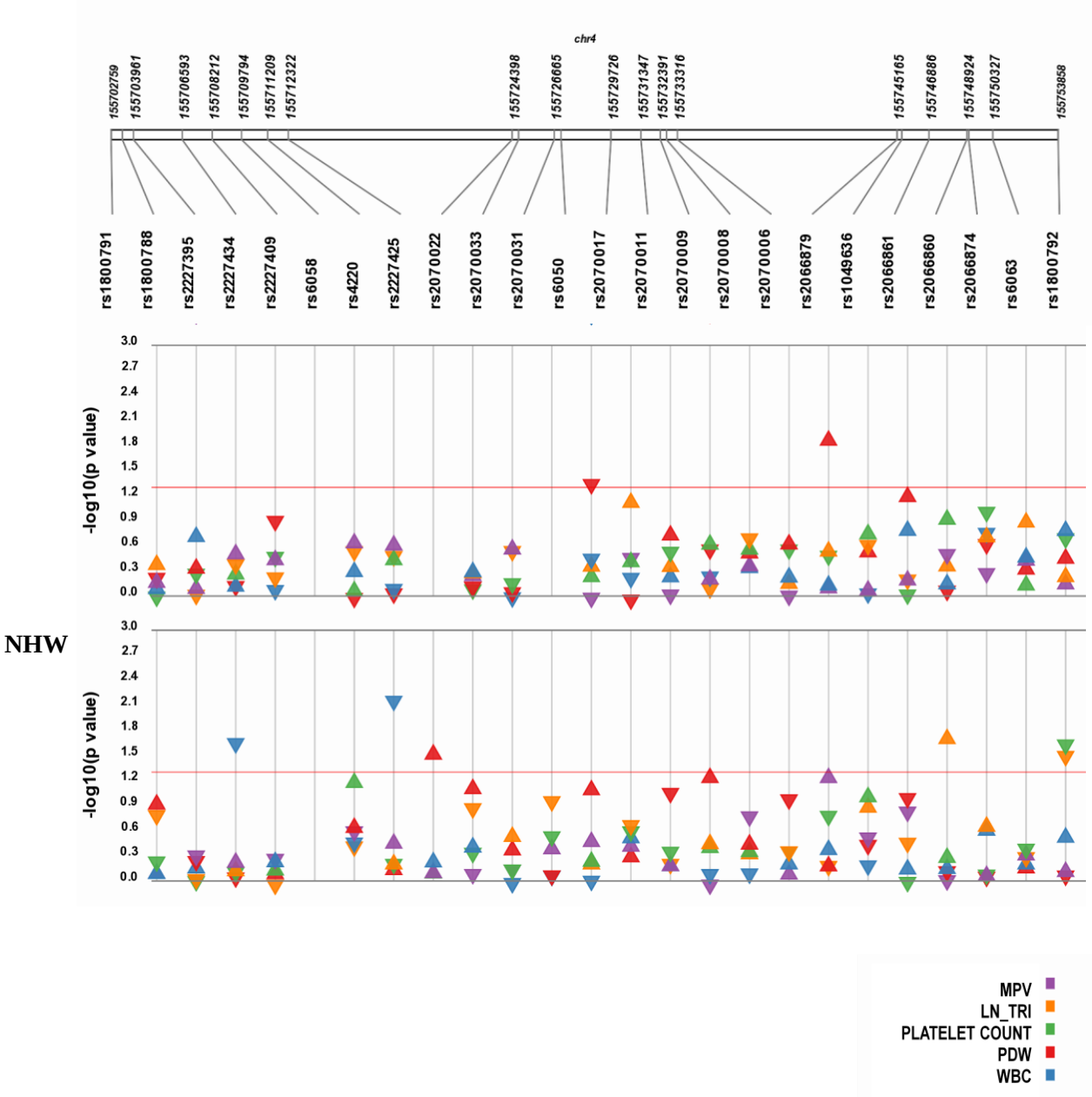


Figure 4.2 Synthesis view plot comparing significant results from tests of association for five hematological traits across NHANES subpopulations. The –log10 of the p-value and the direction of the effect (indicated by arrow direction) are on the y-axis and SNPs are located on the x-axis. Each panel represents a different population from the NHANES III dataset and each color represents the hematological traits tested for each SNP. Points above the red line represent a significant association ($p<0.05$). Abbreviations: non-Hispanic white (NHW), Mexican American (MA), mean platelet volume (MPV), transformed triglycerides (ln_TRI), white blood cell count (WBC), and platelet distribution width (PDW).

Replication

We replicated several single SNP associations from previous studies in non-Hispanic whites from NHANES III (Table 4.6). Replication was defined by statistical significance ($p < 0.05$) and consistent direction of effect with the same SNP in the same ancestral population across studies. Four of the six tests of association replicated at $p<0.05$ and, after accounting for the coded allele, all tests trended in the same direction as previous reports regardless of significance (Table 4.6) . In non-Hispanic whites, we replicated two associations previously reported in a candidate gene study (Reiner et al., 2006a): rs6050 (Thr312Ala; $p = 0.03$) and intronic rs1049636 ($p =0.04$) were associated with decreased and increased plasma fibrinogen, respectively. We also replicated two associations reported from recent GWA studies (Wassel et al., 2010k; Dehghan et al., 2009): *FGG* intronic rs2066861 ($\beta = 0.05$; $p=0.01$) and *FGB* nonsynonymous rs4220 ($\beta = -0.08$; $p=0.002$; Table 4.6).

There were several SNPs in the fibrinogen gene cluster we did not genotype in our original analysis in NHANES III. Our initial analysis was limited to tagSNPs chosen in either CEU or YRI HapMap reference populations (2003). Consequently, there were several SNPs that have been previously described by other candidate gene studies we did not analyze in NHANES III. We sought to replicate and/or generalize these variants in NHANES 99-02. To do so we genotyped two additional fibrinogen cluster variants that were originally identified by candidate gene studies in NHANES 99-02 participants. We did not replicate rs1800790, the only SNP previously identified in Europeans in our non-Hispanic whites from NHANES 99-02 (Table 4.7). This was an unexpected finding since

allele frequencies for rs1800790 are similar between the original population and our study population (MAF= 0.20, original study and MAF=0.16 in non-Hispanic whites). Additionally, we were adequately powered >80% to detect this effect given our sample sizes which is also comparable to the original study.

To replicate variants identified by GWAS, we genotyped three SNPs that were originally associated with fibrinogen and/or other hematological traits in our combined NHANES dataset. Only one of the three SNPs (rs210135) genotyped was previously described in Europeans (the others were originally identified in African Americans). *BAK1* variant, rs210135 was originally associated with decreased platelet count in Europeans (Table 4.8). We replicate this association in non-Hispanic whites from the combined NHANES dataset (Table 4.8).

Generalization

As previously mentioned, we genotyped two additional fibrinogen cluster variants in NHANES 99-02 for generalization studies. We considered a SNP generalized if it had consistent direction of effect and is statistically significant at $p<0.05$ in a different population that the original analysis. One SNP rs2070017 was previously identified by Reiner et al with decreased fibrinogen levels in African Americans. We failed to generalize this association in non-Hispanic whites (Table 4.7); however this association generalized in Mexican Americans from NHANES 99-02 (Table 4.9). The other SNP genotyped in NHANES 99-02, rs1800790, was previously associated with increased fibrinogen levels in Europeans (Reiner et al., 2006a). This SNP did not replicate in non-Hispanic whites (see previous section *"Replication"*; Table 4.7) nor did it generalize in

Mexican Americans (Table 4.9). While rs1800790 was statistically significant in Mexican Americans, the direction of effect was not consistent between studies (β=-0.07, p-value = 0.01; Table 4.9).

We used the combined NHANES dataset to determine the extent to which GWAS-identified variants associated with fibrinogen and/or other hematological traits generalize/replicate in non-Hispanic whites and/or Mexican Americans. We tested total of three SNPs: rs2070017, rs206674, and rs210135 for an association with either fibrinogen or platelet count in non-Hispanic whites and Mexican Americans. Two SNPs rs2070017 and rs206674 were previously identified by GWAS in African Americans and were associated with decreased fibrinogen levels (Wassel et al., 2010d). These variants did not generalize to non-Hispanic whites and rs2070017 had a different direction of effect compared to African Americans (Table 4.8). One SNP, rs2070017, was associated with decreased fibrinogen levels (β = 0.18, p-value = 0.04) and had a consistent magnitude of effect compared to African Americans. As previously mentioned, we genotyped one SNP in the combined dataset that was originally identified in Europeans and is associated with deceased platelet count. We replicated this finding in non-Hispanic whites (previously mentioned) and we generalized this variant in Mexican Americans (β = -5.91, p- value 5.8E-4; Table 4.10).

Table 4.6 Replication of results from previous studies in non-Hispanic whites for plasma fibrinogen levels. We compared results from our tests of association with plasma fibrinogen and all SNPs tested with the results from previous studies. Note that the coded allele may be different across studies. Abbreviations: Candidate gene Association Resource (CARe), Coronary Artery Risk Development in Young Adults (CARDIA), Rotterdam Study (RS), Framingham Heart Study (FHS), Cardiovascular Health Study (CHS), Atherosclerosis Risk in Communities Study (ARIC), Monitoring of Trends and Determinants of Cardiovascular Disease/Cooperative Health Research in the Region Augsburg (MONICA/KORA). Abbreviations: CA= coded allele, NHW = non-Hispanic white

		NHANES III non-Hispanic white			**CARe Wassel et al**			**CARDIA Reiner et al**			**RS/FHS/CHS/ARIC/ MONICA Wassel et al**		
SNP	**Alleles**	**CA**	**β**	**P-value**	**CA**	**β**	**P-value**	**CA**	**β**	**P-value**	**CA**	**β**	**P-value**
rs4220	A/G	G	0.08	2.10E-3	A	0.14	1.04E-34	--	--	--	G	-0.08	2.14E-27
rs2070011	A/G	G	0.02	0.14	--	--	--	A	-3.82	0.005	--	--	--
rs6050	A/G	G	0.04	0.03	--	--	--	G	-7.04	0.001	--	--	--
rs1800788	C/T	C	0.04	0.10	--	--	--	T	-6.41	0.008	--	--	--
rs1049636	C/T	C	0.04	0.04	--	--	--	C	6.16	0.003	--	--	--
rs2066861	A/G	G	0.05	0.01	T	-0.06	4.18E-10	--	--	--	--	--	--

Table 4.7 Generalization/Replication of SNPs identified by a candidate gene study of the fibrinogen gene cluster for plasma fibrinogen in non-Hispanic whites from NHANES 99-02. We performed standard linear regression with previously identified variants within the fibrinogen cluster in non-Hispanic whites from NHANES 99-02. All tests were adjusted for age, sex, smoking status, and BMI. The same coded allele was assigned across studies.

SNP	Gene	Original Study				NHANES 99-02 Non-Hispanic whites		
		Population Studied	n	β	P-value	n	β (SE)	P-value
rs1800790	*FGB*	European	1,765	10.47	<0.001	1,733	-0.01 (0.02)	0.37
rs2070017	*FGA*	African American	1,461	-14.01	3.0E-3	1,647	0.14 (0.11)	0.18

Table 4.8 Replication of GWAS- identified variant rs210135 and generalization of African American GWAS-identified variants in non-Hispanic whites from the combined NHANES. We performed a standard linear regression adjusted for age, sex, BMI, and smoking status. The same coded allele was assigned across studies.

SNP	Gene	Phenotype	Original Study				Combined NHANES Non-Hispanic whites		
			Original Population	n	β	P-value	n	β (SE)	P-value
rs2070017	*FGA*	Fibrinogen	African American	6,657	-0.16	4.82E-9	3,044	0.14 (0.11)	0.20
rs2066874	*FGG*	Fibrinogen	African American	6,657	-0.33	2.86E-11	3,102	-0.12 (0.20)	0.53
rs210135	*BAK1*	Platelet Count	Europeans	13,943	5.43	3.7E-10	6,094	3.10 (1.28)	0.02

Table 4.9 Generalization of SNPs identified by a candidate gene study of the fibrinogen gene cluster for plasma fibrinogen in Mexican Americans from NHANES 99-02. We performed standard linear regression with previously identified variants within the fibrinogen cluster in non-Hispanic whites from NHANES 99-02. All tests were adjusted for age, sex, smoking status, and BMI. The same coded allele was assigned across studies.

SNP	Gene	Original Study				NHANES 99-02 Mexican Americans		
		Population Studied	n	β	P-value	n	β (SE)	P-value
rs1800790	*FGB*	European	1,765	10.47	<0.001	749	-0.07 (0.03)	0.01
rs2070017	*FGA*	African American	1,461	-14.01	3.00E-3	720	-0.24 (0.11)	0.02

Table 4.10 Generalization of GWAS identified variants in Mexican Americans from the combined NHANES dataset. We performed a standard linear regression adjusted for age, sex, BMI, and smoking status. The same coded allele was assigned across studies.

SNP	Gene	Phenotype	Original Study				Combined NHANES Mexican Americans		
			Population Studied	n	β	P-value	n	β (SE)	P-value
rs2070017	*FGA*	Fibrinogen	African American	6,657	-0.16	4.82E-9	1,386	-0.18 (0.09)	0.04
rs2066874	*FGG*	Fibrinogen	African American	6,657	-0.33	2.86E-11	1,376	0.33 (0.22)	0.13
rs210135	*BAK1*	Platelet Count	Europeans	13,943	5.43	3.7E-10	3,545	-5.91 (1.71)	5.8E-4

Pleiotropy

Epidemiologic studies such as NHANES make statistical tests for pleiotropy possible, which may ultimately identify novel genotype-phenotype relationships and pathways (Pendergrass et al., 2011). We tested for pleiotropic effects with GWAS-identified variants and other related traits available in non-Hispanic whites and Mexican Americans from the combined NHANES dataset. For example rs1800562, identified by a GWAS in Europeans and was originally associated with mean cell volume (Soranzo et al., 2009), was tested here for an association with mean platelet volume, fibrinogen, triglycerides, and platelet count. Collectively, a total of 26 tests of pleiotropy were performed in non-Hispanic whites and Mexican Americans and two were significant at a liberal threshold of $p<0.05$ (Table 4.11). In Chapter III section B, we identified an association with intronic variant rs1800562, located in the protein coding hemochromatosis gene *HFE*, with plasma fibrinogen in combined NHANES non-Hispanic blacks ($\beta = 0.25$, p-value = 7.50 E-5; Table 3.5). Likewise, this SNP is also associated with fibrinogen in both non-Hispanic whites and Mexican Americans (table 4.11).

Genetic Modifiers

In Chapter III we identified several gene-gene (GxG) and gene-environment (GxE) interactions with fibrinogen cluster variants in non-Hispanic blacks from NHANES III (Table 3.6 and 3.7). Using the same approach we tested for gene-gene and gene-environment interactions for plasma fibrinogen in non-Hispanic white and Mexican

Americans from NHANES III. Notably, we identified a total of 19 gene-gene interactions and five gene-environment interactions across populations.

We identified a total of 19 GxG interactions that were significant at the liberal significance threshold of $p< 0.05$ out of 552 total tests performed (Tables 4.12 and 4.13). We define a significant GxG interaction here as having significant interaction term ($p<0.05$) from two different genes in the fibrinogen gene cluster.

Greater than 60% (12/19) of the significant GxG interactions were identified in non-Hispanic whites, all of which include a *FGB* SNP in the interaction term (Table 4.12). In fact, all of the significant interaction terms included either *FGB* rs2227395 or *FGB* rs4220, and the resulting associations are nearly identical in non-Hispanic whites (Table 4.12). As previously mentioned in the single-SNP test of association, these SNPs are highly correlated ($r^2 = 0.86$) and likely represent the same genetic effect. After accounting for this correlation, there are only six unique gene-gene interactions that reach statistical significance at $p<0.05$ in non-Hispanic whites. Both rs2227395 and rs4220 had significant main effects in the single-SNP test of association (Figure 4.1).

In Mexican Americans we identified seven significant interaction terms at the $p<0.05$ level. Four of these associations include rs2227395 and rs4220, and may be correlated in our Mexican American population; however there was no evidence proving this. After accounting for the correlation between SNPs, we identified five independent associations at the $p<0.05$ level (Table 4.13).

In addition to gene-gene interactions, we tested for gene-environment (GxE) interactions between fibrinogen cluster genetic variants and sex, age, smoking status, and

BMI separately for an association with plasma fibrinogen in non-Hispanic whites and Mexican Americans (from NHANES III). All three environmental factors we tested have been previously described as significant risk factors for elevated fibrinogen (de Maat, 2001). We only identified five SNP x environment interactions at $p < 0.05$ level out of 200 total tests performed (Table 4.14). Four of the five associations we identified were SNP x sex interactions with nearly similar association results for the interaction term in non-Hispanic whites. One interaction term was significant in Mexican Americans, rs2006879 x BMI, which was associated with fibrinogen levels (Table 4.14).

Table 4.11 Significant pleiotropic effects detected with GWAS-identified variants in non-Hispanic whites and Mexican Americans in the combined NHANES dataset. We performed linear regression with GWAS-identified variants and hematological traits: platelet count, mean platelet volume, and log- transformed triglycerides. Significance was declared at $p<0.05$.

Abbreviations: MPV = mean platelet volume

SNP	Gene	Original Study					Combined NHANES				
		Original Phenotype	**Population**	**n**	**β**	**P-value**	**Combined NHANES population**	**Associated Phenotype**	**n**	**β (SE)**	**P-value**
rs1800562	*HFE*	Mean cell volume	Europeans	13,943	0.37	1.4E-23	non-Hispanic whites	Fibrinogen	3,165	0.06 (0.03)	0.03
rs1800562	*HFE*	Mean cell volume	Europeans	13,943	0.37	1.4E-23	Mexican Americans	Fibrinogen	1,434	-0.12 (0.06)	0.03

Table 4.12 Significant fibrinogen cluster gene-gene interactions in NHANES III non-Hispanic whites (n = 1,385). Using linear regression, we added cross product terms (all pair-wise SNP-SNP) to the regression models for serum fibrinogen and adjusted for age, sex, BMI, and smoking status. Abbreviations: SE = standard error.

Gene-Gene Interaction	SNP-SNP Interaction	SNP 1 Main Effect		SNP 2 Main Effect		Interaction Term	
		Beta (SE)	p-value	Beta (SE)	p-value	Beta (SE)	p-value
FGB x *FGB*	rs1800791 x rs2227395	<0.001 (0.03)	0.89	-0.08 (0.02)	6.7E-4	0.15 (0.06)	0.02
FGA x *FGB*	rs2070006 x rs2227395	0.02 (0.02)	0.20	-0.08 (0.02)	6.7E-4	-0.09 (0.03)	0.005
FGA x *FGB*	rs2070006 x rs4220	0.02 (0.02)	0.20	-0.09 (0.02)	2.1E-3	-0.09 (0.04)	0.01
FGA x *FGB*	rs2070009 x rs2227395	-0.03 (0.02)	0.16	-0.08 (0.02)	6.7E-4	0.09 (0.04)	0.005
FGA x *FGB*	rs2070009 x rs4220	-0.03 (0.02)	0.16	-0.09 (0.02)	2.1E-3	0.09 (0.04)	0.01
FGA x *FGB*	rs2070011 x rs2227395	0.03 (0.02)	0.14	-0.08 (0.02)	6.7E-4	-0.08 (0.04)	0.03
FGA x *FGB*	rs2070011 x rs4220	0.03 (0.02)	0.14	-0.09 (0.02)	2.1E-3	-0.09 (0.04)	0.01
FGA x *FGB*	rs2070022 x rs2227395	-0.03 (0.02)	0.23	-0.08 (0.02)	6.7E-4	0.18 (0.06)	0.002
FGA x *FGB*	rs2070022 x rs4220	-0.03 (0.02)	0.23	-0.09 (0.02)	2.1E-3	0.14 (0.06)	0.02
FGB x *FGG*	rs2227395 x rs2066861	-0.08 (0.02)	6.7E-4	0.05 (0.02)	0.01	-0.13 (0.04)	0.002
FGB x *FGG*	rs4220 x rs2066861	-0.09 (0.02)	2.1E-3	-0.03 (0.02)	0.23	-0.08 (0.04)	0.05
FGA x *FGB*	rs6050 x rs2227395	-0.04 (0.02)	0.04	-0.08 (0.02)	6.7E-4	0.13 (0.04)	0.001

FGA x *FGB*	rs6050 x rs4220	-0.04 (0.02)	0.04	-0.09 (0.02)	2.1E-3	0.09 (0.04)	0.02

Table 4.13 Significant fibrinogen cluster gene-gene interactions in NHANES III Mexican Americans (n = 664). Using linear regression, we added cross product terms (all pair-wise SNP-SNP) to the regression models for serum fibrinogen. Abbreviations: SE = standard error.

Gene-Gene Interaction	SNP-SNP Interaction	SNP 1 Main Effect		SNP 2 Main Effect		Interaction Term	
		Beta (SE)	p-value	Beta (SE)	p-value	Beta (SE)	p-value
FGA x *FGB*	rs2070008 x rs1800791	-0.02 (0.04)	0.57	-0.03 (0.03)	0.26	0.18 (0.08)	0.022
FGA x *FGG*	rs2070033 x rs1049636	-0.16 (0.15)	0.26	0.04 (0.03)	0.08	0.59 (0.30)	0.051
FGA x *FGB*	rs2070033 x rs1800791	-0.16 (0.15)	0.26	-0.03 (0.03)	0.26	-0.66 (0.33)	0.048
FGB x *FGG*	rs2227395 x rs1049636	-0.02 (0.04)	0.58	0.04 (0.03)	0.08	0.17 (0.06)	0.006
FGB x *FGG*	rs2227395 x rs1800792	-0.02 (0.04)	0.58	-0.03 (0.03)	0.36	-0.13 (0.06)	0.025
FGB x *FGG*	rs4220 x rs1049636	-0.04 (0.04)	0.36	0.04 (0.03)	0.08	0.14 (0.07)	0.035
FGB x *FGG*	rs4220 x rs1800792	-0.04 (0.04)	0.36	-0.03 (0.03)	0.36	-0.12 (0.06)	0.032

Table 4.14 Gene-environment interactions with fibrinogen variants and age, sex, BMI, and smoking status in non-Hispanic whites (n = 1,385) and Mexican Americans (n = 664) from NHANES III. Using linear regression we added the multiplicative terms (SNP x environmental factor) to the regression model while simultaneously adjusting for covariates: age, sex, BMI, and smoking status when necessary. Below are the association results for the single SNP or environmental factor tests of association as well as the association results for the interaction term.

NHANES III Population	Interaction	Gene	SNP Main Effect		Environment Main Effect		SNP-Environment Interaction Effect	
			Beta (SE)	p-value	Beta (SE)	p-value	Beta (SE)	p-value
non-Hispanic whites	rs2070017 x sex	*FGA*	0.24 (0.22)	0.28	0.07 (0.03)	0.01	0.03	0.01
Mexican Americans	rs2066879 x BMI	*FGG*	-0.37 (0.24)	0.12	0.02 (0.004)	3.1E-5	-0.28	0.03
non-Hispanic whites	rs2066861 x sex	*FGG*	0.05 (0.02)	0.01	0.06 (0.03)	0.02	-0.08	0.05
non-Hispanic whites	rs2070009 x sex	*FGA*	-0.03 (0.02)	0.16	0.06 (0.03)	0.02	0.07	0.05
non-Hispanic whites	rs6050 x sex	*FGA*	-0.04 (0.02)	0.04	0.06 (0.03)	0.04	0.08	0.05

Discussion

Using a large, diverse study population, we were able to replicate as well as identify two novel associations between SNPs located in the fibrinogen gene cluster and levels of plasma fibrinogen. We were also able to replicate and generalize GWAS-identified variants for fibrinogen and hematological traits in non-Hispanic whites and Mexican Americans from NHANES III. This is the first genetic association performed in Mexican Americans with variants within the fibrinogen gene cluster. As well as the first study to identify both gene-gene and gene-environment interactions with fibrinogen cluster variants in a population setting.

Several genetic variants are associated with increased plasma fibrinogen in all three genes in the fibrinogen cluster. Any one SNP (replicated and novel) significantly associated with plasma fibrinogen in non-Hispanic whites explained 0.3-0.7% of the trait variability. Collectively, the five associations we identified account for 1.6% of the trait variability in plasma fibrinogen in non-Hispanic whites. While we successfully replicated genetic associations reported for plasma fibrinogen in non-Hispanic whites, two variants did not replicate the findings of previous studies (Table 4.6) (Reiner et al., 2006a; Wassel et al., 2010a; Dehghan et al., 2009). We were unable to replicate the association with rs2070011 and rs1800788 in non-Hispanic whites. In comparison to Reiner et al, we were under-powered to detect the same effects with our sample size in non-Hispanic whites. However, we identified two novel associations with plasma fibrinogen in *FGB* and *FGA*. The novel association we identified between plasma fibrinogen and *FGA* is

significant in Mexican Americans and trended towards significant in non-Hispanic blacks (Table 4.5).

This is the first study that characterizes GWAS findings for fibrinogen and hematological traits in Mexican Americans. We were able to successfully generalize two associations originally identified in European or African descent populations to Mexican Americans. One association rs210135 generalized to Mexican Americans and replicated in non-Hispanic whites. In Chapter III this association also generalized in African American. Confirming associations such as rs210135 and other GWAS-identified variants in other populations provides the starting point for understanding the clinical impact these variants may have on disease.

We identified 19 gene-gene associations in non-Hispanic whites and Mexican Americans. One criticism of exhaustively testing for interactions is the biological interpretation of the results. The fibrinogen gene cluster is an ideal candidate for identifying gene-gene interactions since each gene is dependent on the other to produce active fibrin in the blood (Zhang and Redman, 1996; Zhang and Redman, 1994). Notably, all but one of the gene-gene associations we detected includes SNPs in the rate-limiting gene *FGB.* The *FGB* gene plays an important role in fibrinogen production. Little is known about function of genetic variation in the other genes, *FGA* and *FGG*, in the fibrinogen gene cluster. Our gene-gene interaction data imply that there are several interactions involving these genes with the rate limiting gene, *FGB.*

We also identified several SNP x sex interactions in non-Hispanic whites. There have been several reports of an association with sex or gender specific hormone levels and fibrinogen (Jeff et al., 2011a)**;** (Middeldorp, 2005; Voetsch and Loscalzo.J, 2003; Evans et al., 1999). We did not detect significant interactions with age, BMI, or smoking status. It is important to note that several genes we identified are in high linkage disequilibrium with each other resulting in redundant results for both populations.

This study, which accesses the diverse NHANES datasets, has many strengths. One strength is the breadth of laboratory measurements available for tests of association. Until recently, most genetic association studies have ignored pleiotropy and have limited tests of association to a single trait or pathway. Pleiotropy is defined here and elsewhere as a locus affecting multiple traits (Wagner and Zhang, 2011). In a statistical setting, pleiotropy is detected as a single genetic variant associated with multiple traits, but this setting does not establish causality without further experiments. Epidemiologic studies such as NHANES make statistical tests for pleiotropy possible, which may ultimately identify novel genotype-phenotype relationships and pathways (Pendergrass et al., 2011). Our study is unique in that we also investigated the pleiotropic effects of the fibrinogen cluster and GWAS-identified variants with fibrinogen and other hematological traits. This is important given the complex pathophysiology of cardiovascular disease and the role hematological traits play in the development of disease. Several studies have been published on genetic variation with levels of plasma fibrinogen but little is known about how these same variants impact hematological

traits. It is known that elevated levels of plasma fibrinogen are observed for several disease states such as inflammation, deep venous thrombosis, and arterial thrombosis (David Ginsburg, 2005). However, it is still unclear as to whether elevated levels of plasma fibrinogen are a result of fibrinogen itself or other intermediate phenotypes underlying the pathophysiology. For example, arterial thrombosis, a common characteristic and intermediate phenotype for cardiovascular disease, is very complex and is likely to involve the fibrinogen gene cluster. However, unlike venous thrombosis, which is mainly due to the accumulation fibrin clots, arterial thrombosis is primarily due to platelet adhesion and it is likely to involve hematological traits tested in this study (Feinbloom and Bauer, 2005; Lane and Grant, 2000; Lowe, 2008; Mackman, 2008; Lowe, 2008). We identified genetic associations with fibrinogen SNPs and white blood cell count, platelet distribution width, and platelet count suggesting possible pleiotropic effects. These findings support the idea that variants in the fibrinogen gene cluster may also affect the regulation of hematological traits.

Another major strength of our study is its diversity. Although there have been studies performed in European and African descent populations, none have included Mexican Americans. The characterization of variants in non-European descent populations has considerable scientific benefits. For example in Mexican Americans, we identified a novel association between plasma fibrinogen levels and rs2070022 that also trended towards significance in non-Hispanic blacks (Chapter III, section A). We also were able to report a consistent direction of effect for rs1049636 across all the three NHANES populations despite sample size. In

addition to plasma fibrinogen levels, we identified significant associations with serum triglycerides in Mexican Americans.

A weakness of the present study is power and sample size. While NHANES III contains over 33,000 study participants, our study population sample size was limited for several reasons. The Centers for Disease Control and Prevention (CDC) only collected DNA samples from a subset of the entire dataset. DNA was not collected from participants who reported having hemophilia or chemotherapy within four weeks of collection (Gunter E. et al., 1996). Of these samples, plasma fibrinogen levels in NHANES III was only measured for participants > 40 years of age (Gunter E. et al., 1996), which drastically reduces our sample size with plasma fibrinogen compared to the hematological traits. Despite the small sample sizes for plasma fibrinogen, we were able to successfully replicate several known (GWAS identified and fibrinogen cluster candidate gene variants) associations at a liberal significance threshold of $p<0.05$. We also identified potentially novel gene-gene and gene-environment interactions at the same significance threshold, but because of multiple testing and the increased risk of false positive findings, these novel associations will require replication in other studies. In addition to limited power due to sample size, we were limited to only successfully genotyped SNPs, whih represent a small portion of the SNPs identified by GWAS for replication and generalization studies.

In summary, we have identified both previous and novel associations within the fibrinogen gene cluster with plasma fibrinogen levels, as well as confirmed

previous GWAS associations for platelet count in NHANES populations. Notably, we generalized GWAS variants originally identified in African Americans in non-Hispanic whites and Mexican Americans. We also identified several genetic modifiers of plasma fibrinogen in NHANES III. Identifying these associations will provide important information about intermediate phenotypes with the ultimate goal of predicting cardiovascular disease outcome.

CHAPTER V

CONCLUSIONS AND FUTURE DIRECTIONS

Conclusions

The experiments described here were designed to identify common genetic variants that explain or contribute to the variability of common quantitative traits associated with cardiovascular disease (CVD). The genetics of cardiovascular disease has been well-studied; however a large portion of the genetic variance is still unexplained. The use of quantitative traits in genotype-phenotype studies can be more powerful compared with disease outcomes given these traits are more robust to phenotypic heterogeneity or measurement error. Furthermore, the use of these easily and uniformly measured quantitative traits in genetic studies improves the reproducibility of results across studies, which is important for clinical validation.

Risk of CVD is not uniformly distributed across populations, as African Americans have more risk factors for CVD compared with European Americans (Lloyd-Jones et al., 2009). Unfortunately the genetics of quantitative traits associated with CVD, such as electrocardiographic and fibrinogen/hematological, traits have been mostly limited to populations of European descent. Using several approaches and diverse study populations such as African Americans; we identified and generalized genetic associations in candidate genes as well as other loci in the genome that are associated with ECG and/or fibrinogen and hematological traits.

Candidate gene association studies are based on previous biological knowledge of a gene and have been useful in identifying genetic risk factors for several common diseases and/or traits. In Chapter II section A, we focused on one candidate gene, *SCN5A*, and its association with ECG traits in >3,000 African Americans from the Jackson Heart Study. We tested 65 variants in the predominant sodium channel gene expressed in heart, *SCN5A*, for an association with nine ECG traits. We were able to identify 13 independent associations in this gene with PR interval, QRS duration, and P duration at $p < 1.0 \times 10^{-4}$. These variants individually explain up to 2% of the variation in ECG traits in African Americans. To confirm these associations, we tested for replication for five of the significant *SCN5A* associations from the JHS in non-Hispanic blacks in NHANES III. We were not able to replicate any of these associations most likely due to the limited sample size of non-Hispanic blacks in NHANES III, which is only 16% of the sample size compared to JHS. Thus, on average we were only powered to detect 20% of the effect estimates identified in JHS in non-Hispanic blacks from NHANES III for any ECG trait. In section A of Chapter IV, we characterized *SCN5A* associations in non-Hispanic whites and Mexican Americans from NHANES III. Two of these associations successfully generalized in either population. Results from these studies suggest *SCN5A* associations may be population specific but future studies with large sample sizes need to be performed to confirm associations in African Americans and generalizability of these associations in other populations.

Genes that encode the fibrinogen polypeptide have been extensively studied in European descent populations and are associated with fibrinogen levels in the

blood. There are few studies that test for this association in African Americans. In section A of Chapter III, we tested 25 fibrinogen cluster variants for an association with fibrinogen levels with non-Hispanic blacks from NHANES III. We were able to replicate two associations in the fibrinogen cluster previously identified either by candidate gene or GWA studies in Europeans with fibrinogen levels at $p<0.05$. We also identified a novel association with rs2070022 and decreased fibrinogen levels ($p=0.007$). These SNPs (replicated and novel) explain 1.7% of the trait variability for fibrinogen levels. In section B of Chapter IV, we replicated four previous associations, as well as identified one novel association with serum fibrinogen in non-Hispanic whites. In Mexican Americans, rs2070022, which is also associated with non-Hispanic blacks, was associated with decreased fibrinogen levels. In summary, we have identified both previous and novel associations within the fibrinogen gene cluster with plasma fibrinogen levels across three NHANES III populations.

Genes that are associated with complex phenotypes such as CVD are often involved in multiple pathways and potentially have an effect on multiple phenotypes. In Chapters III and IV, we tested for pleiotropic effects with SNPs in the fibrinogen gene cluster and hematological traits. We also tested fibrinogen variants for an association with serum triglycerides, which has been consistently associated with CVD risk factors such as atherosclerosis. In section A of Chapter III, we identified three pleiotropicassociations between fibrinogen variants and platelet count, platelet distribution width, and mean platelet volume in non-Hispanic blacks. We performed the same tests in section B of Chapter IV focused

on non-Hispanic whites and Mexican Americans. We were able to identify one pleiotropic effect in non-Hispanic whites and three in Mexican Americans. None of these associations were observed across all three populations; nevertheless, these pleiotropic findings are novel and suggest that these genes may affect several phenotypes involved in common disease pathways. The biological explanations for these pleiotropic associations remain to be elucidated and warrant further study.

The need for genetic association studies in more informative populations such as African Americans is vital for the identification of casual variants for complex diseases. To address this concern, we generalized several GWAS associations, originally identified in European and/or Asian descent populations, in two independent African American populations for ECG and fibrinogen/hematological traits (section B in Chapters II and III). There were many associations that did not generalize most likely due low power resulting from the small sample sizes and differences in allele frequency in African Americans. Despite this limitation we were able to generalize variants that have been replicated in multiple studies in our African Americans. For example, SNPs located in the *NOS1AP* gene has been replicated several GWAS of QT interval in European and/or Asian populations. Associations in this gene are also associated with QT interval in African American populations. Likewise, these associations also generalized in Mexican Americans and replicated in non-Hispanic whites. In section B of Chapter III, using African Americans from NHANES III, we confirmed the performance of a custom-context chip, the IBC BeadChip, which is designed to capture SNPs with low allele frequencies (particularly those in African

descent populations). We were able to replicate two of these associations: rs2070017 (β = -0.08, p= 0.02) and rs2066874 (β = -0.28, p= 1.5 E-4) with decreased fibrinogen. In addition to tests of association, we extensively characterized GWAS-identified loci for these traits by describing allele frequency and linkage disequilibrium patterns in African Americans. Understanding the genetic architecture of GWAS-identified variants in African descent populations may eventually lead to the identification of casual variants.

In addition to the single SNP tests of association described above, another approach to account for the unexplained heritability is to test for genetic modifiers. In section C of Chapters II and III, we test for gene-gene and gene-environment interactions for ECG and fibrinogen/hematological traits. Using this approach we identified several gene-gene interactions with rate-limiting gene *FGB* with variants in *FGA* and *FGB* in non-Hispanic blacks. Many of these associations generalized in non-Hispanic whites (section B, Chapter IV). We also identified several gene-environment interactions with all NHANES sub-population with age, sex, body mass index, and smoking with at least one fibrinogen variant.

In conclusion, the data described from this work have identified several genetic risk factors in clinically relevant CVD phenotypes in African Americans. We identified several variants with *SCN5A* that are associated with ECG traits in African Americas, some of which generalize in other populations. Similarly, novel associations with the fibrinogen gene cluster were identified with fibrinogen levels, a number of which also have pleiotropic effects with hematological phenotypes. This work also confirms several GWAS variants originally identified in European

and/or Asian descent populations in African Americans. Finally, this work identifies associations that do generalize in African Americans, suggesting the need for future analysis such as molecular studies, more genetic association studies, or clinical studies to determine the effect of these putative disease loci. Collectively, the data presented in these chapters provide the framework for genetic association studies of CVD in African Americans and add to the limited genetics research in diverse populations.

Future Directions

The future goals expanding upon the work described here could include the exploration of different approaches to identify novel genotype-phenotype associations in African Americans for ECG and fibrinogen/hematological traits. Our analysis was limited to three approaches: candidate gene study, genome-wide association study, and testing for genetic modifiers. While we were successful in identifying novel and previously-reported associations, there are several approaches we did not consider. Additionally, this work can be continued by using more study diverse populations with large sample sizes, which would increase our power to detect associations that we were underpowered to detect.

One such approach that we did not consider in this present body of work is admixture mapping. Admixed populations are derived from a combination of one or multiple ancestral populations (Patterson et al., 2004b; Smith et al., 2004; Winkler et al., 2010b). Gene flow in admixed populations, such as African Americans, is fairly recent and chromosomes are mosaics of European and African descent (Winkler et al., 2010a). An admixture scan searches for regions in the genome that have a higher representation from one ancestral population compared with the other ancestral population. To identify putative disease causing loci, the overrepresentation of one ancestral population is usually the population with higher incidence for the disease. This approach can be used to identify disease loci in our African American study populations missed by traditional candidate gene and GWA study designs. Currently we are using this approach as part of the electronic

Medical Records and Genomics (eMERGE) network in African Americans from the Vanderbilt Genome-Electronic Records (VGER) Project and Northwestern University NUgene Project. We performed an admixture scan on >930 SNPs to identify disease loci for ECG traits. The admixture scan revealed a novel region on chromosome 5 *DPYSL3* that is possibly associated with PR interval. Follow-up studies of this region need to be performed to confirm the association with PR interval in African Americans.

Fine-mapping is also an alternative approach which is primarily used to identify true 'causal' variants. Unlike our candidate gene approach, fine mapping requires dense genotyping or sequencing of a region of interest. In this study design, SNP selection is not necessarily based on linkage disequilibrium alone. Fine-mapping studies are usually hypothesis-driven, and investigators select a gene within the associated region that is suspected to be biologically relevant to the phenotype being studied. For example, this approach would be useful to follow-up the *NOS1AP* region, where the SNP rs10494366 is consistently replicated across studies (Section B, Chapter II and section A, Chapter IV), but has no obvious biological relevance to QT interval. Performing fine-mapping studies in this gene in a large sample of African Americans can help identify the putative casual variant in *NOS1AP* and other genes identified by GWAS with unknown biological function. Additionally fine-mapping studies will provide a better understanding of the GWAS-identified variants (from section B, Chapter II) that fail to generalize in African Americans.

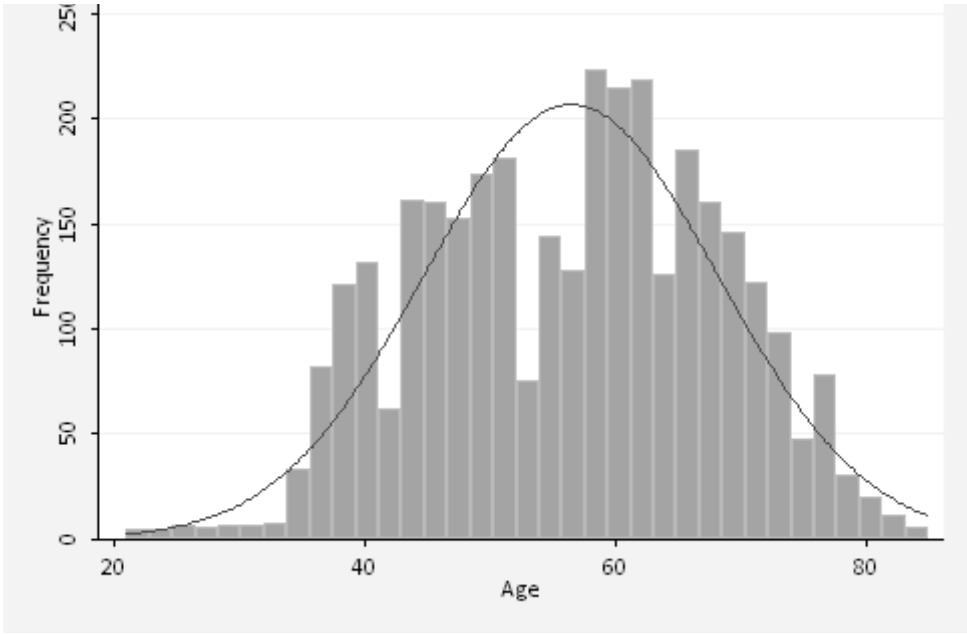
Perhaps one of the major limitations of this work was power. For several analyses, especially those for ECG traits, we were underpowered to generalize previously-identified association and to detect novel associations. We were underpowered particularly for gene-gene and gene-environment interactions, which typically have very small genetic effects and thus require larger sample sizes compared with studies designed to detect the marginal effects only. One remedy for this problem is to increase the sample size of the study population at the individual-level data, if possible. Another alternative would be to perform a meta-analysis, which combine association results across several study populations. Both solutions are possible extension of the work presented here.

The African American study population mentioned in section B of Chapter II was collected from two biobanks, VGER and NUgene. Samples from VGER are a part of the Vanderbilt biobanks BioVU, which accrues new samples every week. In fact there are an additional 600 African American samples in BioVU for PR interval since this present study was conducted. Several analyses include samples from NHANES participants and are a part of EAGLE, a study site within PAGE. Unfortunately, NHANES is a cross-sectional study and samples are collected at one time point for each survey, therefore additional samples cannot be added to a survey at one time point. But since NHANES is collected every year we can potentially increase our sample size for future analysis of traits such as fibrinogen which are measured each year. Another alternative to improve power in our current analysis would be to perform a meta-analysis with African Americans at other

study sites within PAGE that have data for ECG and fibrinogen/hematological traits.

Complex traits such as ECG traits and fibrinogen levels can be influenced by current medications or dietary intake. Environmental variables tested in the present study were limited to age and sex for gene-environment interactions for ECG traits and age, sex, body mass index, and smoking status for fibrinogen levels. NHANES is rich in nutritional data and several dietary measurements can be tested for gene-environment interactions that alter fibrinogen levels. Here we did not test for drug interactions with candidate genes that may alter ECG traits or fibrinogen levels. Information on current medications is available for individuals in BioVU and can be tested in future analysis for possible gene-drug interactions.

Collectively, the findings presented in this work identify several novel genetic associations for ECG and fibrinogen/hematological traits in African Americans. This work also shows the need for more studies in diverse populations for these traits. Lastly, this work illustrates the full spectrum of genotype-phenotype associations, as well as their modifiers, are needed to elucidate the biological pathways that may be relevant to the prevention, intervention, and/or treatment of cardiovascular disease in a population and/or personalized setting.



Appendix A. **Age distribution of JHS participants in the study population (n= 3,054).** Age in the study population follows a normal distribution. The mean age is 56 years and ranges from 21 to 85 with a standard deviation of 11.73 years.

Appendix B. Location, Hardy Weinberg Equilibrium, minor allele, minor allele frequency and marker genotyping efficiency for 72 SNPs in the *SCN5A* gene. Using thresholds of HWE p >0.0001, MAF >0.05, GE >95%, five SNPs (bold) did not meet these criteria were removed from further analyses.

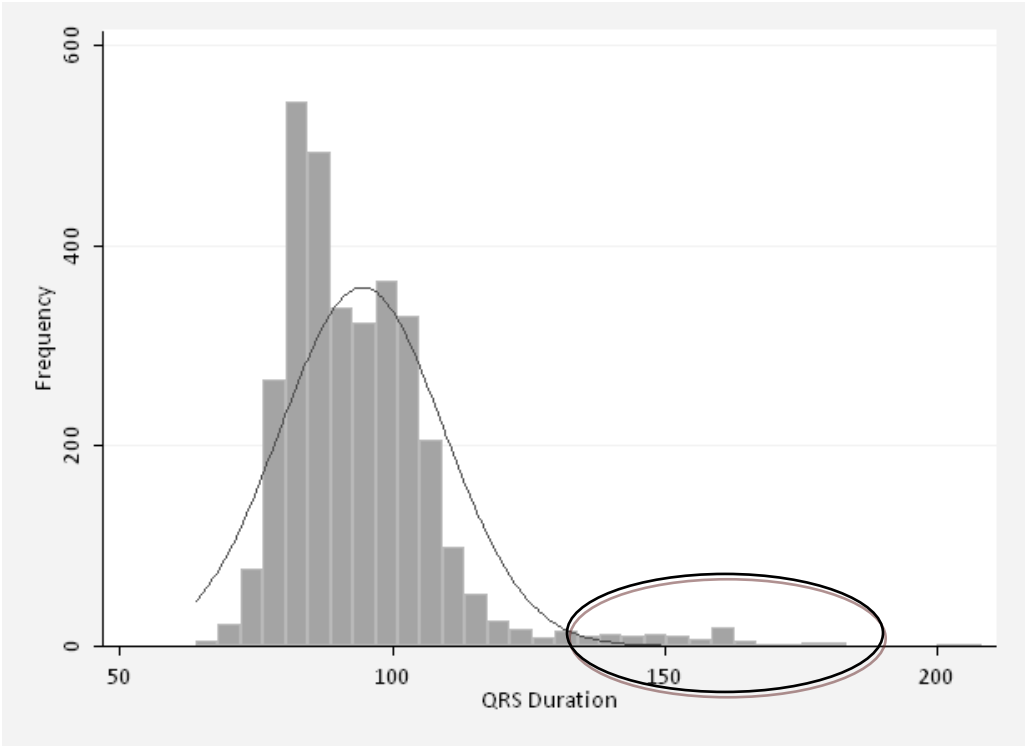
SNP	Location	HWE p-value	Minor Allele	MAF	Genotype Efficiency %
rs11129796	Intron	0.94	T	0.15	99.8
rs11710077	Intron	0.93	T	0.11	99.8
rs12053903	Intron	0.004	T	0.19	99.8
rs12491987	Intron	0.08	T	0.16	99.8
rs12498069	Intron	0.11	A	0.18	99.8
rs13073578	Intron	0.52	G	0.18	99.8
rs13084981	Intron	0.36	T	0.09	99.8
rs13314361	Intron	0.31	A	0.15	99.8
rs13315133	Intron	0.79	C	0.20	99.8
rs1805124	His->Arg	0.23	G	0.29	99.8
rs1805126	Asp->Asp	0.08	T	0.35	99.8
rs3918389	Gly-> Arg	1.00	G	**0.00**	99.8
rs3922843	Intron	0.50	T	0.24	99.8
rs3922844	Intron	0.01	G	0.41	99.8
rs3934936	Intron	0.85	T	0.35	99.8
rs3935472	Intron	0.75	A	0.33	99.8
rs4130467	Intron	0.31	G	0.22	99.8
rs41313691	Ser->Tyr	0.80	T	**0.04**	99.8
rs4131778	Intron	0.06	A	0.44	99.8
rs6599214	Intron	0.30	G	0.12	99.8
rs6599219	Intron	0.82	A	0.37	99.8
rs6599221	Intron	0.05	T	0.09	99.7
rs6599222	Intron	0.47	C	0.21	99.8
rs6599223	Intron	0.95	T	0.18	99.8
rs6599228	Intron	0.83	C	0.46	99.8
rs6599229	Intron	0.34	A	0.46	99.8
rs6599230	Ala->Ala	0.26	T	0.11	99.8
rs6763048	Intron	0.03	G	0.24	99.8
rs6768664	Intron	0.30	C	0.36	99.8
rs6770569	Intron	0.75	C	0.50	99.8
rs6786119	Intron	**2.83E-14**	C	0.33	99.5
rs6791924	Arg -> Ser	0.46	A	0.09	99.8
rs6793245	Intron	0.29	G	0.41	99.8

rs6795580	Intron	0.86	C	0.46	99.8
rs6797133	Intron	0.09	A	0.31	99.8
rs6799868	Intron	0.42	T	0.40	99.8
rs7372712	Intron	0.74	T	0.09	99.8
rs7372739	Intron	0.16	A	0.05	99.8
rs7373102	Intron	0.01	T	0.27	99.7
rs7373819	Intron	0.57	G	0.14	99.8
rs7374138	Intron	0.07	G	0.23	99.8
rs7374540	Intron	0.003	A	0.25	99.8
rs7374605	Intron	1.00	C	0.22	99.8
rs7427106	Intron	0.15	G	0.06	99.8
rs7427447	Intron	0.22	C	0.36	99.8
rs7428779	Intron	0.59	T	0.18	99.8
rs7428882	Intron	0.32	C	0.46	99.8
rs7429322	Intron	0.65	A	0.06	99.8
rs7433206	Intron	0.89	A	0.16	99.8
rs7620661	Intron	0.54	T	0.22	99.8
rs7624535	Intron	0.70	G	0.24	99.8
rs7626962	Ser->Tyr	0.73	T	0.08	99.8
rs7627552	Intron	0.91	A	0.19	99.8
rs7629265	Intron	1.00	T	0.08	99.8
rs7633974	Intron	0.53	C	0.34	99.8
rs7637849	Intron	0.67	A	0.21	99.8
rs7638909	Intron	**1.52E-15**	G	0.50	99.5
rs7645173	Intron	0.62	A	0.10	99.8
rs7645358	Intron	0.86	A	0.43	99.8
rs9311190	Intron	1.00	G	0.06	99.8
rs9311191	Intron	**2.39E-05**	C	0.15	99.8
rs9311195	Intron	0.18	G	0.16	99.8
rs9812912	Intron	0.71	T	0.14	99.8
rs9822819	Intron	0.48	A	0.31	99.8
rs9831389	Intron	0.28	C	0.36	99.8
rs9832586	Intron	0.31	T	0.08	99.8
rs9832895	Intron	0.97	C	0.32	99.8
rs9833086	Intron	0.69	G	0.45	99.8
rs9853984	Intron	0.58	A	0.20	99.8
rs9856587	Intron	0.11	T	0.41	99.8
rs9876660	Intron	0.69	C	0.28	99.8

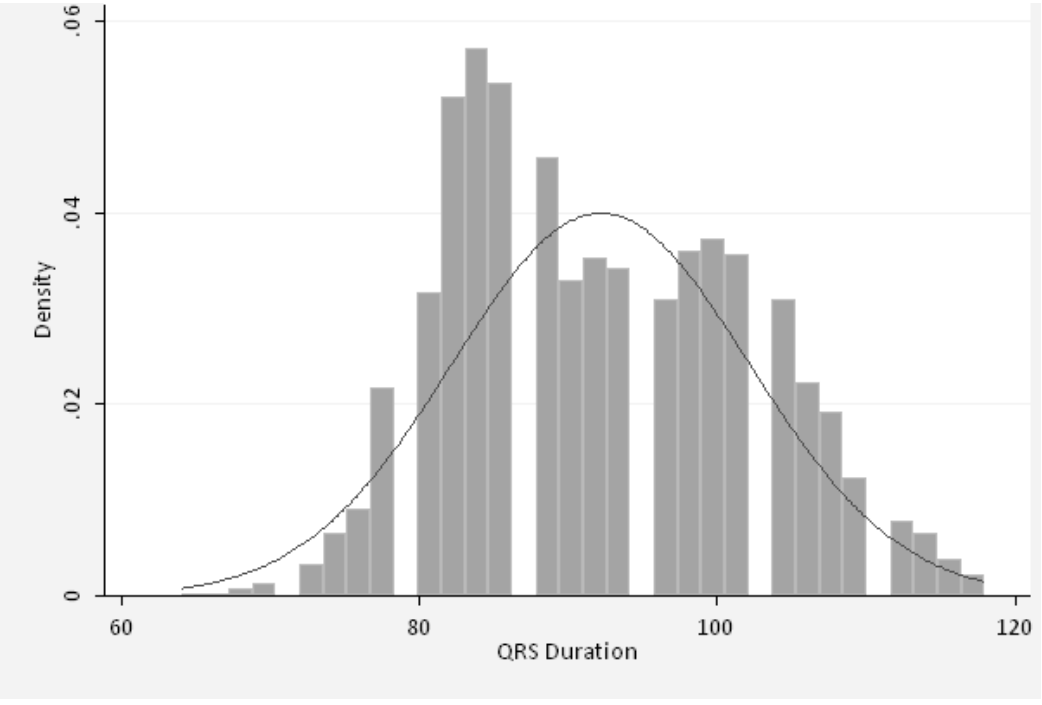
Appendix C. Correlation matrix for ECG traits. Pair wise correlations of 11 ECG traits were performed to determine if multiple traits are correlated. T & QT durations and PR duration II & VI are significantly correlated (r^2=0.93 and 0.87 respectively (bold).

	P Dur	**PR lead II**	**PR lead VI**	**QRS Dur**	**QT Dur**	**Qtc Dur**	**T Dur**	**P Axis**	**QRS Axis**	**T Axis**	**Heart Rate**
P Dur	1.00										
PR lead II	0.48	1.00									
PR lead VI	0.47	**0.93**	1.00								
QRS Dur	0.24	0.12	0.13	1.00							
QT Dur	0.16	0.12	0.12	0.29	1.00						
QTc Dur	0.02	-0.003	-0.003	0.27	0.37	1.00					
T Dur	0.05	0.06	0.06	-0.16	**0.87**	0.24	1.00				
P Axis	-0.007	0.03	0.01	0.007	-0.05	0.07	-0.06	1.00			
QRS Axis	-0.07	-0.05	-0.06	-0.16	-0.06	-0.09	0.01	0.17	1.00		
T Axis	0.008	0.02	0.01	-0.003	0.02	0.04	0.02	0.10	0.03	1.00	
Heart Rate	-0.14	-0.12	-0.13	-0.06	-0.65	0.44	-0.63	0.11	-0.02	0.01	1.00

A.



B.



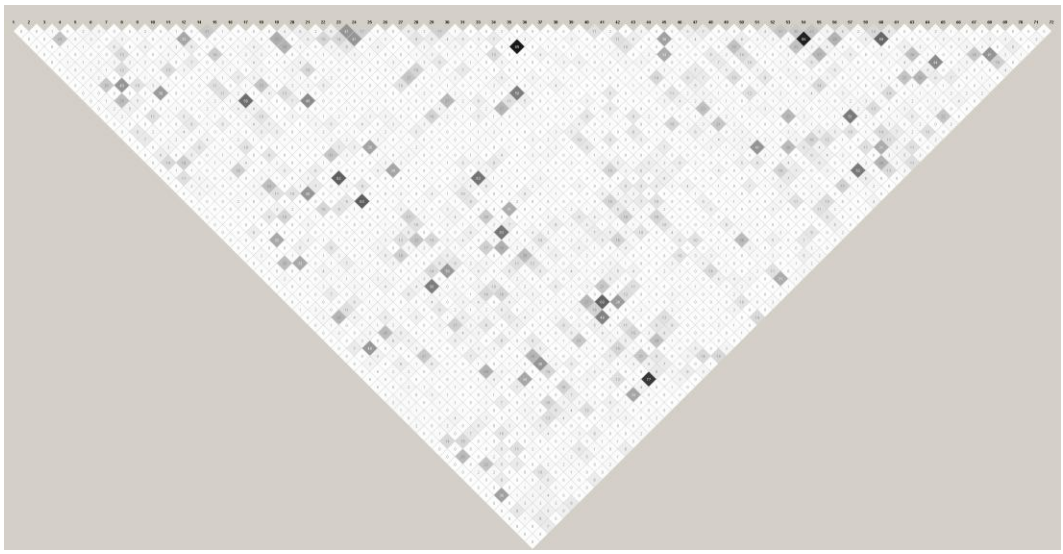
Appendix D. **Distribution of QRS duration values in unrelated JHS study participants.** A. QRS duration distribution including all JHS participants with a QRS

measurement. The histogram is skewed to the right of the normal distribution (skewness=2.19). There were a total of 169 outliers with a QRS duration >120m/sec that were excluded from the analysis with QRS duration. B. Gaussian distribution of QRS duration in unrelated JHS study participants after the removal of outliers.

Appendix E. Beta coefficients and p-values for single SNP tests of association with ECG axes and heart rate in unrelated JHS participants. Results from linear regressions adjusting for age, sex, and European ancestry was performed for 65 SNPs (n=3,054) assuming an additive genetic model are shown for five quantitative ECG traits. There were no significant associations at $p<1.0x10^{-4}$ between the genotyped *SCN5A* SNPs and ECG axes and heart rate measurements.

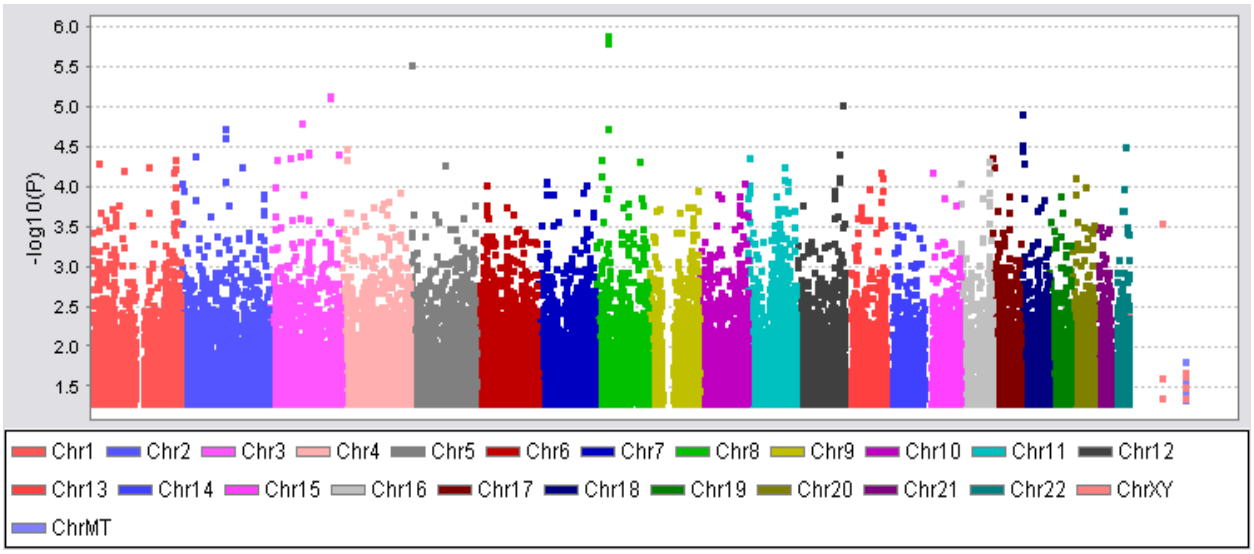
SNP	QTC DURATION		P AXIS		QRS AXIS		T AXIS		HEART RATE	
	β	p-value	β	p-value	β	p-value	β	p-value	β	p-value
rs11129796	0.07	0.94	1.04	0.17	1.32	0.22	0.88	0.54	0.32	0.40
rs11710077	0.99	0.34	-1.37	0.11	0.61	0.62	-2.43	0.14	-0.61	0.15
rs12053903	-0.86	0.28	-0.48	0.45	1.20	0.19	-0.03	0.98	-0.65	0.04
rs12491987	0.54	0.54	-0.90	0.21	0.24	0.81	-3.75	0.01	-0.53	0.14
rs12498069	0.46	0.58	-0.61	0.36	-2.19	0.02	-1.82	0.16	-0.08	0.82
rs13073578	-0.06	0.95	-0.40	0.55	0.61	0.53	-1.67	0.20	0.28	0.41
rs13084981	-0.15	0.90	-0.64	0.48	-1.88	0.14	-3.23	0.06	-0.04	0.94
rs13314361	-1.04	0.24	-0.63	0.38	-0.76	0.46	-1.16	0.40	-0.23	0.52
rs13315133	0.30	0.70	-0.06	0.93	-0.30	0.74	-3.29	0.01	-0.21	0.52
rs1805124	0.14	0.84	1.03	0.08	1.08	0.19	-1.72	0.13	-0.15	0.61
rs1805126	-0.22	0.75	-0.40	0.46	1.67	0.03	-0.68	0.52	-0.04	0.87
rs3922843	0.54	0.47	0.81	0.19	1.48	0.09	-2.58	0.03	-0.08	0.79
rs3922844	-1.34	0.04	0.09	0.87	-0.63	0.40	1.69	0.09	-0.21	0.42
rs3934936	0.77	0.26	-0.05	0.93	-1.82	0.02	-0.70	0.51	0.05	0.87
rs3935472	-0.22	0.75	0.90	0.11	-0.13	0.88	1.10	0.31	0.33	0.24
rs4130467	-1.36	0.08	0.95	0.13	-0.01	1.00	1.87	0.12	-0.30	0.34
rs4131778	-0.47	0.47	0.13	0.81	0.52	0.49	0.59	0.56	-0.37	0.16
rs6599214	-1.29	0.20	1.80	0.03	0.85	0.46	0.66	0.67	0.01	0.99
rs6599219	-0.32	0.63	0.28	0.61	1.02	0.19	-0.89	0.39	-0.01	0.98
rs6599221	-0.54	0.63	2.94	0.001	1.90	0.15	4.14	0.02	-0.39	0.40
rs6599222	-0.99	0.21	1.32	0.04	0.16	0.86	3.33	0.01	-0.26	0.41
rs6599223	-0.70	0.40	0.85	0.21	0.88	0.37	2.77	0.03	-0.16	0.64
rs6599228	-0.72	0.27	0.20	0.71	0.94	0.21	0.15	0.88	0.03	0.90
rs6599229	-0.93	0.15	-1.00	0.06	0.26	0.74	0.50	0.63	-0.36	0.18
rs6599230	1.86	0.08	-0.70	0.42	-1.34	0.28	2.21	0.19	0.14	0.75
rs6763048	0.48	0.52	-0.52	0.39	0.44	0.61	-1.63	0.16	0.37	0.22
rs6768664	0.13	0.84	0.26	0.64	0.14	0.86	0.84	0.42	-0.72	0.01

rs6770569	-0.003	0.99	0.67	0.21	0.40	0.60	-0.41	0.69	-0.26	0.33
rs6791924	-0.21	0.85	-1.82	0.04	-0.17	0.89	-1.39	0.42	0.45	0.32
rs6793245	-0.47	0.47	-0.35	0.51	1.59	0.04	-1.74	0.09	-0.44	0.10
rs6795580	0.98	0.13	-0.28	0.60	-0.65	0.39	-0.06	0.95	-0.10	0.70
rs6797133	-0.12	0.86	0.88	0.11	1.71	0.03	0.83	0.44	-0.17	0.55
rs6799868	-0.18	0.78	-0.31	0.56	1.22	0.11	-1.77	0.08	-0.48	0.07
rs7372712	-0.64	0.57	0.87	0.35	1.07	0.42	-0.21	0.91	0.94	0.05
rs7373102	0.58	0.42	-0.50	0.41	0.23	0.79	1.02	0.37	-0.50	0.09
rs7373819	-1.36	0.14	-0.92	0.22	-0.67	0.53	1.77	0.22	-0.13	0.72
rs7374138	0.31	0.69	-0.14	0.82	1.51	0.09	-0.22	0.85	0.40	0.20
rs7374540	-0.09	0.91	-1.05	0.08	0.95	0.27	-0.92	0.43	0.11	0.71
rs7374605	-0.38	0.62	-0.31	0.63	-1.29	0.15	0.40	0.74	0.37	0.24
rs7427106	-0.38	0.78	0.82	0.46	-0.52	0.75	0.96	0.65	-0.27	0.63
rs7427447	0.35	0.60	0.21	0.70	-0.69	0.39	0.10	0.93	0.59	0.04
rs7428779	0.59	0.48	-0.33	0.63	0.84	0.40	-2.70	0.04	-0.15	0.67
rs7428882	0.71	0.27	0.55	0.30	-0.11	0.89	-1.15	0.26	0.03	0.91
rs7429322	0.03	0.98	-0.16	0.88	-1.03	0.51	-4.11	0.05	-1.10	0.04
rs7433206	-0.05	0.96	-1.13	0.11	-1.02	0.32	1.63	0.24	0.71	0.05
rs7620661	1.51	0.05	-1.52	0.02	-0.21	0.81	-0.51	0.68	0.01	0.97
rs7624535	-0.30	0.69	-0.02	0.98	1.61	0.07	0.68	0.56	-0.29	0.35
rs7626962	1.53	0.20	-0.74	0.45	0.68	0.62	-0.75	0.69	0.91	0.06
rs7627552	0.93	0.26	-0.72	0.28	0.38	0.69	-0.35	0.79	1.01	0.003
rs7629265	1.19	0.32	-0.62	0.53	0.08	0.95	-1.53	0.42	0.98	0.05
rs7633974	-0.62	0.36	-0.57	0.31	0.24	0.76	0.99	0.36	-0.04	0.89
rs7637849	0.15	0.85	-1.06	0.10	-0.03	0.98	-0.08	0.95	0.27	0.40
rs7645173	-0.05	0.96	2.18	0.01	1.69	0.18	-0.94	0.58	-0.17	0.70
rs7645358	-0.54	0.41	0.16	0.76	0.78	0.31	0.01	0.99	0.14	0.61
rs9311190	-0.34	0.80	1.81	0.10	2.04	0.19	-0.77	0.71	-0.09	0.87
rs9311195	-1.74	0.05	0.49	0.49	0.24	0.82	0.82	0.56	0.46	0.20
rs9812912	1.23	0.19	0.25	0.74	0.22	0.84	-2.21	0.13	-0.64	0.10
rs9822819	0.02	0.97	-0.73	0.20	-0.89	0.27	0.34	0.76	-0.50	0.08
rs9831389	0.25	0.71	-0.40	0.47	-0.64	0.41	-0.22	0.84	0.61	0.03
rs9832586	0.05	0.97	1.95	0.06	1.67	0.24	-0.34	0.86	0.93	0.07
rs9832895	-0.63	0.36	0.73	0.19	0.09	0.91	0.62	0.57	0.16	0.58
rs9833086	-0.06	0.92	-0.34	0.53	-0.28	0.71	1.10	0.28	-0.30	0.27
rs9853984	-0.17	0.84	-1.05	0.11	-0.93	0.33	2.15	0.09	0.28	0.40
rs9856587	-0.28	0.67	0.12	0.82	-0.65	0.39	-0.56	0.58	-0.15	0.57
rs9876660	-0.15	0.83	-0.08	0.90	-0.38	0.65	0.12	0.92	-0.12	0.69

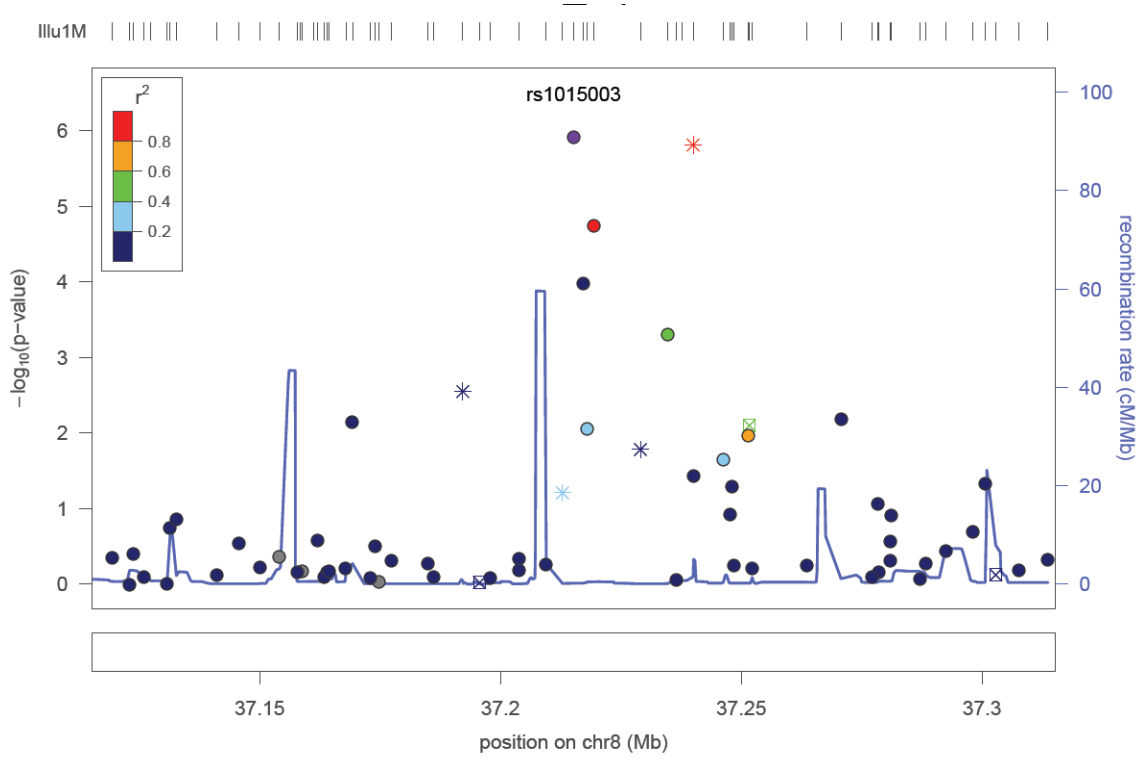


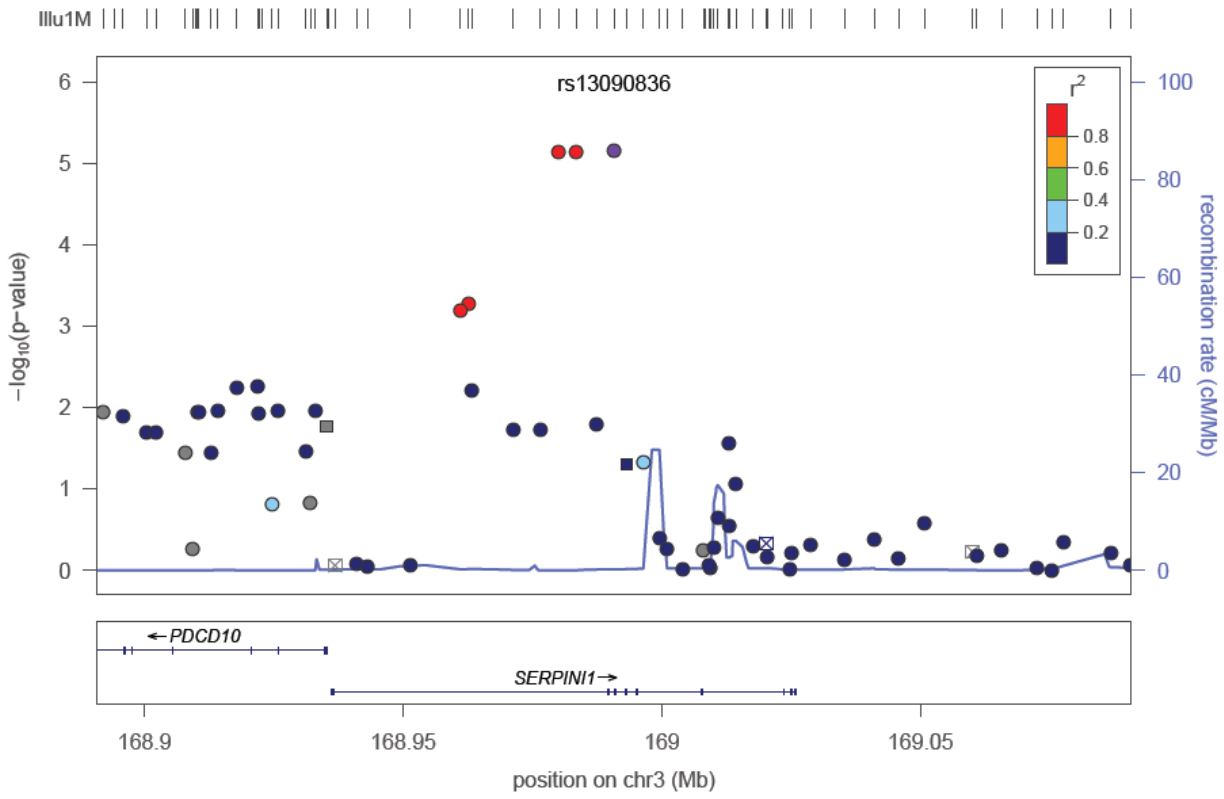
Appendix F. Linkage disequilibrium plots of SNPs in unrelated JHS participants. Linkage disequilibrium is presented as r^2 using Haploview (Barrett et al., 2005). SNPs rs6793245/rs3935472, rs7626962/rs7627552, rs9833086/rs4130467 are in LD (r^2>0.77).

1.


-log10(P)
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
Chr1 Chr2 Chr3 Chr4 Chr5 Chr6 Chr7 Chr8 Chr9 Chr10 Chr11 Chr12
Chr13 Chr14 Chr15 Chr16 Chr17 Chr18 Chr19 Chr20 Chr21 Chr22 ChrXY
ChrMT

2.


Illu1M
rs1015003
r²
0.8
0.6
0.4
0.2
−log₁₀(p−value)
recombination rate (cM/Mb)
37.15
37.2
37.25
37.3
position on chr8 (Mb)



3.

Appendix G. Genome-wide association results for heart rate in African Americans (n= 455). (1) Manhattan plot of GWAS results for heart rate. Association results are plotted for each maker on the x-axis across the genome, the –log of the p-value is plotted on the y axis. **(2)Locus Zoom Plot for the most significant association.** We plotted the local association results around 100kb of the most significant SNP rs1015003 from our GWAS of heart rate. The –log of the p-value is plotted on the y-axis and each SNP is plotted by location on the x-axis. Pair-wise linkage disequilibrium with the index SNP, measured by r2, is indicated by color. **(3) Locus Zoom Plot for the *SERPINI1* region.** We plotted the local association results around 100kb of the most significant SNP (rs13090836) region in the *SERPINI1* region.

Appendix H. SNP-trait associations significant in African Americans (p<0.05) but trend in the opposite direction compared with previously identified association. We compared association results in African Americans to the association results from the original study accounting for the coded allele. This table only represents associations that are statistically significant in African Americans (p<0.05) but have discordant effect sizes (expressed as a beta coefficient) compared to the original study. Abbreviations: CAF= coded allele frequency

Trait	SNP	Coded Allele	Location	Original Study					African Americans (n=455)		
				Population	Sample size	P-value	β	CAF	P-value	β	CAF
PR interval	rs1937332	G	*FAS,CH25H*	European	2,334	1.00E-10	2.6	0.47	0.007	-3.48	0.49
PR interval	rs7602460	A	Intergenic	European	2,334	6.00E-06	2.42	0.40	0.012	-3.38	0.43
PR interval	rs12053903	T	*SCN5A*	European	13,685	1.00E-07	-6.59	0.72	0.01	3.76	0.23
PR interval	rs4944092	G	*WNT11*	European	28,517	3.00E-08	-1.19	0.32	0.02	2.96	0.45
QT interval	rs12576239	T	*KCNQ1*	European	13,685	1.00E-15	1.75	0.13	4.4E-04	-8.53	0.15
Heart rate	rs1024020	A	Intergenic	European	2,325	7.00E-14	0.18	0.21	0.02	-1.63	0.38

Appendix I. Effect size and p-values for GWAS-identified associations with a consistent direction of effect in African Americans compared to European descent populations. We compared the association results from GWAS-identified variants to African American association results for SNPs that have a consistent direction of effect but did not meet our significance threshold ($p<0.05$) for all ECG traits.

Trait	SNP	Coded Allele	Gene/Location	Original Study					African Americans (n=455)		
				Population	Sample Size	P-value	Effect	CAF	P-value	Effect	CAF
QT Interval	rs10494366	G	*NOS1AP*	European	12,670	5.00E-22	12.2	0.39	0.88	-0.28	0.60
QRS duration	rs11129795	G	*SCN5A*	European	12,670	5.00E-10	-8.24	0.77	0.73	0.25	0.84
QT Interval	rs12053903	C	*SCN5A*	European	13,685	1.00E-06	-1.23	0.34	0.27	2.45	0.77
PR interval	rs3807989	A	*CAV1*	European	12,670	7.00E-13	6.40	0.4	0.46	-0.94	0.62
QT Interval	rs10919071	A	*ATP1B1*	European	15,842	1.00E-15	2.05	0.87	0.93	-0.46	0.97
PR interval	rs1371867	C	*RNF19A, ANKRD46*	European	2,334	3.00E-10	2.32	0.42	0.28	-1.64	0.74
PR interval	rs251253	C	*NKX2-5, C5orf41*	European	28,517	9.00E-13	-1.49	0.4	0.53	0.85	0.64
Heart rate	rs2670321	C	Intergenic	European	2,325	2.00E-06	-0.16	0.27	0.7	0.31	0.63
Heart rate	rs281868	G	*SLC35F1*	European	38,991	4.00E-10	6.30	0.5	0.41	-0.64	0.53
QT Interval	rs3807375	T	*KCNH2*	European	12,670	5.00E-11	11.95	0.35	0.32	-1.95	0.69
QT interval	rs652889	A	*PTPRG*	European	2,325	8.00E-07	-0.15	0.36	0.44	1.43	0.60
Ventricular conduction	rs7562790	G	*CRIM1*	European	40,407	8.00E-09	0.39	0.4	0.75	-0.18	0.59
QT	rs11129795	A	*SCN5A*	European	15,842	5.00E-14	-1.27	0.23	0.05	-4.77	0.16

Interval											
QT interval	rs12143842	T	*NOS1AP*	European	13,685	1.00E-10	3.15	0.26	0.59	1.50	0.12
QT Interval	rs3825214	G	*TBX5*	European	12,670	1.00E-07	5.88	0.22	0.57	1.20	0.23
QT Interval	rs4725982	T	*KCNH2*	European (and Indian Asian)	13,685	5.00E-16	1.58	0.22	0.03	4.79	0.23
QRS duration	rs6795970	A	*SCN10A*	European	12,670	4.00E-09	5.17	0.36	0.01	2.43	0.08
PR interval	rs6795970	A	*SCN10A*	European (and Indian Asian)	12,670	4.00E-09	5.17	0.36	0.13	3.67	0.08
PR interval	rs6800541	C	*SCN10A*	European	28,517	1.00E-13	3.77	0.4	0.2	3.07	0.09
QT Interval	rs8049607	T	*LITAF*	European	13,685	5.00E-15	1.23	0.49	0.26	2.00	0.45
QT interval	rs10488031	A	*ELMO1*	European	2,325	2.00E-06	-0.26	0.07	0.93	0.20	0.17
Heart rate	rs10496166	A	Intergenic	European	2,325	3.00E-14	0.18	0.13	0.43	0.60	0.43
Heart rate	rs10514995	G	Intergenic	European	2,325	5.00E-10	0.15	0.36	0.2	1.00	0.48
Heart rate	rs11154022	A	*GJA1*	European	2,325	7.00E-08	0.59	0.33	0.56	0.56	0.17
Ventricular conduction	rs11710077	T	*SCN5A*	European	40,407	1.00E-06	0.44	0.21	0.004	2.42	0.11
Ventricular conduction	rs11848785	G	*SIPA1L1*	European	40,407	1.00E-10	0.50	0.27	0.2	1.27	0.089
PR interval	rs11897119	C	*MEIS1*	European	28,517	5.00E-11	1.36	0.39	0.76	0.39	0.37
Heart rate	rs12554086	A	Intergenic	European	2,325	1.00E-10	-0.16	0.39	0.93	-0.09	0.15
QRS duration	rs1321311	T	*CDKN1A*	European	12,670	3.00E-10	6.52	0.21	0.17	0.77	0.39
QT interval	rs1348582	G	*PDE3A*	European	2,325	1.00E-06	0.20	0.18	0.59	1.00	0.34

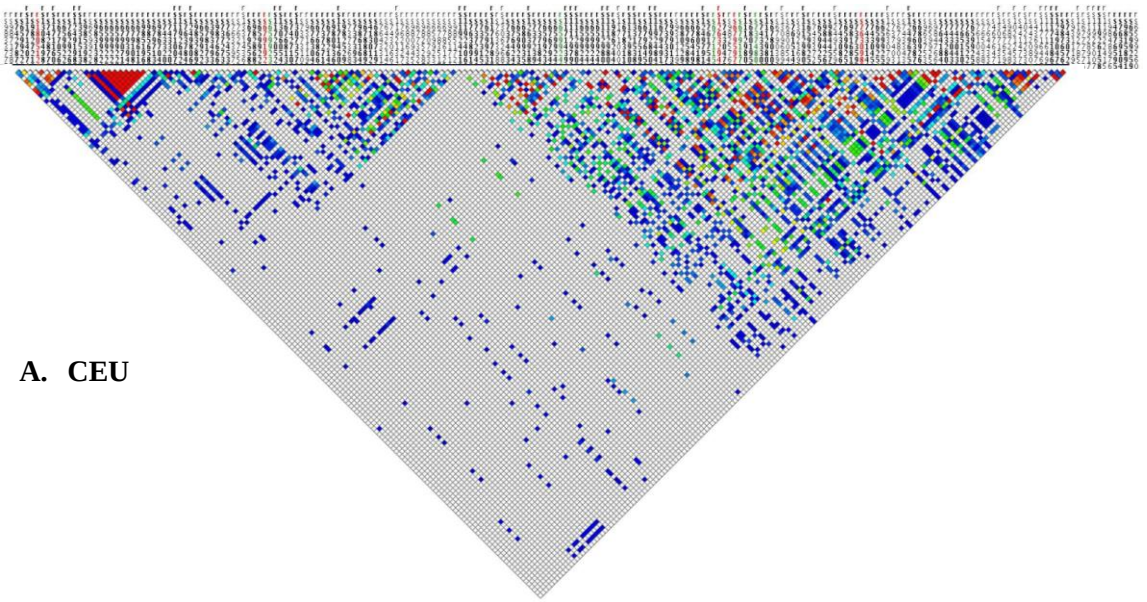
Heart rate	rs1541010	A	*FRMD4A*	European	2,325	1.00E-06	0.15	0.28	0.5	0.55	0.31
Heart rate	rs174547	C	*FADS1*	European	38,991	2.00E-09	6.20	0.33	0.43	1.03	0.1
QT interval	rs2478333	A	Intergenic	European	2,325	4.00E-08	0.17	0.33	0.21	4.30	0.08
QT interval	rs2650951	A	Intergenic	European	2,325	1.00E-06	-0.29	0.05	0.84	-0.40	0.28
Heart rate	rs3117035	A	Intergenic	European	2,325	1.00E-06	-0.14	0.48	0.11	-1.54	0.22
Heart rate	rs314370	C	*SLC12A9*	European	38,991	6.00E-10	7.60	0.19	0.28	1.58	0.07
QT interval	rs4318720	A	Intergenic	European	2,325	8.00E-07	0.28	0.09	0.49	1.95	0.12
PR interval	rs470490	G	*C18orf20, LOC643448*	European	2,334	5.00E-06	2.45	0.39	0.95	0.10	0.16
QT interval	rs6845865	G	*ARHGAP10*	European	2,325	7.00E-07	0.19	0.21	0.74	0.63	0.49
PR interval	rs7660702	T	*ARHGAP24*	European	12,670	3.00E-17	8.46	0.74	0.6	0.65	0.44
QT interval	rs8015016	G	*TCL6*	European	2,325	5.00E-07	-0.18	0.2	0.4	-1.90	0.22
Heart rate	rs885389	A	*GPR133*	European	2,325	4.00E-08	-0.17	0.3	0.55	-0.51	0.3
Ventricular conduction	rs9912468	G	*PRKCA*	European	40,407	1.00E-08	0.39	0.43	0.97	0.02	0.32

Appendix J. Coded Allele Frequencies for GWAS-identified SNPs in African Americans compared to the original study population. We calculated allele frequencies for all GWAS-identified SNPs in African Americans after accounting for the coded allele.

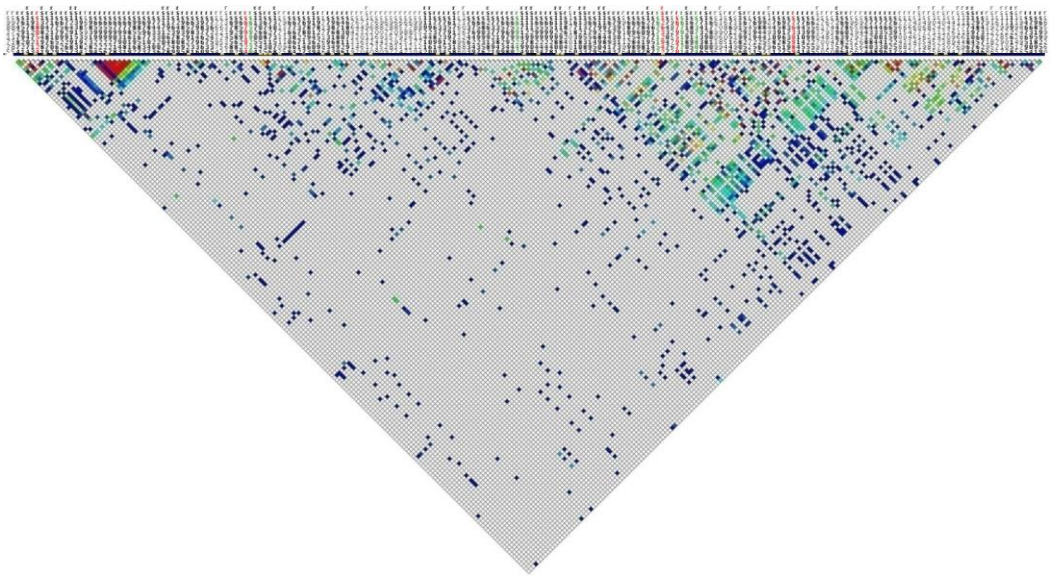
Original Study population	Reference (PubMed ID)	SNP	Coded Allele	CAF Original Study	CAF African Americans
European	20031603	rs1024020	A	0.21	0.38
European	20031603	rs10488031	A	0.07	0.17
Icelandic	20062063	rs10494366	G	0.39	0.60
European	16648850	rs10496166	A	0.13	0.43
European	20031603	rs10514995	G	0.36	0.48
European	21076409	rs10850409	A	0.27	0.23
European	19305409	rs10919071	A	0.87	0.97
European	19305409	rs11129795	A	0.23	0.16
European	20639392	rs11154022	A	0.33	0.17
European	21076409	rs11710077	T	0.21	0.11
European	21076409	rs11848785	G	0.27	0.09
European	20062060	rs11897119	C	0.39	0.37
European	19305409	rs11970286	T	0.44	0.24
European	19305408	rs12029454	A	0.15	0.26
European	19305408	rs12053903	C	0.34	0.77
European	19305408	rs12143842	T	0.24	0.12
European	19305409	rs12296050	T	0.20	0.50
European	20031603	rs12476289	A	0.07	0.01
European	20031603	rs12552736	G	0.06	0.22
European	20031603	rs12554086	A	0.39	0.15
European	19305408	rs12576239	T	0.13	0.15
European	21076409	rs13165478	A	0.36	0.53
Icelandic	20062063	rs1321311	T	0.21	0.39
European	20031603	rs13300284	A	0.04	0.013
European	20031603	rs1348582	G	0.18	0.34
European	21041692	rs1371867	C	0.42	0.74
Indian Asians	20062061	rs1415259	A	0.39	0.40
European	20031603	rs1447537	A	0.31	0.60
European	20031603	rs1484948	G	0.32	0.80
European	20031603	rs1533317	A	0.46	0.82

European	20031603	rs1541010	A	0.28	0.31
European	21076409	rs1733724	A	0.25	0.05
European	20639392	rs174547	C	0.33	0.10
European	21909110	rs17608766	C	0.16	0.02
European	20031603	rs17706439	A	0.17	0.15
European	19305409	rs17779747	T	0.35	0.09
European	21041692	rs1937332	G	0.47	0.49
Korsae	19389651	rs2070488	A	0.33	0.23
European	19305408	rs2074518	T	0.46	0.18
European	20639392	rs223116	A	0.24	0.76
European	21076409	rs2242285	A	0.42	0.26
Korsae	19389651	rs2461751	G	0.44	0.68
European	20031603	rs2478333	A	0.33	0.08
European	20062060	rs251253	C	0.40	0.64
European	20031603	rs2650951	A	0.05	0.28
European	20031603	rs2670321	C	0.27	0.63
European	20031603	rs2717128	G	0.14	0.15
European	20639392	rs281868	G	0.50	0.53
European	20031603	rs2880058	G	0.32	0.70
European	20031603	rs3110127	A	0.33	0.56
European	20031603	rs3117035	A	0.48	0.22
European	20639392	rs314370	C	0.19	0.07
Icelandic	20062063	rs365990	G	0.34	0.61
Icelandic	20062063	rs37062	G	0.24	0.47
Icelandic	20062063	rs3807375	T	0.35	0.69
European	20062060	rs3807989	A	0.40	0.62
Icelandic	20062063	rs3825214	G	0.22	0.22
European	21076409	rs4074536	C	0.29	0.44
European	20031603	rs4132509	A	0.21	0.30
European	20031603	rs4318720	A	0.09	0.12
European	20031603	rs4352210	A	0.37	0.44
European	20639392	rs452036	A	0.36	0.61
European	19305409	rs4657178	T	0.33	0.37
European	21076409	rs4687718	A	0.14	0.55
European	21041692	rs470490	G	0.39	0.16
European	20062061	rs4725982	T	0.22	0.23
European	20062060	rs4944092	G	0.32	0.45
European	20031603	rs652889	A	0.36	0.60
Icelandic	20062063	rs6795970	A	0.36	0.08

Indian Asians	20062061	rs6795970	A	0.36	0.08
European	20062060	rs6800541	C	0.40	0.09
European	20031603	rs6845865	G	0.21	0.49
Icelandic	20062063	rs727957	T	0.19	0.03
European	20031603	rs7318731	A	0.48	0.37
European	20031603	rs744016	A	0.20	0.12
European	21076409	rs7562790	G	0.40	0.59
European	20031603	rs7601713	A	0.23	0.14
European	21041692	rs7602460	A	0.40	0.43
Icelandic	20062063	rs7660702	T	0.74	0.44
European	20031603	rs7728043	G	0.48	0.52
European	21076409	rs7784776	G	0.43	0.38
European	20031603	rs789852	A	0.08	0.96
European	20031603	rs8015016	G	0.20	0.22
European	19305408	rs8049607	T	0.46	0.45
European	21076409	rs883079	C	0.29	0.31
European	20031603	rs885389	A	0.30	0.30
European	21076409	rs9851724	C	0.33	0.14
European	21076409	rs991014	T	0.42	0.98
European	21076409	rs9912468	G	0.43	0.32

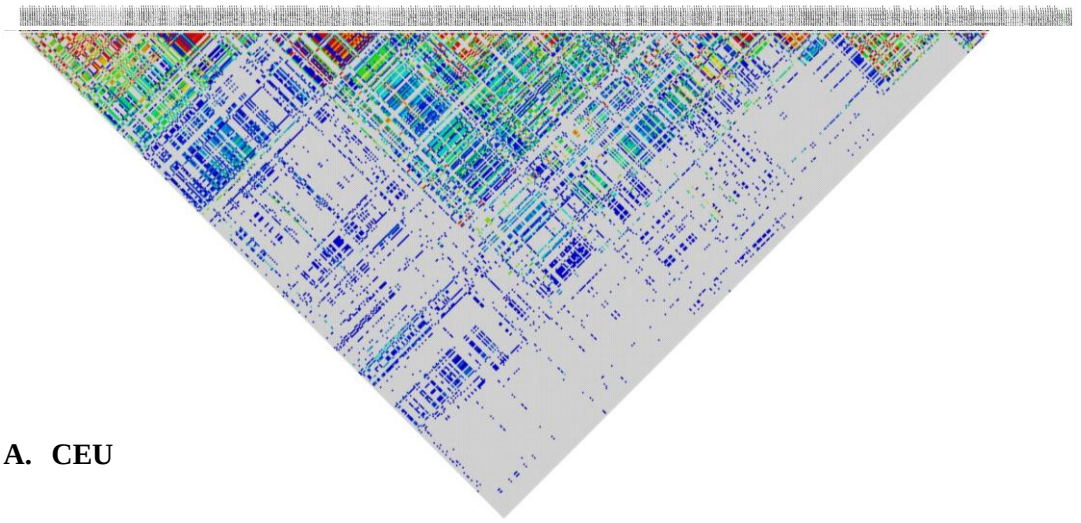


A. CEU

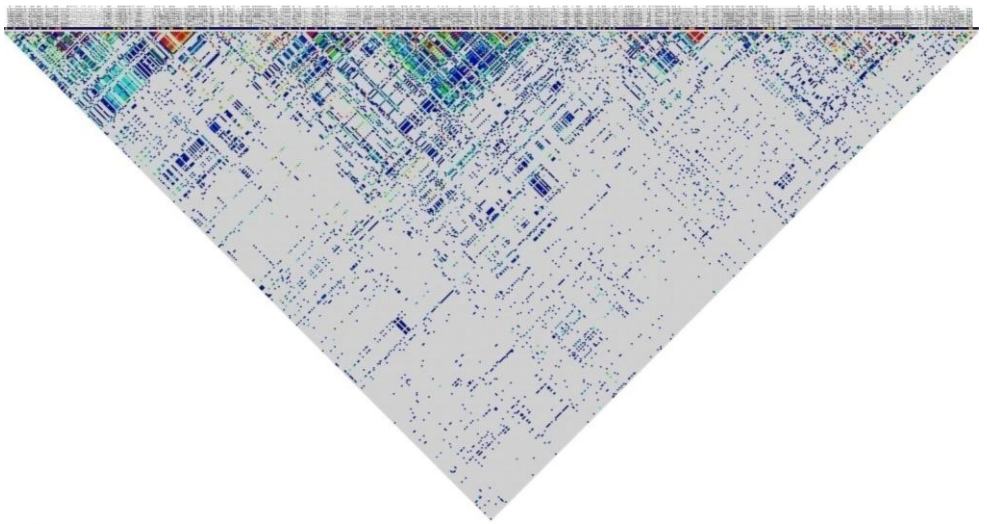


B.YRI

Appendix K. Linkage Disequilibrium plots around the *SCN5A/SCN10A* region in YRI HapMap samples compared to CEU samples.



A. CEU



B.YRI

Appendix L. Linkage Disequilibrium plots around the *NOS1AP* region in YRI HapMap samples compared to CEU samples.

Appendix M. Correlation matrix for fibrinogen and hematological traits for all NHANES III populations. Pair wise correlations (r^2) of all traits tested in Chapter III (Section A) and IV (Section B) were performed to determine if multiple traits are correlated in all three NHANES III populations. A. non-Hispanic whites, B. non-Hispanic blacks, and C. Mexican Americans.

A. NHANES III non-Hispanic whites

Variable	Fibrinogen	Platelet Count	ln Triglycerides	Platelet Distribution Width	Mean Platelet Volume	White Blood Cell Count
Fibrinogen	1.00	0.09	0.14	0.05	0.02	0.17
Platelet Count	0.09	1.00	0.04	-0.27	-0.37	0.26
ln Triglycerides	0.14	0.04	1.00	0.11	0.01	0.14
Platelet Distribution Width	0.05	-0.27	0.11	1.00	0.10	-0.05
Mean Platelet Volume	0.02	-0.37	0.01	0.10	1.00	0.03
White Blood Cell Count	0.17	0.26	0.14	-0.05	0.03	1.00

B. NHANES III non-Hispanic blacks

Variable	Fibrinogen	Platelet Count	ln Triglycerides	Platelet Distribution Width	Mean Platelet Volume	White Blood Cell Count
Fibrinogen	1.00	0.11	0.13	0.03	0.02	0.09
Platelet Count	0.11	1.00	0.09	-0.26	-0.47	0.21
ln Triglycerides	0.13	0.09	1.00	0.07	-0.05	0.16
Platelet Distribution Width	0.03	-0.26	0.07	1.00	0.19	-0.08
Mean Platelet Volume	0.02	-0.47	-0.05	0.19	1.00	0.02
White Blood Cell Count	0.09	0.21	0.16	-0.08	0.02	1.00

C. NHANES III Mexican Americans						
Variable	Fibrinogen	Platelet Count	ln Triglycerides	Platelet Distribution Width	Mean Platelet Volume	White Blood Cell Count
Fibrinogen	1.00	0.10	0.10	0.02	-0.04	0.21
Platelet Count	0.10	1.00	0.06	-0.30	-0.46	0.32
ln Triglycerides	0.10	0.06	1.00	0.04	-0.01	0.10
Platelet Distribution Width	0.02	-0.30	0.04	1.00	0.22	-0.08
Mean Platelet Volume	-0.04	-0.46	-0.01	0.22	1.00	-0.01
White Blood Cell Count	0.21	0.32	0.10	-0.08	-0.01	1.00

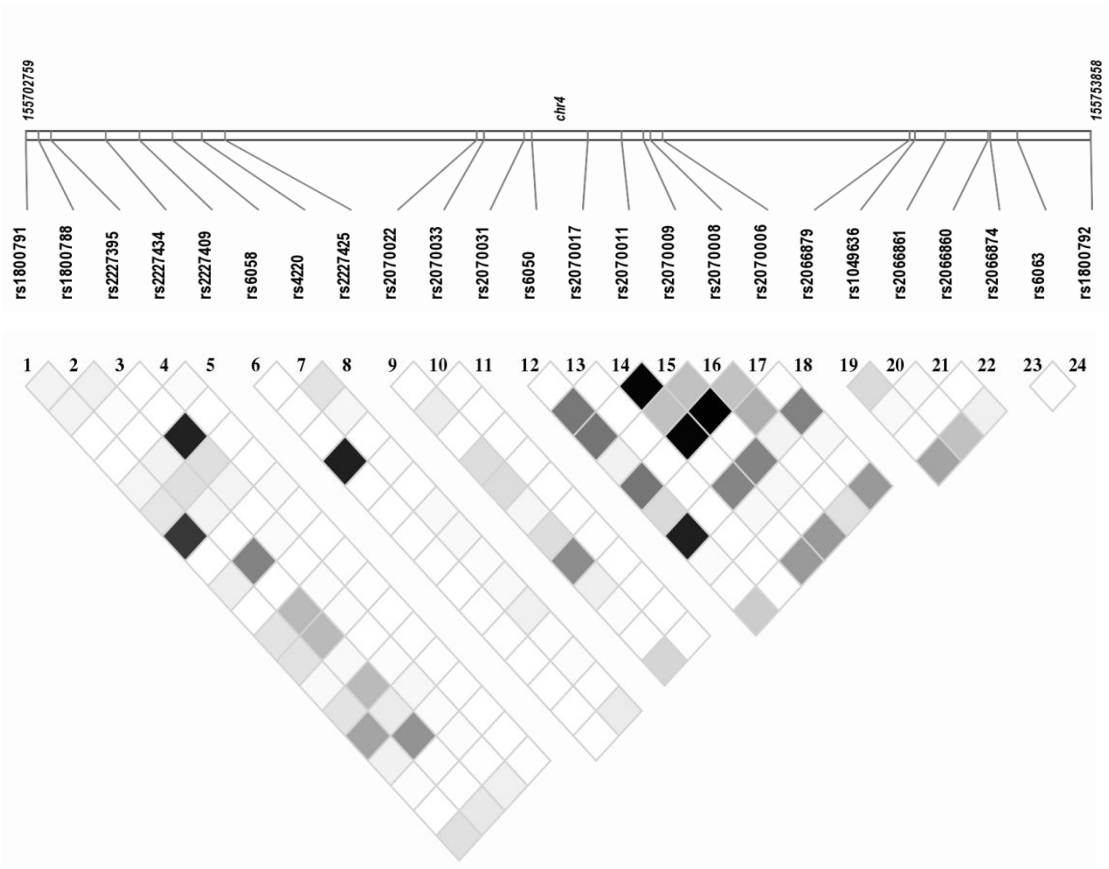
Appendix N. Coded allele frequencies across all three NHANES III study sub-populations. We selected the same coded allele across all three populations in our analysis. In some cases the major allele was coded as the effect allele in the regression model.Abbreviation: aoded allele frequency (CAF),

SNP	Alleles	Coded Allele	CAF Non-Hispanic white	CAF Non-Hispanic black	CAF Mexican Americans
rs2070022	A/G	G	0.67	0.87	0.67
rs2070033	A/G	G	0.99	0.93	0.99
rs2070031	A/G	G	<0.001	0.007	<0.001
rs6050	A/G	G	0.25	0.37	0.25
rs2070017	A/G	G	0.99	0.90	0.99
rs2070011	A/G	G	0.63	0.75	0.63
rs2070009	A/C	C	0.39	0.54	0.39
rs2070008	A/G	G	0.13	0.11	0.13
rs2070006	A/G	G	0.62	0.51	0.62
rs1800791	A/G	G	0.72	0.88	0.72
rs1800788	A/G	G	0.76	0.90	0.76
rs2227395	A/G	G	0.86	0.90	0.86
rs2227409	A/C	C	<0.001	<0.001	<0.001
rs2227434	C/T	T	0.99	0.99	0.99
rs6058	A/C	C	0.99	0.93	0.99
rs4220	A/G	G	0.86	0.91	0.86
rs2227425	C/G	G	<0.001	0.02	<0.001
rs2066879	C/T	C	0.99	0.97	0.99
rs1049636	C/T	C	0.47	0.24	0.47
rs2066861	A/G	G	0.77	0.71	0.77

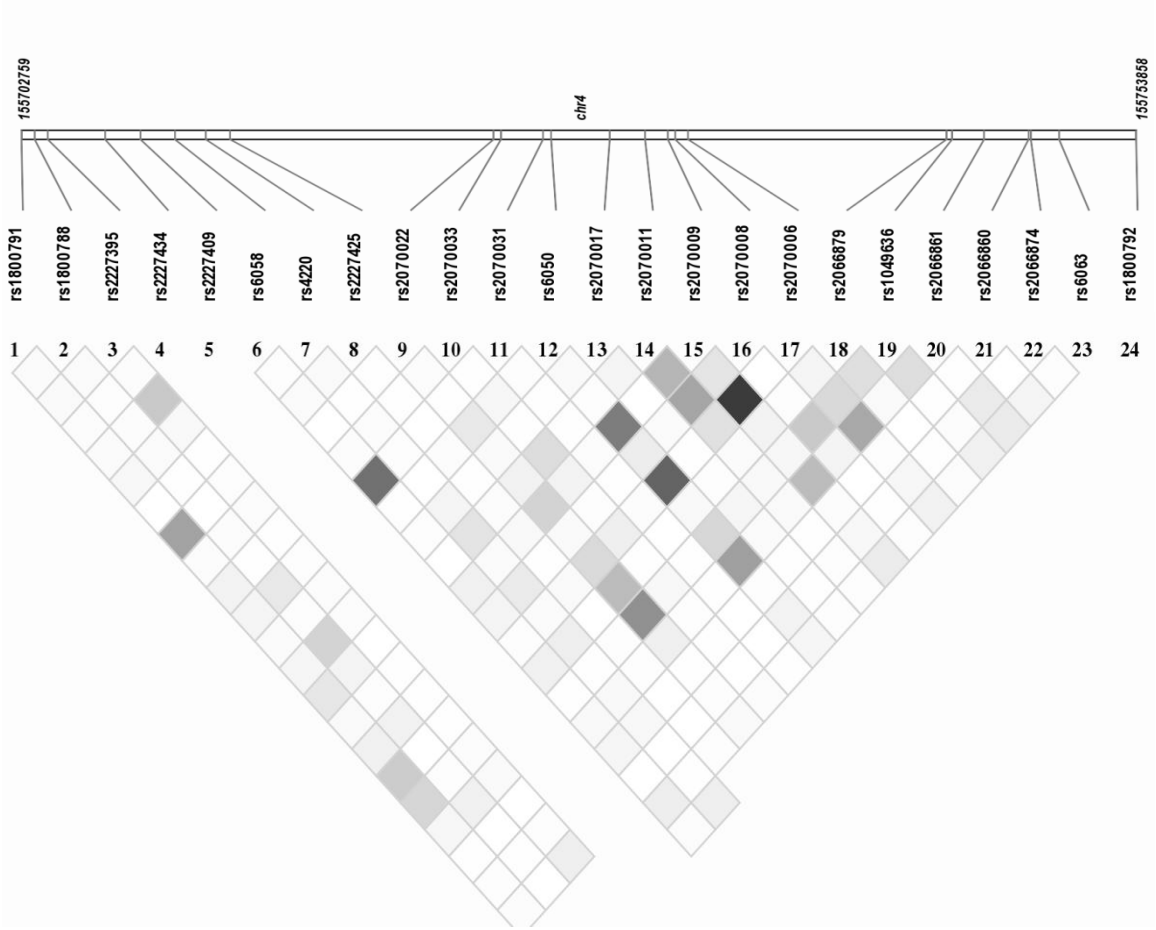
rs2066860	A/G	G	0.99	0.49	0.99
rs2066874	A/G	G	<0.001	0.02	<0.001
rs6063	A/G	G	0.99	0.99	0.99
rs1800792	A/G	G	0.27	0.05	0.27

Appendix O. Pair-wise linkage disequilibrium plots for fibrinogen SNPs across each study population. Pair-wise LD (r^2) was calculated separately for each NHANES III sub-population (indicated by letter).

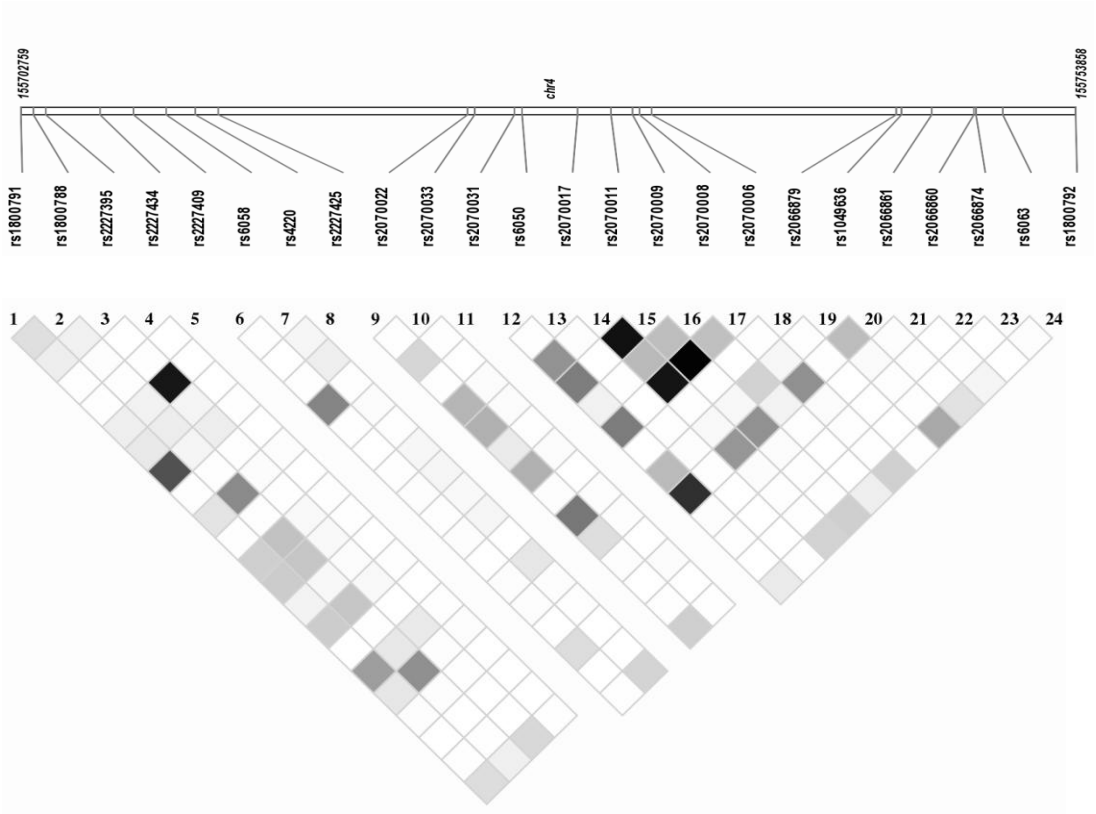
A. Non-Hispanic whites



B. non-Hispanic blacks



C. Mexican Americans



Appendix P. Association results for all fibrinogen SNPs and traits tested in each NHANES III sub-population. We performed linear regression adjusting for age, sex, BMI and smoking status for 25 SNPs, assuming an additive model for six hematological traits. Note some tests of association were not performed in sub-populations with low minor allele frequencies and/or variants that were not in Hardy-Weinberg Equilibrium. Abbreviations: NHW= Non-Hispanic White, NHB= Non-Hispanic Black, MA= Mexican American, MPV= mean platelet count, FB = fibrinogen, PDW= platelet distribution width, PC = platelet count, WBC = white blood cell count, LN_Tri = natural logarithm of triglycerides.

SNP	Trait	NHANES III Population	N	β	SE	p-value	95% CI
rs1049636	FB	NHW	1392	0.04	0.02	0.04	0, 0.08
rs1049636	FB	MA	664	0.04	0.03	0.08	-0.01, 0.09
rs1049636	FB	NHB	626	0.06	0.03	0.07	0, 0.13
rs1800788	FB	NHB	627	-0.02	0.05	0.64	-0.11, 0.07
rs1800788	FB	MA	659	0.00	0.03	0.91	-0.06, 0.06
rs1800788	FB	NHW	1389	0.04	0.02	0.11	-0.01, 0.09
rs1800791	FB	NHB	627	-0.08	0.04	0.07	-0.16, 0.01
rs1800791	FB	MA	661	-0.03	0.03	0.26	-0.09, 0.03
rs1800791	FB	NHW	1391	0.00	0.03	0.89	-0.05, 0.05
rs1800792	FB	MA	666	-0.03	0.03	0.36	-0.08, 0.03
rs1800792	FB	NHB	627	-0.01	0.04	0.71	-0.08, 0.06
rs1800792	FB	NHW	1395	0.01	0.02	0.72	-0.03, 0.04
rs2066860	FB	NHW	1393	-0.04	0.05	0.36	-0.13, 0.05
rs2066860	FB	NHB	627	0.05	0.18	0.79	-0.3, 0.4
rs2066860	FB	MA	665	0.12	0.10	0.26	-0.08, 0.31
rs2066861	FB	MA	663	0.04	0.03	0.24	-0.02, 0.1
rs2066861	FB	NHW	1393	0.05	0.02	0.01	0.01, 0.1
rs2066861	FB	NHB	625	0.05	0.03	0.12	-0.01, 0.11
rs2066874	FB	MA	666	Not tested			
rs2066874	FB	NHB	626	-0.28	0.08	6.78E-04	-0.45, -0.12
rs2066874	FB	NHW	1396	0.27	0.49	0.58	-0.69, 1.23
rs2066879	FB	NHW	1397	-0.38	0.34	0.28	-1.05, 0.3
rs2066879	FB	MA	666	-0.37	0.24	0.12	-0.84, 0.1
rs2066879	FB	NHB	627	-0.13	0.07	0.07	-0.26, 0.01

rs2070006	FB	NHB	620	0.01	0.03	0.85	-0.05, 0.06
rs2070006	FB	NHW	1391	0.02	0.02	0.20	-0.01, 0.06
rs2070006	FB	MA	660	0.03	0.03	0.28	-0.02, 0.08
rs2070008	FB	MA	661	-0.02	0.04	0.57	-0.1, 0.05
rs2070008	FB	NHB	627	-0.01	0.05	0.75	-0.1, 0.08
rs2070008	FB	NHW	1390	0.01	0.03	0.61	-0.04, 0.07
rs2070009	FB	NHW	1387	-0.03	0.02	0.16	-0.06, 0.01
rs2070009	FB	MA	662	-0.03	0.03	0.22	-0.09, 0.02
rs2070009	FB	NHB	621	0.00	0.03	0.94	-0.06, 0.05
rs2070011	FB	NHB	626	0.02	0.03	0.45	-0.04, 0.09
rs2070011	FB	NHW	1387	0.03	0.02	0.14	-0.01, 0.07
rs2070011	FB	MA	663	0.03	0.03	0.29	-0.02, 0.08
rs2070017	FB	NHB	625	0.08	0.05	0.08	-0.01, 0.17
rs2070017	FB	MA	666	0.14	0.13	0.28	-0.11, 0.39
rs2070017	FB	NHW	1397	0.24	0.22	0.28	-0.19, 0.67
rs2070022	FB	NHB	623	-0.08	0.04	0.06	-0.17, 0
rs2070022	FB	MA	659	-0.07	0.03	0.01	-0.13, -0.02
rs2070022	FB	NHW	1393	-0.03	0.02	0.23	-0.08, 0.02
rs2070031	FB	NHW	1397	Not tested			
rs2070031	FB	NHB	627	0.16	0.16	0.32	-0.16, 0.47
rs2070031	FB	MA	666	0.54	0.48	0.26	-0.41, 1.49
rs2070033	FB	MA	666	-0.16	0.15	0.26	-0.45, 0.12
rs2070033	FB	NHW	1395	-0.05	0.24	0.85	-0.52, 0.43
rs2070033	FB	NHB	627	0.07	0.05	0.16	-0.03, 0.17
rs2227395	FB	NHW	1391	-0.08	0.02	6.94E-04	-0.13, -0.03
rs2227395	FB	MA	665	-0.02	0.04	0.58	-0.1, 0.05
rs2227395	FB	NHB	625	-0.02	0.05	0.61	-0.11, 0.07
rs2227409	FB	MA	666	Not tested			
rs2227409	FB	NHB	627	Not tested			
rs2227409	FB	NHW	1396	Not tested			
rs2227425	FB	MA	666	Not tested			
rs2227425	FB	NHW	1397	Not tested			
rs2227425	FB	NHB	627	-0.03	0.10	0.78	-0.21, 0.16
rs2227434	FB	NHW	1396	0.16	0.20	0.42	-0.23, 0.55

rs2227434	FB	NHB	627	0.27	0.50	0.59	-0.71, 1.25
rs2227434	FB	MA	666	0.50	0.28	0.07	-0.04, 1.04
rs4220	FB	NHW	1391	-0.08	0.02	2.08E-03	-0.13, -0.03
rs4220	FB	MA	666	-0.04	0.04	0.36	-0.11, 0.04
rs4220	FB	NHB	625	0.08	0.05	0.10	-0.02, 0.18
rs6050	FB	NHW	1389	-0.04	0.02	0.04	-0.09, 0
rs6050	FB	MA	661	-0.04	0.03	0.19	-0.1, 0.02
rs6050	FB	NHB	624	0.02	0.03	0.52	-0.04, 0.08
rs6058	FB	MA	666	-0.17	0.20	0.38	-0.56, 0.21
rs6058	FB	NHW	1397	-0.05	0.24	0.85	-0.52, 0.43
rs6058	FB	NHB	627	0.07	0.05	0.18	-0.03, 0.17
rs6063	FB	NHB	627	Not tested			
rs6063	FB	NHW	1397	-0.03	0.12	0.80	-0.26, 0.2
rs6063	FB	MA	666	0.11	0.21	0.60	-0.31, 0.53
rs1049636	LN_Tri	NHW	1459	-0.01	0.01	0.22	-0.04, 0.01
rs1049636	LN_Tri	NHB	1320	0.01	0.02	0.38	-0.02, 0.04
rs1049636	LN_Tri	MA	1004	0.02	0.01	0.15	-0.01, 0.04
rs1800788	LN_Tri	MA	1000	0.00	0.02	0.84	-0.03, 0.03
rs1800788	LN_Tri	NHW	1457	0.00	0.01	0.86	-0.03, 0.03
rs1800788	LN_Tri	NHB	1325	0.02	0.02	0.46	-0.03, 0.06
rs1800791	LN_Tri	NHB	1321	-0.03	0.02	0.19	-0.07, 0.01
rs1800791	LN_Tri	MA	1004	-0.02	0.01	0.15	-0.05, 0.01
rs1800791	LN_Tri	NHW	1459	0.01	0.02	0.48	-0.02, 0.04
rs1800792	LN_Tri	MA	1005	-0.03	0.01	0.03	-0.06, 0
rs1800792	LN_Tri	NHB	1325	-0.03	0.02	0.08	-0.06, 0
rs1800792	LN_Tri	NHW	1460	0.00	0.01	0.67	-0.02, 0.03
rs2066860	LN_Tri	NHW	1460	0.02	0.03	0.50	-0.04, 0.08
rs2066860	LN_Tri	NHB	1324	0.04	0.08	0.56	-0.1, 0.19
rs2066860	LN_Tri	MA	1005	0.13	0.06	0.02	0.02, 0.24
rs2066861	LN_Tri	MA	1000	-0.02	0.02	0.31	-0.05, 0.01
rs2066861	LN_Tri	NHW	1458	-0.01	0.01	0.56	-0.03, 0.02
rs2066861	LN_Tri	NHB	1319	-0.01	0.01	0.70	-0.03, 0.02
rs2066874	LN_Tri	NHB	1324	0.03	0.04	0.37	-0.04, 0.11
rs2066874	LN_Tri	MA	1006	0.34	0.30	0.25	-0.24, 0.92
rs2066874	LN_Tri	NHW	1462	0.37	0.30	0.22	-0.22, 0.96
rs2066879	LN_Tri	MA	1006	-0.06	0.11	0.59	-0.28, 0.16
rs2066879	LN_Tri	NHB	1325	-0.03	0.03	0.34	-0.09, 0.03
rs2066879	LN_Tri	NHW	1463	0.21	0.21	0.33	-0.21, 0.63

rs2070006	LN_Tri	MA	998	-0.01	0.01	0.39	-0.04, 0.01
rs2070006	LN_Tri	NHW	1452	0.00	0.01	0.81	-0.02, 0.02
rs2070006	LN_Tri	NHB	1311	0.00	0.01	0.89	-0.02, 0.03
rs2070008	LN_Tri	NHW	1455	-0.02	0.02	0.18	-0.05, 0.01
rs2070008	LN_Tri	MA	1004	0.01	0.02	0.54	-0.03, 0.05
rs2070008	LN_Tri	NHB	1325	0.01	0.02	0.67	-0.03, 0.05
rs2070009	LN_Tri	NHW	1451	0.00	0.01	0.72	-0.03, 0.02
rs2070009	LN_Tri	MA	1000	0.01	0.01	0.41	-0.02, 0.04
rs2070009	LN_Tri	NHB	1319	0.01	0.01	0.64	-0.02, 0.03
rs2070011	LN_Tri	MA	1001	-0.01	0.01	0.56	-0.03, 0.02
rs2070011	LN_Tri	NHB	1325	0.00	0.01	0.86	-0.03, 0.03
rs2070011	LN_Tri	NHW	1455	0.01	0.01	0.50	-0.01, 0.03
rs2070017	LN_Tri	MA	1006	-0.11	0.08	0.19	-0.26, 0.05
rs2070017	LN_Tri	NHB	1322	0.02	0.02	0.46	-0.03, 0.06
rs2070017	LN_Tri	NHW	1462	0.37	0.21	0.09	-0.05, 0.78
rs2070022	LN_Tri	NHB	1323	-0.03	0.02	0.19	-0.06, 0.01
rs2070022	LN_Tri	MA	1002	-0.02	0.01	0.12	-0.05, 0.01
rs2070022	LN_Tri	NHW	1460	0.01	0.01	0.60	-0.02, 0.04
rs2070031	LN_Tri	NHW	1463	Not tested			
rs2070031	LN_Tri	MA	1006	-0.50	0.30	0.10	-1.08, 0.09
rs2070031	LN_Tri	NHB	1325	0.05	0.08	0.52	-0.1, 0.2
rs2070033	LN_Tri	NHW	1462	-0.24	0.21	0.26	-0.66, 0.18
rs2070033	LN_Tri	NHB	1324	-0.02	0.03	0.54	-0.07, 0.03
rs2070033	LN_Tri	MA	1006	0.08	0.08	0.34	-0.08, 0.24
rs2227395	LN_Tri	NHW	1460	-0.01	0.01	0.37	-0.04, 0.02
rs2227395	LN_Tri	MA	1003	0.00	0.02	0.86	-0.03, 0.04
rs2227395	LN_Tri	NHB	1323	0.01	0.02	0.54	-0.03, 0.05
rs2227409	LN_Tri	MA	1006	Not tested			
rs2227409	LN_Tri	NHW	1462	Not tested			
rs2227409	LN_Tri	NHB	1325	0.04	0.19	0.84	-0.34, 0.42
rs2227425	LN_Tri	MA	1006	Not tested			
rs2227425	LN_Tri	NHW	1463	Not tested			
rs2227425	LN_Tri	NHB	1325	-0.02	0.04	0.59	-0.11, 0.06
rs2227434	LN_Tri	NHW	1460	-0.08	0.12	0.53	-0.32, 0.16
rs2227434	LN_Tri	MA	1006	0.00	0.15	0.98	-0.3, 0.29
rs2227434	LN_Tri	NHB	1323	0.01	0.34	0.98	-0.65, 0.67
rs4220	LN_Tri	NHW	1454	-0.02	0.01	0.29	-0.04, 0.01
rs4220	LN_Tri	MA	1004	0.01	0.02	0.72	-0.03, 0.04
rs4220	LN_Tri	NHB	1323	0.01	0.02	0.74	-0.04, 0.05
rs6050	LN_Tri	NHB	1317	0.00	0.01	0.81	-0.03, 0.02

rs6050	LN_Tri	NHW	1454	0.01	0.01	0.50	-0.02, 0.03
rs6050	LN_Tri	MA	997	0.01	0.02	0.71	-0.02, 0.04
rs6058	LN_Tri	NHW	1463	-0.24	0.21	0.26	-0.66, 0.18
rs6058	LN_Tri	MA	1006	-0.11	0.11	0.34	-0.33, 0.11
rs6058	LN_Tri	NHB	1324	-0.03	0.03	0.26	-0.08, 0.02
rs6063	LN_Tri	MA	1006	-0.11	0.15	0.46	-0.4, 0.18
rs6063	LN_Tri	NHB	1325	0.05	0.19	0.79	-0.33, 0.43
rs6063	LN_Tri	NHW	1463	0.11	0.08	0.15	-0.04, 0.27
rs1049636	MPV	MA	1674	-0.04	0.03	0.27	-0.1, 0.03
rs1049636	MPV	NHW	2305	0.00	0.03	0.97	-0.05, 0.06
rs1049636	MPV	NHB	1652	0.02	0.04	0.56	-0.05, 0.1
rs1800788	MPV	MA	1668	-0.03	0.04	0.44	-0.11, 0.05
rs1800788	MPV	NHW	2302	0.00	0.03	0.93	-0.06, 0.07
rs1800788	MPV	NHB	1657	0.00	0.06	0.96	-0.12, 0.11
rs1800791	MPV	NHB	1652	-0.05	0.05	0.35	-0.15, 0.05
rs1800791	MPV	MA	1671	0.00	0.04	0.93	-0.07, 0.08
rs1800791	MPV	NHW	2304	0.01	0.04	0.78	-0.06, 0.08
rs1800792	MPV	NHB	1657	0.01	0.04	0.74	-0.07, 0.1
rs1800792	MPV	NHW	2308	0.01	0.03	0.82	-0.05, 0.06
rs1800792	MPV	MA	1679	0.01	0.04	0.88	-0.07, 0.08
rs2066860	MPV	NHB	1656	-0.14	0.20	0.47	-0.53, 0.25
rs2066860	MPV	NHW	2308	-0.08	0.07	0.28	-0.21, 0.06
rs2066860	MPV	MA	1679	-0.02	0.14	0.87	-0.31, 0.26
rs2066861	MPV	MA	1673	-0.06	0.04	0.13	-0.14, 0.02
rs2066861	MPV	NHB	1651	0.00	0.04	0.93	-0.07, 0.08
rs2066861	MPV	NHW	2307	0.01	0.03	0.74	-0.05, 0.07
rs2066874	MPV	NHW	2311	-0.66	0.91	0.47	-2.45, 1.13
rs2066874	MPV	NHB	1654	-0.12	0.10	0.27	-0.32, 0.09
rs2066874	MPV	MA	1680	0.02	0.99	0.99	-1.92, 1.95
rs2066879	MPV	NHB	1657	-0.10	0.09	0.24	-0.27, 0.07
rs2066879	MPV	NHW	2313	0.07	0.65	0.91	-1.19, 1.34
rs2066879	MPV	MA	1679	0.60	0.33	0.07	-0.04, 1.25
rs2070006	MPV	NHB	1642	-0.01	0.03	0.82	-0.08, 0.06
rs2070006	MPV	NHW	2300	0.00	0.03	0.88	-0.06, 0.05
rs2070006	MPV	MA	1665	0.00	0.03	0.95	-0.07, 0.07
rs2070008	MPV	MA	1673	-0.07	0.05	0.15	-0.17, 0.03
rs2070008	MPV	NHB	1657	-0.02	0.05	0.69	-0.13, 0.09
rs2070008	MPV	NHW	2301	0.03	0.04	0.50	-0.05, 0.11
rs2070009	MPV	NHB	1648	-0.02	0.03	0.59	-0.09, 0.05
rs2070009	MPV	MA	1669	0.00	0.03	0.98	-0.07, 0.07

rs2070009	MPV	NHW	2296	0.01	0.03	0.71	-0.04, 0.06
rs2070011	MPV	NHW	2299	-0.01	0.03	0.85	-0.06, 0.05
rs2070011	MPV	NHB	1656	0.01	0.04	0.74	-0.06, 0.09
rs2070011	MPV	MA	1672	0.01	0.03	0.75	-0.06, 0.08
rs2070017	MPV	NHW	2312	-0.41	0.41	0.31	-1.22, 0.39
rs2070017	MPV	NHB	1653	0.08	0.06	0.13	-0.03, 0.19
rs2070017	MPV	MA	1680	0.16	0.21	0.44	-0.25, 0.57
rs2070022	MPV	NHB	1651	-0.05	0.05	0.36	-0.15, 0.05
rs2070022	MPV	MA	1665	-0.01	0.04	0.73	-0.08, 0.06
rs2070022	MPV	NHW	2307	0.02	0.03	0.65	-0.05, 0.08
rs2070031	MPV	NHW	2313	Not tested			
rs2070031	MPV	NHB	1657	0.11	0.20	0.59	-0.29, 0.5
rs2070031	MPV	MA	1680	0.71	0.99	0.47	-1.23, 2.65
rs2070033	MPV	MA	1680	-0.11	0.23	0.65	-0.56, 0.35
rs2070033	MPV	NHB	1656	-0.05	0.07	0.49	-0.18, 0.08
rs2070033	MPV	NHW	2311	0.35	0.35	0.31	-0.33, 1.03
rs2227395	MPV	NHB	1654	-0.07	0.05	0.20	-0.18, 0.04
rs2227395	MPV	MA	1676	0.02	0.05	0.68	-0.08, 0.11
rs2227395	MPV	NHW	2304	0.03	0.03	0.36	-0.04, 0.1
rs2227409	MPV	MA	1680	Not tested			
rs2227409	MPV	NHW	2312	Not tested			
rs2227409	MPV	NHB	1657	1.11	0.57	0.05	0, 2.22
rs2227425	MPV	NHW	2313	Not tested			
rs2227425	MPV	NHB	1657	-0.12	0.12	0.32	-0.34, 0.11
rs2227425	MPV	MA	1680	0.11	0.99	0.91	-1.82, 2.05
rs2227434	MPV	MA	1680	-0.31	0.44	0.48	-1.17, 0.56
rs2227434	MPV	NHW	2310	0.25	0.31	0.42	-0.35, 0.85
rs2227434	MPV	NHB	1654	0.68	0.69	0.32	-0.67, 2.04
rs4220	MPV	NHW	2300	0.04	0.04	0.28	-0.03, 0.11
rs4220	MPV	MA	1676	0.04	0.05	0.41	-0.06, 0.14
rs4220	MPV	NHB	1655	0.04	0.06	0.56	-0.08, 0.15
rs6050	MPV	NHB	1649	-0.01	0.04	0.77	-0.08, 0.06
rs6050	MPV	NHW	2298	0.00	0.03	0.93	-0.06, 0.06
rs6050	MPV	MA	1667	0.03	0.04	0.39	-0.04, 0.11
rs6058	MPV	MA	1680	-0.38	0.31	0.23	-0.99, 0.23
rs6058	MPV	NHB	1656	-0.02	0.07	0.79	-0.15, 0.12
rs6058	MPV	NHW	2313	0.42	0.37	0.26	-0.32, 1.15
rs6063	MPV	NHW	2313	0.15	0.19	0.42	-0.22, 0.53
rs6063	MPV	MA	1680	0.19	0.33	0.56	-0.45, 0.84
rs6063	MPV	NHB	1657	0.47	0.49	0.34	-0.49, 1.43

rs1049636	PC	NHB	1623	1.47	2.47	0.55	-3.38, 6.33
rs1049636	PC	NHW	2284	2.21	1.75	0.21	-1.21, 5.64
rs1049636	PC	MA	1658	3.27	2.06	0.11	-0.77, 7.3
rs1800788	PC	NHW	2281	-1.50	2.12	0.48	-5.66, 2.66
rs1800788	PC	MA	1652	-0.33	2.45	0.89	-5.14, 4.47
rs1800788	PC	NHB	1628	1.47	3.64	0.69	-5.67, 8.62
rs1800791	PC	NHB	1623	-3.20	3.28	0.33	-9.63, 3.23
rs1800791	PC	MA	1655	-1.50	2.33	0.52	-6.08, 3.07
rs1800791	PC	NHW	2283	-0.28	2.26	0.90	-4.72, 4.15
rs1800792	PC	MA	1663	-5.32	2.30	0.02	-9.83, -0.81
rs1800792	PC	NHW	2287	-2.20	1.64	0.18	-5.41, 1.02
rs1800792	PC	NHB	1628	1.11	2.79	0.69	-4.37, 6.58
rs2066860	PC	MA	1663	4.72	8.94	0.60	-12.82, 22.25
rs2066860	PC	NHW	2287	6.35	4.27	0.14	-2.03, 14.73
rs2066860	PC	NHB	1627	15.09	12.26	0.22	-8.96, 39.13
rs2066861	PC	NHW	2286	-0.36	1.91	0.85	-4.12, 3.39
rs2066861	PC	MA	1657	-0.27	2.49	0.91	-5.15, 4.6
rs2066861	PC	NHB	1622	2.45	2.31	0.29	-2.07, 6.98
rs2066874	PC	NHW	2290	-95.82	56.04	0.09	-205.71, 14.07
rs2066874	PC	MA	1664	-19.61	60.31	0.75	-137.9, 98.68
rs2066874	PC	NHB	1625	-4.86	6.58	0.46	-17.76, 8.04
rs2066879	PC	NHW	2292	-41.85	39.61	0.29	-119.52, 35.82
rs2066879	PC	MA	1663	-29.11	20.11	0.15	-68.56, 10.33
rs2066879	PC	NHB	1628	-3.08	5.40	0.57	-13.67, 7.5
rs2070006	PC	NHW	2280	-1.97	1.69	0.25	-5.29, 1.35
rs2070006	PC	MA	1649	-1.82	2.14	0.39	-6.01, 2.37
rs2070006	PC	NHB	1613	2.39	2.20	0.28	-1.92, 6.69
rs2070008	PC	NHB	1628	0.35	3.45	0.92	-6.41, 7.11
rs2070008	PC	MA	1657	2.06	3.13	0.51	-4.08, 8.2
rs2070008	PC	NHW	2282	2.47	2.45	0.31	-2.34, 7.27
rs2070009	PC	NHB	1619	-1.61	2.17	0.46	-5.87, 2.65
rs2070009	PC	MA	1653	1.59	2.13	0.45	-2.58, 5.77
rs2070009	PC	NHW	2276	1.86	1.69	0.27	-1.47, 5.18
rs2070011	PC	NHW	2278	-1.91	1.70	0.26	-5.24, 1.43
rs2070011	PC	MA	1656	-1.81	2.14	0.40	-6.02, 2.39
rs2070011	PC	NHB	1627	0.73	2.43	0.77	-4.05, 5.5
rs2070017	PC	MA	1664	-15.32	12.67	0.23	-40.17, 9.53
rs2070017	PC	NHB	1624	-8.12	3.47	0.02	-14.93, -1.31
rs2070017	PC	NHW	2291	19.29	25.11	0.44	-29.94, 68.53
rs2070022	PC	MA	1649	-1.83	2.22	0.41	-6.17, 2.52

rs2070022	PC	NHW	2286	-0.71	2.14	0.74	-4.91, 3.48
rs2070022	PC	NHB	1622	-0.60	3.17	0.85	-6.82, 5.63
rs2070031	PC	NHW	2292	Not tested			
rs2070031	PC	MA	1664	-68.26	60.53	0.26	-186.99, 50.47
rs2070031	PC	NHB	1628	-11.92	12.50	0.34	-36.45, 12.6
rs2070033	PC	NHW	2290	-10.37	21.22	0.63	-51.97, 31.24
rs2070033	PC	MA	1664	-6.50	14.29	0.65	-34.52, 21.53
rs2070033	PC	NHB	1628	2.10	4.18	0.62	-6.1, 10.29
rs2227395	PC	MA	1660	-1.17	2.97	0.69	-7, 4.66
rs2227395	PC	NHW	2284	1.05	2.13	0.62	-3.12, 5.22
rs2227395	PC	NHB	1625	2.40	3.36	0.47	-4.18, 8.99
rs2227409	PC	MA	1664	Not tested			
rs2227409	PC	NHW	2291	Not tested			
rs2227409	PC	NHB	1628	-43.77	35.16	0.21	-112.73, 25.18
rs2227425	PC	NHW	2292	Not tested			
rs2227425	PC	MA	1664	6.23	60.34	0.92	-112.12, 124.58
rs2227425	PC	NHB	1628	7.07	7.29	0.33	-7.22, 21.36
rs2227434	PC	NHB	1625	-37.06	43.04	0.39	-121.48, 47.35
rs2227434	PC	NHW	2289	-19.33	18.72	0.30	-56.05, 17.38
rs2227434	PC	MA	1664	5.31	27.02	0.84	-47.68, 58.3
rs4220	PC	NHB	1626	-5.91	3.73	0.11	-13.23, 1.41
rs4220	PC	MA	1660	-1.80	3.04	0.55	-7.75, 4.16
rs4220	PC	NHW	2279	1.79	2.21	0.42	-2.55, 6.13
rs6050	PC	NHB	1620	-1.89	2.21	0.39	-6.22, 2.44
rs6050	PC	NHW	2277	0.81	1.90	0.67	-2.91, 4.53
rs6050	PC	MA	1651	1.05	2.44	0.67	-3.73, 5.84
rs6058	PC	NHW	2292	0.33	22.92	0.99	-44.61, 45.27
rs6058	PC	NHB	1628	3.29	4.30	0.44	-5.13, 11.72
rs6058	PC	MA	1664	35.76	20.13	0.08	-3.72, 75.24
rs6063	PC	NHB	1628	-27.15	30.44	0.37	-86.85, 32.54
rs6063	PC	MA	1664	-18.24	20.14	0.37	-57.74, 21.27
rs6063	PC	NHW	2292	2.36	11.74	0.84	-20.66, 25.38
rs1049636	PDW	MA	1656	-0.02	0.02	0.33	-0.05, 0.02
rs1049636	PDW	NHB	1623	0.00	0.02	0.93	-0.05, 0.04
rs1049636	PDW	NHW	2277	0.01	0.01	0.34	-0.01, 0.04
rs1800788	PDW	MA	1649	-0.01	0.02	0.51	-0.06, 0.03
rs1800788	PDW	NHW	2274	0.01	0.02	0.53	-0.02, 0.04
rs1800788	PDW	NHB	1628	0.04	0.03	0.23	-0.03, 0.11
rs1800791	PDW	NHW	2276	-0.01	0.02	0.54	-0.05, 0.02
rs1800791	PDW	NHB	1623	0.02	0.03	0.55	-0.04, 0.08

rs1800791	PDW	MA	1652	0.03	0.02	0.14	-0.01, 0.07
rs1800792	PDW	NHB	1628	-0.01	0.03	0.62	-0.06, 0.04
rs1800792	PDW	MA	1660	-0.01	0.02	0.76	-0.05, 0.03
rs1800792	PDW	NHW	2280	0.01	0.01	0.40	-0.01, 0.04
rs2066860	PDW	NHB	1627	-0.08	0.11	0.49	-0.3, 0.14
rs2066860	PDW	MA	1660	-0.03	0.08	0.68	-0.19, 0.12
rs2066860	PDW	NHW	2280	-0.01	0.03	0.77	-0.08, 0.06
rs2066861	PDW	MA	1654	-0.04	0.02	0.09	-0.08, 0.01
rs2066861	PDW	NHB	1622	-0.04	0.02	0.09	-0.08, 0.01
rs2066861	PDW	NHW	2279	0.03	0.02	0.07	0, 0.06
rs2066874	PDW	NHW	2283	-0.56	0.45	0.21	-1.44, 0.32
rs2066874	PDW	MA	1661	-0.14	0.54	0.79	-1.21, 0.92
rs2066874	PDW	NHB	1625	0.08	0.06	0.16	-0.03, 0.2
rs2066879	PDW	NHB	1628	-0.02	0.05	0.63	-0.12, 0.07
rs2066879	PDW	MA	1660	0.06	0.18	0.75	-0.3, 0.41
rs2066879	PDW	NHW	2285	0.77	0.32	0.02	0.15, 1.39
rs2070006	PDW	MA	1646	-0.03	0.02	0.09	-0.07, 0.01
rs2070006	PDW	NHB	1613	-0.01	0.02	0.50	-0.05, 0.03
rs2070006	PDW	NHW	2272	0.01	0.01	0.27	-0.01, 0.04
rs2070008	PDW	NHB	1628	0.00	0.03	0.98	-0.06, 0.06
rs2070008	PDW	NHW	2273	0.02	0.02	0.35	-0.02, 0.06
rs2070008	PDW	MA	1654	0.02	0.03	0.41	-0.03, 0.08
rs2070009	PDW	NHW	2268	-0.02	0.01	0.25	-0.04, 0.01
rs2070009	PDW	NHB	1619	0.00	0.02	0.94	-0.04, 0.04
rs2070009	PDW	MA	1650	0.04	0.02	0.07	0, 0.07
rs2070011	PDW	MA	1653	-0.03	0.02	0.08	-0.07, 0
rs2070011	PDW	NHB	1627	-0.01	0.02	0.73	-0.05, 0.04
rs2070011	PDW	NHW	2271	0.02	0.01	0.21	-0.01, 0.04
rs2070017	PDW	NHW	2284	-0.01	0.20	0.97	-0.4, 0.39
rs2070017	PDW	NHB	1624	0.05	0.03	0.11	-0.01, 0.11
rs2070017	PDW	MA	1661	0.06	0.11	0.59	-0.16, 0.29
rs2070022	PDW	NHW	2279	-0.01	0.02	0.69	-0.04, 0.03
rs2070022	PDW	NHB	1622	-0.01	0.03	0.75	-0.07, 0.05
rs2070022	PDW	MA	1646	0.03	0.02	0.09	-0.01, 0.07
rs2070031	PDW	NHW	2285	Not tested			
rs2070031	PDW	MA	1661	-0.16	0.54	0.76	-1.23, 0.9
rs2070031	PDW	NHB	1628	0.17	0.12	0.16	-0.07, 0.41
rs2070033	PDW	NHW	2283	-0.04	0.17	0.80	-0.38, 0.29
rs2070033	PDW	NHB	1627	-0.02	0.04	0.61	-0.09, 0.06
rs2070033	PDW	MA	1661	0.09	0.13	0.49	-0.16, 0.34

rs2227395	PDW	NHW	2276	-0.01	0.02	0.68	-0.04, 0.03
rs2227395	PDW	MA	1657	-0.01	0.03	0.80	-0.06, 0.05
rs2227395	PDW	NHB	1625	0.02	0.03	0.56	-0.04, 0.08
rs2227409	PDW	MA	1661	Not tested			
rs2227409	PDW	NHW	2284	Not tested			
rs2227409	PDW	NHB	1628	0.52	0.32	0.11	-0.11, 1.14
rs2227425	PDW	NHW	2285	Not tested			
rs2227425	PDW	NHB	1628	-0.08	0.07	0.22	-0.21, 0.05
rs2227425	PDW	MA	1661	1.14	0.54	0.03	0.08, 2.21
rs2227434	PDW	NHW	2282	-0.25	0.16	0.11	-0.56, 0.06
rs2227434	PDW	MA	1661	0.02	0.24	0.94	-0.46, 0.5
rs2227434	PDW	NHB	1625	0.36	0.39	0.36	-0.41, 1.13
rs4220	PDW	NHW	2272	0.00	0.02	0.83	-0.04, 0.03
rs4220	PDW	MA	1657	0.01	0.03	0.84	-0.05, 0.06
rs4220	PDW	NHB	1626	0.04	0.03	0.20	-0.02, 0.11
rs6050	PDW	NHW	2270	-0.03	0.02	0.04	-0.06, 0
rs6050	PDW	NHB	1620	-0.01	0.02	0.52	-0.05, 0.03
rs6050	PDW	MA	1649	0.04	0.02	0.09	-0.01, 0.08
rs6058	PDW	NHB	1627	-0.03	0.04	0.40	-0.11, 0.04
rs6058	PDW	NHW	2285	-0.02	0.18	0.93	-0.38, 0.34
rs6058	PDW	MA	1661	0.19	0.17	0.26	-0.14, 0.53
rs6063	PDW	MA	1661	0.05	0.18	0.80	-0.31, 0.4
rs6063	PDW	NHW	2285	0.06	0.09	0.54	-0.13, 0.24
rs6063	PDW	NHB	1628	0.61	0.28	0.03	0.06, 1.15
rs1049636	WBC	MA	1680	-0.04	0.07	0.58	-0.17, 0.1
rs1049636	WBC	NHW	2307	-0.02	0.07	0.82	-0.16, 0.13
rs1049636	WBC	NHB	1654	0.02	0.08	0.78	-0.14, 0.18
rs1800788	WBC	NHB	1659	-0.10	0.12	0.42	-0.34, 0.14
rs1800788	WBC	MA	1674	0.02	0.08	0.80	-0.14, 0.18
rs1800788	WBC	NHW	2304	0.11	0.09	0.22	-0.07, 0.29
rs1800791	WBC	NHB	1654	-0.14	0.11	0.20	-0.35, 0.08
rs1800791	WBC	MA	1677	0.01	0.08	0.93	-0.15, 0.16
rs1800791	WBC	NHW	2306	0.01	0.10	0.93	-0.18, 0.2
rs1800792	WBC	NHB	1659	-0.01	0.09	0.89	-0.2, 0.17
rs1800792	WBC	MA	1685	0.07	0.08	0.34	-0.08, 0.22
rs1800792	WBC	NHW	2310	0.09	0.07	0.19	-0.04, 0.23
rs2066860	WBC	NHB	1658	0.01	0.41	0.97	-0.8, 0.82
rs2066860	WBC	NHW	2310	0.04	0.18	0.81	-0.31, 0.4
rs2066860	WBC	MA	1685	0.07	0.30	0.81	-0.51, 0.65
rs2066861	WBC	NHB	1653	-0.03	0.08	0.71	-0.18, 0.12

rs2066861	WBC	MA	1679	0.02	0.08	0.82	-0.14, 0.18
rs2066861	WBC	NHW	2309	0.11	0.08	0.19	-0.05, 0.27
rs2066874	WBC	NHW	2313	-3.38	2.38	0.16	-8.06, 1.29
rs2066874	WBC	NHB	1656	0.06	0.22	0.77	-0.37, 0.49
rs2066874	WBC	MA	1686	2.19	2.03	0.28	-1.8, 6.18
rs2066879	WBC	NHB	1659	0.29	0.18	0.11	-0.07, 0.64
rs2066879	WBC	NHW	2315	0.33	1.69	0.84	-2.97, 3.63
rs2066879	WBC	MA	1685	0.48	0.68	0.48	-0.85, 1.81
rs2070006	WBC	NHB	1644	-0.05	0.07	0.53	-0.19, 0.1
rs2070006	WBC	NHW	2302	0.03	0.07	0.68	-0.11, 0.17
rs2070006	WBC	MA	1671	0.03	0.07	0.71	-0.11, 0.17
rs2070008	WBC	MA	1679	-0.04	0.11	0.72	-0.24, 0.17
rs2070008	WBC	NHB	1659	0.05	0.11	0.66	-0.17, 0.27
rs2070008	WBC	NHW	2303	0.07	0.10	0.52	-0.14, 0.27
rs2070009	WBC	NHW	2298	-0.05	0.07	0.52	-0.19, 0.09
rs2070009	WBC	MA	1675	-0.02	0.07	0.74	-0.16, 0.12
rs2070009	WBC	NHB	1650	0.06	0.07	0.42	-0.08, 0.2
rs2070011	WBC	NHB	1658	-0.11	0.08	0.20	-0.26, 0.05
rs2070011	WBC	MA	1678	0.02	0.07	0.75	-0.12, 0.16
rs2070011	WBC	NHW	2301	0.03	0.07	0.67	-0.11, 0.17
rs2070017	WBC	NHW	2314	-0.66	1.07	0.54	-2.75, 1.44
rs2070017	WBC	NHB	1655	0.10	0.12	0.40	-0.13, 0.33
rs2070017	WBC	MA	1686	0.40	0.43	0.35	-0.44, 1.24
rs2070022	WBC	NHB	1653	0.02	0.11	0.85	-0.19, 0.23
rs2070022	WBC	NHW	2309	0.05	0.09	0.58	-0.13, 0.23
rs2070022	WBC	MA	1671	0.06	0.07	0.45	-0.09, 0.2
rs2070031	WBC	NHW	2315	Not tested			
rs2070031	WBC	MA	1686	-0.59	2.04	0.77	-4.6, 3.41
rs2070031	WBC	NHB	1659	-0.28	0.42	0.51	-1.1, 0.55
rs2070033	WBC	NHB	1658	-0.12	0.14	0.40	-0.39, 0.15
rs2070033	WBC	NHW	2313	-0.09	0.90	0.92	-1.86, 1.68
rs2070033	WBC	MA	1686	-0.04	0.48	0.94	-0.98, 0.91
rs2227395	WBC	MA	1682	-0.23	0.10	0.02	-0.43, -0.04
rs2227395	WBC	NHW	2306	0.01	0.09	0.87	-0.16, 0.19
rs2227395	WBC	NHB	1656	0.10	0.11	0.37	-0.12, 0.32
rs2227409	WBC	MA	1686	Not tested			
rs2227409	WBC	NHW	2314	Not tested			
rs2227409	WBC	NHB	1659	-1.27	1.18	0.28	-3.6, 1.05
rs2227425	WBC	NHW	2315	Not tested			
rs2227425	WBC	NHB	1659	-0.05	0.24	0.84	-0.53, 0.43

rs2227425	WBC	MA	1686	0.85	2.04	0.68	-3.14, 4.84
rs2227434	WBC	NHW	2312	-0.25	0.80	0.75	-1.81, 1.31
rs2227434	WBC	NHB	1656	0.23	1.45	0.87	-2.61, 3.08
rs2227434	WBC	MA	1686	0.39	0.91	0.67	-1.4, 2.17
rs4220	WBC	MA	1682	-0.28	0.10	0.01	-0.48, -0.08
rs4220	WBC	NHW	2302	-0.03	0.09	0.73	-0.22, 0.15
rs4220	WBC	NHB	1657	0.07	0.13	0.58	-0.18, 0.31
rs6050	WBC	NHW	2300	-0.08	0.08	0.32	-0.24, 0.08
rs6050	WBC	MA	1673	-0.01	0.08	0.88	-0.17, 0.15
rs6050	WBC	NHB	1651	0.00	0.07	0.98	-0.15, 0.14
rs6058	WBC	MA	1686	-0.66	0.65	0.31	-1.92, 0.61
rs6058	WBC	NHB	1658	-0.05	0.14	0.75	-0.32, 0.23
rs6058	WBC	NHW	2315	0.54	0.97	0.58	-1.37, 2.45
rs6063	WBC	MA	1686	0.25	0.68	0.72	-1.09, 1.58
rs6063	WBC	NHW	2315	0.43	0.50	0.39	-0.55, 1.41
rs6063	WBC	NHB	1659	0.64	1.03	0.53	-1.37, 2.65

Appendix Q. Candidate SNPs list for replication of GWAS-identified variants in combined NHANES dataset. Below is a list of SNPs proposed for genotyping for section B in Chapter III. Candidate SNPs had to be either previously identified by a genome-wide association study or well-characterized in the literature by candidate gene studies. SNPs also had to be associated with fibrinogen or hematological related traits. Abbreviations: MPV= mean platelet count, FB = fibrinogen, MCH = mean corpuscular hemoglobin concentration, RBC = red blood cell count, MPC = mean platelet count, PC = platelet count, WBC = white blood cell count, AfAms= African Americans, GWAS = genome-wide association study, CG = candidate gene study, NR= not reported, CAF= Coded Allele Frequency.

SNP	Gene	Phenotype	Population	CAF	p-value	Study design	PubMed ID
rs7961894	*WDR66*	MPV	Europeans	0.12	7.24x10-48	GWAS	19110211
rs12485738	*ARHGEF3*	MPV	Europeans	0.36	3.81x10-27	GWAS	19110211
rs2138852	*TAOK1*	MPV	Europeans	0.49	7.19x10-28	GWAS	19110211
rs2070016	*FGA*	FB	Europeans	NR	1.00E-03	CG	16706972
rs1800790	*FGB*	FB	Europeans	NR	<0.001	CG	16706972
rs2227421	*FGB*	FB	Europeans	NR	4.00E-03	CG	16706972
rs2070017	*FGA*	FB	AfAms	NR	3.00E-03	CG	16706972
rs1800792	*FGB*	FB	AfAms	NR	2.00E-02	CG	16706972
rs2066874	*FGG*	FB	AfAms	NR	6.00E-03	CG	16706972
rs1800789	*FGB*	FB	Europeans	NR	1.75E-30	GWAS	20031576
rs2522056	*IRF1*	FB	Europeans	NR	1.30E-15	GWAS	20031576
rs511154	*PCCB*	FB	Europeans	NR	5.99E-10	GWAS	20031576
rs1539019	near *NLRP3*	FB	Europeans	NR	1.00E-08	GWAS	20031576
rs342293	Intergenic	MPV	Europeans	0.46	1.08E-24	GWAS	19221038
rs5756506	*TMPRSS6*	MCH	Europeans	NR	1.2E-6	GWAS	19820697
rs11970772	*BYSL/CCND3*	RBC	Europeans	NR	4.7E-7	GWAS	19820697
rs1800562	*HFW*	RBC	Europeans	NR	5.9E-11	GWAS	19820697
rs9609565	*FBX07*	RBC	Europeans	NR	8.2E-7	GWAS	19820697
rs9402686	*HBS1L-MYB*	RBC	Europeans	NR	9.1E-17	GWAS	19820697
rs7385804	*TFR2*	RBC	Europeans	NR	4.7E-6	GWAS	19820697
rs10914144	*DNM3*	MPC	Europeans	NR	2.9E-7	GWAS	19820697
rs11071720	*TPM1*	MPC	Europeans	NR	6.5E-7	GWAS	19820697
rs11602954	*BET1L*	MPC	Europeans	NR	1.9E-6	GWAS	19820697
rs12485738	*ARHGEF3*	MPC	Europeans	NR	1.5E-8	GWAS	19820697
rs1668873	*TMCC2*	MPC	Europeans	NR	3.3E-10	GWAS	19820697
rs2138852	*TAOK1*	MPC	Europeans	NR	2.5E-9	GWAS	19820697
rs2393967	*JMJD1C*	MPC	Europeans	NR	2.3E-8	GWAS	19820697
rs342293	*PIK3CG*	MPC	Europeans	NR	6.8E-13	GWAS	19820697
rs6136489	*SIRPA*	MPC	Europeans	NR	1.3E-6	GWAS	19820697
rs674316	*EHD3*	MPC	Europeans	NR	7.4E-8	GWAS	19820697
rs7961894	*WDR66*	MPC	Europeans	NR	8.2E-19	GWAS	19820697

rs893001	*CD226*	MPC	Europeans	NR	8.3E-8	GWAS	19820697
rs11065987	*ATXN2*	PC	Europeans	NR	8.3E-9	GWAS	19820697
rs11066301	*PTPN11*	PC	Europeans	NR	2.3E-9	GWAS	19820697
rs210135	*BAK1*	PC	Europeans	0.12	2.6E-7	GWAS	19278955
rs385893	*AK3*	PC	Europeans	NR	5.6E-7	GWAS	19820697
rs17609240	*GSDMA*	WBC	Europeans	NR	1.2E-6	GWAS	19820697

Appendix R. Minor allele frequency comparison for *SCN5A* SNPs in NHANES III populations and Jackson Heart Study To compare allele frequencies, we calculated Fst for all *SCN5A* SNPs, comparing each NHANES III sub-population studied and African Americans to the Jackson Heart Study. Abbreviations: minor allele frequency (MAF),

SNP	Alleles	Minor Allele	MAF Jackson Heart Study (n = 3054)	MAF non-Hispanic whites (n~1150)	MAF Mexican Americans (n~620)	JHS vs. NHW Fst	JHS vs. MA Fst
rs7374138	C/G	G	0.23	0.16	0.19	0.045	0.050
rs7637849	G/A	A	0.21	<0.01	0.03	0.019	0.004
rs11129796	C/T	T	0.15	0.05	0.04	0.051	0.037
rs7629265	C/T	T	0.08	0.003	0.01	0.015	0.125
rs6768664	A/C	C	0.36	0.46	0.45	0.061	0.053

REFERENCES

Centers for Disease Control and Prevention Third National Health and Nutrition Examination Survey, 1988-94, Plan and Operations Procedures Manual. 1996.

(2003). The International HapMap Project. Nature *426*, 789-796.

Centers for Disease Control and Prevention Plan and Operation of the Third National Health and Nutrition Examination Survey, 1988-94. 2004. Bethesda, MD.

Heart Disease: Symptoms, Diagnosis, and Treatment. NIH Medline Plus 4[1], 25-27. 12-1-2009a. Greensboro, NC, Vitality Communications.
Ref Type: Magazine Article

World Health Statistics 2009. 2009b. Geneva, Switzerland, World Health Organization.

Causes of Death 2008: World Health Organization. 2011. Geneva, Switzerland, World Health Organization.

Abrahamsen,B., Zhao,J., Asante,C.O., Cendan,C.M., Marsh,S., Martinez-Barbera,J.P., Nassar,M.A., Dickenson,A.H., and Wood,J.N. (2008). The cell and molecular basis of mechanical, cold, and inflammatory pain. Science *321*, 702-705.

Abriel,H. and Kass,R.S. (2005). Regulation of the voltage-gated cardiac sodium channel Nav1.5 by interacting proteins. Trends Cardiovasc. Med *15*, 35-40.

Adeyemo,A., Gerry,N., Chen,G., Herbert,A., Doumatey,A., Huang,H., Zhou,J., Lashley,K., Chen,Y., Christman,M., and Rotimi,C. (2009). A genome-wide association study of hypertension and blood pressure in African Americans. PLoS. Genet. *5*, e1000564.

Aehlert,B. (2006). Basic Electrophysiology. In ECGs Made Easy, L.Hoenycutt and K.Tomber , eds. (St. Louis, MO: Mosby elsevier), pp. 28-79.

Akopian,A.N., Souslova,V., England,S., Okuse,K., Ogata,N., Ure,J., Smith,A., Kerr,B.J., McMahon,S.B., Boyce,S., Hill,R., Stanfa,L.C., Dickenson,A.H., and Wood,J.N. (1999). The tetrodotoxin-resistant sodium channel SNS has a specialized function in pain pathways. Nat. Neurosci. *2*, 541-548.

Akylbekova,E.L., Crow,R.S., Johnson,W.D., Buxbaum,S.G., Njemanze,S., Fox,E., Sarpong,D.F., Taylor,H.A., and Newton-Cheh,C. (2009). Clinical correlates and

heritability of QT interval duration in blacks: the Jackson Heart Study. Circ. Arrhythm. Electrophysiol. *2*, 427-432.

Arking,D.E. and Chakravarti,A. (2009). Understanding cardiovascular disease through the lens of genome-wide association studies. Trends Genet. *25*, 387-394.

Arking,D.E., Khera,A., Xing,C., Kao,W.H., Post,W., Boerwinkle,E., and Chakravarti,A. (2009). Multiple independent genetic factors at NOS1AP modulate the QT interval in a multi-ethnic population. PLoS. One. *4*, e4333.

Arking,D.E., Pfeufer,A., Post,W., Kao,W.H., Newton-Cheh,C., Ikeda,M., West,K., Kashuk,C., Akyol,M., Perz,S., Jalilzadeh,S., Illig,T., Gieger,C., Guo,C.Y., Larson,M.G., Wichmann,H.E., Marban,E., O'Donnell,C.J., Hirschhorn,J.N., Kaab,S., Spooner,P.M., Meitinger,T., and Chakravarti,A. (2006). A common genetic variant in the NOS1 regulator NOS1AP modulates cardiac repolarization. Nat. Genet. *38*, 644-651.

Barasch,E., Benderly,M., Graff,E., Behar,S., Reicher-Reiss,H., Caspi,A., Pelled,B., Reisin,L., Roguin,N., and Goldbourt,U. (1995). Plasma fibrinogen levels and their correlates in 6457 coronary heart disease patients. The Bezafibrate Infarction Prevention (BIP) Study. J. Clin. Epidemiol. *48*, 757-765.

Baroudi,G., Acharfi,S., Larouche,C., and Chahine,M. (2002). Expression and intracellular localization of an SCN5A double mutant R1232W/T1620M implicated in Brugada syndrome. Circ. Res. *90*, E11-E16.

Baroudi,G., Pouliot,V., Denjoy,I., Guicheney,P., Shrier,A., and Chahine,M. (2001). Novel mechanism for Brugada syndrome: defective surface localization of an SCN5A mutant (R1432G). Circ. Res. *88*, E78-E83.

Barrett,J.C., Fry,B., Maller,J., and Daly,M.J. (2005). Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. *21*, 263-265.

Benson,D.W., Macrae,C.A., Vesely,M.R., Walsh,E.P., Seidman,J.G., Seidman,C.E., and Satler,C.A. (1996). Missense mutation in the pore region of HERG causes familial long QT syndrome. Circulation *93*, 1791-1795.

Best,L.G., North,K.E., Li,X., Palmieri,V., Umans,J.G., MacCluer,J., Laston,S., Haack,K., Goring,H., Diego,V.P., Almasy,L., Lee,E.T., Tracy,R.P., and Cole,S. (2008). Linkage study of fibrinogen levels: the Strong Heart Family Study. BMC. Med. Genet. *9*, 77.

Bezzina,C., Veldkamp,M.W., van den Berg,M.P., Postma,A.V., Rook,M.B., Viersma,J.W., van Langen,I.M., Tan-Sindhunata,G., Bink-Boelkens,M.T., Der Hout,A.H., Mannens,M.M., and Wilde,A.A. (1999). A single Na(+) channel mutation causing both long-QT and Brugada syndromes. Circ. Res. *85*, 1206-1213.

Bezzina,C.R., Shimizu,W., Yang,P., Koopmann,T.T., Tanck,M.W., Miyamoto,Y., Kamakura,S., Roden,D.M., and Wilde,A.A. (2006). Common sodium channel promoter

haplotype in Asian subjects underlies variability in cardiac conduction. Circulation. *113*, 338-344.

Bhagavatula,M.R., Fan,C., Shen,G.Q., Cassano,J., Plow,E.F., Topol,E.J., and Wang,Q. (2004). Transcription factor MEF2A mutations in patients with coronary artery disease. Hum. Mol. Genet. *13*, 3181-3188.

Blomback,B., Hessel,B., Hogg,D., and Therkildsen,L. (1978). A two-step fibrinogen--fibrin transition in blood coagulation. Nature *275*, 501-505.

Boekholdt,S.M., Bijsterveld,N.R., Moons,A.H., Levi,M., Buller,H.R., and Peters,R.J. (2001). Genetic variation in coagulation and fibrinolytic proteins and their relation with acute myocardial infarction: a systematic review. Circulation *104*, 3063-3068.

Boos,C.J. and Lip,G.Y. (2006). Platelet activation and cardiovascular outcomes in acute coronary syndromes. J Thromb Haemost. *4*, 2542-2543.

Burke,A., Creighton,W., Mont,E., Li,L., Hogan,S., Kutys,R., Fowler,D., and Virmani,R. (2005). Role of SCN5A Y1102 polymorphism in sudden cardiac death in blacks. Circulation *112*, 798-802.

Burt,V.L., Whelton,P., and Roccella,E.J. (1995). Prevalence of hypertension in the US adult population: Results from the Third National Health and Nutrition Examination Survey, 1988-91. Hypertension *25*, 305-313.

Bush,W.S., Crawford,D.C., Alexander,C., George,A.L., Jr., Roden,D.M., and Ritchie,M.D. (2009). Genetic variation in the rhythmonome: ethnic variation and haplotype structure in candidate genes for arrhythmias. Pharmacogenomics *10*, 1043-1053.

Campbell-Jenkins,B.W., Addison,C.C., Young,L., Anugu,P., Wilson,G., and Sarpong,D. (2009). Development of the Jackson Heart Study Coordinating Center. Int. J Environ. Res. Public Health *6*, 1597-1608.

Carlson,C.S., Eberle,M.A., Rieder,M.J., Yi,Q., Kruglyak,L., and Nickerson,D.A. (2004). Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. Am. J Hum. Genet. *74*, 106-120.

Carpenter,M.A., Crow,R., Steffes,M., Rock,W., Heilbraun,J., Evans,G., Skelton,T., Jensen,R., and Sarpong,D. (2004). Laboratory, reading center, and coordinating center data management methods in the Jackson Heart Study. Am. J Med. Sci. *328*, 131-144.

Chambers,J.C., Zhao,J., Terracciano,C.M., Bezzina,C.R., Zhang,W., Kaba,R., Navaratnarajah,M., Lotlikar,A., Sehmi,J.S., Kooner,M.K., Deng,G., Siedlecka,U., Parasramka,S., El-Hamamsy,I., Wass,M.N., Dekker,L.R., de Jong,J.S., Sternberg,M.J., McKenna,W., Severs,N.J., de,S.R., Wilde,A.A., Anand,P., Yacoub,M., Scott,J., Elliott,P.,

Wood,J.N., and Kooner,J.S. (2010). Genetic variation in SCN10A influences cardiac conduction. Nat. Genet. *42*, 149-152.

Chang,M.H., Lindegren,M.L., Butler,M.A., Chanock,S.J., Dowling,N.F., Gallagher,M., Moonesinghe,R., Moore,C.A., Ned,R.M., Reichler,M.R., Sanders,C.L., Welch,R., Yesupriya,A., and Khoury,M.J. (2009). Prevalence in the United States of selected candidate gene variants: Third National Health and Nutrition Examination Survey, 1991-1994. Am. J Epidemiol. *169*, 54-66.

Chen,Q.Y., Kirsch,G.E., Zhang,D.M., Brugada,R., Brugada,J., Brugada,P., Potenza,D., Moya,A., Borggrefe,M., Breithardt,G., Ortizlopez,R., Wang,Z., Antzelevitch,C., Obrien,R.E., Schulzebahr,E., Keating,M.T., Towbin,J.A., and Wang (1998). Genetic basis and molecular mechanism for idiopathic ventricular fibrillation. Nature *392*, 293-296.

Cheng,C.Y., Kao,W.H., Patterson,N., Tandon,A., Haiman,C.A., Harris,T.B., Xing,C., John,E.M., Ambrosone,C.B., Brancati,F.L., Coresh,J., Press,M.F., Parekh,R.S., Klag,M.J., Meoni,L.A., Hsueh,W.C., Fejerman,L., Pawlikowska,L., Freedman,M.L., Jandorf,L.H., Bandera,E.V., Ciupak,G.L., Nalls,M.A., Akylbekova,E.L., Orwoll,E.S., Leak,T.S., Miljkovic,I., Li,R., Ursin,G., Bernstein,L., Ardlie,K., Taylor,H.A., Boerwinckle,E., Zmuda,J.M., Henderson,B.E., Wilson,J.G., and Reich,D. (2009). Admixture mapping of 15,280 African Americans identifies obesity susceptibility loci on chromosomes 5 and X. PLoS. Genet. *5*, e1000490.

Collet,J.P., Soria,J., Mirshahi,M., Hirsch,M., Dagonnet,F.B., Caen,J., and Soria,C. (1993). Dusart syndrome: a new concept of the relationship between fibrin clot architecture and fibrin clot degradability: hypofibrinolysis related to an abnormal clot structure. Blood *82*, 2462-2469.

Cook,D.G., Cappuccio,F.P., Atkinson,R.W., Wicks,P.D., Chitolie,A., Nakandakare,E.R., Sagnella,G.A., and Humphries,S.E. (2001). Ethnic differences in fibrinogen levels: the role of environmental factors and the beta-fibrinogen gene. Am. J. Epidemiol. *153*, 799-806.

Crawford,D.C., Akey,D.T., and Nickerson,D.A. (2005). The patterns of natural variation in human genes. Annu. Rev. Genomics Hum. Genet. *6*, 287-312.

Crawford,D.C., Sanders,C.L., Qin,X., Smith,J.D., Shephard,C., Wong,M., Witrak,L., Rieder,M.J., and Nickerson,D.A. (2006). Genetic variation is associated with C-reactive protein levels in the Third National Health and Nutrition Examination Survey. Circulation *114*, 2458-2465.

Crook,E.D. and Taylor,H. (2003). Traditional and nontraditional risk factors for cardiovascular and renal disease in African Americans (Part 2): a project of the Jackson Heart Study investigators. Am. J Med. Sci. *325*, 305-306.

Crosslin,D.R., McDavid,A., Weston,N., Nelson,S.C., Zheng,X., Hart,E., de,A.M., Kullo,I.J., McCarty,C.A., Doheny,K.F., Pugh,E., Kho,A., Hayes,M.G., Pretel,S., Saip,A., Ritchie,M.D., Crawford,D.C., Crane,P.K., Newton,K., Li,R., Mirel,D.B., Crenshaw,A., Larson,E.B., Carlson,C.S., and Jarvik,G.P. (2011). Genetic variants associated with the white blood cell count in 13,923 subjects in the eMERGE Network. Hum. Genet.(In Press)

Danesh,J. and Lewington,S. (1998). Plasma homocysteine and coronary heart disease: systematic review of published epidemiological studies. J Cardiovasc. Risk *5*, 229-232.

Daum, P., Estergreen, J., and Wener.M. Laboratory Procedure Manual:Rate of Clot Formation on the STA-Compact. 1-12. 2000.

Davi,G. and Patrono,C. (2007). Platelet activation and atherothrombosis. N. Engl. J Med. *357*, 2482-2494.

David Ginsburg (2005). Genetic Risk Factors for Arterial Thrombosis and Inflammation. Hematology 442-444.

Davis,R.L., Holohan,P.D., Shrimpton,A.E., Tatum,A.H., Daucher,J., Collins,G.H., Todd,R., Bradshaw,C., Kent,P., Feiglin,D., Rosenbaum,A., Yerby,M.S., Shaw,C.M., Lacbawan,F., and Lawrence,D.A. (1999a). Familial encephalopathy with neuroserpin inclusion bodies. Am. J Pathol. *155*, 1901-1913.

Davis,R.L., Shrimpton,A.E., Holohan,P.D., Bradshaw,C., Feiglin,D., Collins,G.H., Sonderegger,P., Kinter,J., Becker,L.M., Lacbawan,F., Krasnewich,D., Muenke,M., Lawrence,D.A., Yerby,M.S., Shaw,C.M., Gooptu,B., Elliott,P.R., Finch,J.T., Carrell,R.W., and Lomas,D.A. (1999b). Familial dementia caused by polymerization of mutant neuroserpin. Nature *401*, 376-379.

de Maat,M.P. (2001). Effects of diet, drugs, and genes on plasma fibrinogen levels. Ann. N. Y. Acad. Sci. *936*, 509-521.

de Maat,M.P. (2005). Genetic variation in the fibrinogen gene cluster. Thromb. Haemost. *93*, 401-402.

Dehghan,A., Yang,Q., Peters,A., Basu,S., Bis,J.C., Rudnicka,A.R., Kavousi,M., Chen,M.H., Baumert,J., Lowe,G.D., McKnight,B., Tang,W., de,M.M., Larson,M.G., Eyhermendy,S., McArdle,W.L., Lumley,T., Pankow,J.S., Hofman,A., Massaro,J.M., Rivadeneira,F., Kolz,M., Taylor,K.D., van Duijn,C.M., Kathiresan,S., Illig,T., Aulchenko,Y.S., Volcik,K.A., Johnson,A.D., Uitterlinden,A.G., Tofler,G.H., Gieger,C., Psaty,B.M., Couper,D.J., Boerwinkle,E., Koenig,W., O'Donnell,C.J., Witteman,J.C., Strachan,D.P., Smith,N.L., and Folsom,A.R. (2009). Association of novel genetic Loci with circulating fibrinogen levels: a genome-wide association study in 6 population-based cohorts. Circ. Cardiovasc. Genet. *2*, 125-133.

Dekker,L.R.C., Bezzina,C.R., Henriques,J.P.S., Tanck,M.W., Koch,K.T., Alings,M.W., Arnold,A.E.R., de Boer,M.J., Gorgels,A.P.M., Michels,H.R., Verkerk,A., Verheugt,F.W.A., Zijlstra,F., and Wilde,A.A.M. (2006). Familial Sudden Death Is an Important Risk Factor for Primary Ventricular Fibrillation: A Case-Control Study in Acute Myocardial Infarction Patients. Circulation *114*, 1140-1145.

Denny,J.C., Crawford,D.C., Ritchie,M.D., Bielinski,S.J., Basford,M.A., Bradford,Y., Chai,H.S., Bastarache,L., Zuvich,R., Peissig,P., Carrell,D., Ramirez,A.H., Pathak,J., Wilke,R.A., Rasmussen,L., Wang,X., Pacheco,J.A., Kho,A.N., Hayes,M.G., Weston,N., Matsumoto,M., Kopp,P.A., Newton,K.M., Jarvik,G.P., Li,R., Manolio,T.A., Kullo,I.J., Chute,C.G., Chisholm,R.L., Larson,E.B., McCarty,C.A., Masys,D.R., Roden,D.M., and de,A.M. (2011). Variants near FOXE1 are associated with hypothyroidism and other thyroid conditions: using electronic medical records for genome- and phenome-wide studies. Am. J. Hum. Genet. *89*, 529-542.

Denny,J.C., Ritchie,M.D., Basford,M.A., Pulley,J.M., Bastarache,L., Brown-Gentry,K., Wang,D., Masys,D.R., Roden,D.M., and Crawford,D.C. (2010a). PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. Bioinformatics. *26*, 1205-1210.

Denny,J.C., Ritchie,M.D., Crawford,D.C., Schildcrout,J.S., Ramirez,A.H., Pulley,J.M., Basford,M.A., Masys,D.R., Haines,J.L., and Roden,D.M. (2010b). Identification of genomic predictors of atrioventricular conduction: using electronic medical records as a tool for genome science. Circulation *122*, 2016-2021.

Deo,R.C., Reich,D., Tandon,A., Akylbekova,E., Patterson,N., Waliszewska,A., Kathiresan,S., Sarpong,D., Taylor,H.A., Jr., and Wilson,J.G. (2009). Genetic differences between the determinants of lipid profile phenotypes in African and European Americans: the Jackson Heart Study. PLoS. Genet. *5*, e1000342.

Distante,S. (2006). Genetic predisposition to iron overload: prevalence and phenotypic expression of hemochromatosis-associated HFE-C282Y gene mutation. Scand. J. Clin. Lab Invest *66*, 83-100.

Distante,S., Berg,J.P., Lande,K., Haug,E., and Bell,H. (2000). HFE gene mutation (C282Y) and phenotypic expression among a hospitalised population in a high prevalence area of haemochromatosis. Gut *47*, 575-579.

Duggan,D., Zheng,S.L., Knowlton,M., Benitez,D., Dimitrov,L., Wiklund,F., Robbins,C., Isaacs,S.D., Cheng,Y., Li,G., Sun,J., Chang,B.L., Marovich,L., Wiley,K.E., Balter,K., Stattin,P., Adami,H.O., Gielzak,M., Yan,G., Sauvageot,J., Liu,W., Kim,J.W., Bleecker,E.R., Meyers,D.A., Trock,B.J., Partin,A.W., Walsh,P.C., Isaacs,W.B., Gronberg,H., Xu,J., and Carpten,J.D. (2007). Two genome-wide association studies of aggressive prostate cancer implicate putative prostate tumor suppressor gene DAB2IP. J Natl. Cancer Inst. *99*, 1836-1844.

Dumitrescu,L., Ritchie,M.D., Brown-Gentry,K., Pulley,J.M., Basford,M., Denny,J.C., Oksenberg,J.R., Roden,D.M., Haines,J.L., and Crawford,D.C. (2010). Assessing the accuracy of observer-reported ancestry in a biorepository linked to electronic medical records. Genet. Med.

Ensrud,K. and Grimm,R.H., Jr. (1992). The white blood cell count and risk for coronary heart disease. Am. Heart J *124*, 207-213.

Evans,D.M., Frazer,I.H., and Martin,N.G. (1999). Genetic and environmental causes of variation in basal levels of blood cells. Twin. Res. *2*, 250-257.

Feinbloom,D. and Bauer,K.A. (2005). Assessment of hemostatic risk factors in predicting arterial thrombotic events. Arterioscler Thromb Vasc Biol *25*, 2043-2053.

Feinleib,M., Garrison,R.J., Fabsitz,R., Christian,J.C., Hrubec,Z., Borhani,N.O., Kannel,W.B., Rosenman,R., Schwartz,J.T., and Wagner,J.O. (1977). The NHLBI twin study of cardiovascular disease risk factors: methodology and summary of results. Am. J Epidemiol. *106*, 284-285.

Folsom,A.R., Aleksic,N., Ahn,C., Boerwinkle,E., and Wu,K.K. (2001). Beta-fibrinogen gene -455G/A polymorphism and coronary heart disease incidence: the Atherosclerosis Risk in Communities (ARIC) Study. Ann. Epidemiol. *11*, 166-170.

Ford,E.S., Li,C., Zhao,G., Pearson,W.S., and Capewell,S. (2009). Trends in the prevalence of low risk factor burden for cardiovascular disease among United States adults. Circulation *120*, 1181-1188.

Friedlander,Y., Siscovick,D.S., Weinmann,S., Austin,M.A., Psaty,B.M., Lemaitre,R.N., Arbogast,P., Raghunathan,T.E., and Cobb,L.A. (1998). Family history as a risk factor for primary cardiac arrest. Circulation *97*, 155-160.

Fries,D., Innerhofer,P., and Schobersberger,W. (2009). Time for changing coagulation management in trauma-related massive bleeding. Curr. Opin. Anaesthesiol. *22*, 267-274.

Gauderman, W. J. and Morrison, J. M. QUANTO 1.1: A computer program for power and sample size calculations for genetic-epidemiology studies. 2006.
Ref Type: Computer Program

Gellens,M.E., George,A.L., Jr., Chen,L.Q., Chahine,M., Horn,R., Barchi,R.L., and Kallen,R.G. (1992). Primary structure and functional expression of the human cardiac tetrodotoxin-insensitive voltage-dependent sodium channel. Proc. Natl. Acad. Sci. *89*, 554-558.

George,A.L., Jr. (2005). Inherited disorders of voltage-gated sodium channels. J Clin. Invest *115*, 1990-1999.

George,A.L., Jr., Iyer,G.S., Kleinfield,R., Kallen,R.G., and Barchi,R.L. (1993). Genomic organization of the human skeletal muscle sodium channel gene. Genomics *15*, 598-606.

George,A.L., Jr., Varkony,T.A., Drabkin,H.A., Han,J., Knops,J.F., Finley,W.H., Brown,G.B., Ward,D.C., and Haas,M. (1995). Assignment of the human heart tetrodotoxin-resistant voltage-gated Na+ channel alpha-subunit gene (SCN5A) to band 3p21. Cytogenet. Cell Genet. *68*, 67-70.

Go,A.S., Hylek,E.M., Phillips,K.A., Chang,Y., Henault,L.E., Selby,J.V., and Singer,D.E. (2001). Prevalence of diagnosed atrial fibrillation in adults: national implications for rhythm management and stroke prevention: the AnTicoagulation and Risk Factors in Atrial Fibrillation (ATRIA) Study. JAMA *285*, 2370-2375.

Gouas,L., Nicaud,V., Berthet,M., Forhan,A., Tiret,L., Balkau,B., and Guicheney,P. (2005b). Association of KCNQ1, KCNE1, KCNH2 and SCN5A polymorphisms with QTc interval length in a healthy population. Eur. J Hum. Genet. *13*, 1213-1222.

Gouas,L., Nicaud,V., Chaouch,S., Berthet,M., Forhan,A., Tichet,J., Tiret,L., Balkau,B., and Guicheney,P. (2007). Confirmation of associations between ion channel gene SNPs and QTc interval duration in healthy subjects. Eur. J Hum. Genet. *15*, 974-979.

Grady,B.J., Torstenson,E., Dudek,S.M., Giles,J., Sexton,D., and Ritchie,M.D. (2010). Finding unique filter sets in plato: a precursor to efficient interaction analysis in gwas data. Pac. Symp. Biocomput. 315-326.

Green,D., Ruth,K.J., Folsom,A.R., and Liu,K. (1994). Hemostatic factors in the Coronary Artery Risk Development in Young Adults (CARDIA) Study. Arterioscler. Thromb. *14*, 686-693.

Gretarsdottir,S., Baas,A.F., Thorleifsson,G., Holm,H., den,H.M., de Vries,J.P., Kranendonk,S.E., Zeebregts,C.J., van Sterkenburg,S.M., Geelkerken,R.H., van Rij,A.M., Williams,M.J., Boll,A.P., Kostic,J.P., Jonasdottir,A., Jonasdottir,A., Walters,G.B., Masson,G., Sulem,P., Saemundsdottir,J., Mouy,M., Magnusson,K.P., Tromp,G., Elmore,J.R., Sakalihasan,N., Limet,R., Defraigne,J.O., Ferrell,R.E., Ronkainen,A., Ruigrok,Y.M., Wijmenga,C., Grobbee,D.E., Shah,S.H., Granger,C.B., Quyyumi,A.A., Vaccarino,V., Patel,R.S., Zafari,A.M., Levey,A.I., Austin,H., Girelli,D., Pignatti,P.F., Olivieri,O., Martinelli,N., Malerba,G., Trabetti,E., Becker,L.C., Becker,D.M., Reilly,M.P., Rader,D.J., Mueller,T., Dieplinger,B., Haltmayer,M., Urbonavicius,S., Lindblad,B., Gottsater,A., Gaetani,E., Pola,R., Wells,P., Rodger,M., Forgie,M., Langlois,N., Corral,J., Vicente,V., Fontcuberta,J., Espana,F., Grarup,N., Jorgensen,T., Witte,D.R., Hansen,T., Pedersen,O., Aben,K.K., de,G.J., Holewijn,S., Folkersen,L., Franco-Cereceda,A., Eriksson,P., Collier,D.A., Stefansson,H., Steinthorsdottir,V., Rafnar,T., Valdimarsson,E.M., Magnadottir,H.B., Sveinbjornsdottir,S., Olafsson,I., Magnusson,M.K., Palmason,R., Haraldsdottir,V., Andersen,K., Onundarson,P.T., Thorgeirsson,G., Kiemeney,L.A., Powell,J.T., Carey,D.J., Kuivaniemi,H., Lindholt,J.S., Jones,G.T., Kong,A., Blankensteijn,J.D., Matthiasson,S.E., Thorsteinsdottir,U., and

Stefansson,K. (2010). Genome-wide association study identifies a sequence variant within the DAB2IP gene conferring susceptibility to abdominal aortic aneurysm. Nat. Genet. *42*, 692-697.

Grody,W.W. (2003). Molecular genetic risk screening. Annu. Rev. Med. *54*, 473-490.

Grody,W.W., Griffin,J.H., Taylor,A.K., Korf,B.R., and Heit,J.A. (2001). American College of Medical Genetics consensus statement on factor V Leiden mutation testing. Genet. Med. *3*, 139-148.

Grundy,S.M. (1980). Diet and coronary heart disease. Compr. Ther. *6*, 28-34.

Grundy,S.M., Pasternak,R., Greenland,P., Smith S Jr, and Fuster,V. (1999). Assessment of cardiovascular risk by use of multiple-risk-factor assessment equations: a statement for healthcare professionals from the American Heart Association and the American College of Cardiology. Circulation *100*, 1481-1492.

Gunter E., Lewis, B., and Koncikowski, S. Laboratory Procedures Used for the Third National Health and Nutrition Examination Survey (NHANES III), 1988-1994. 537-546. 1996.

Hadley,T.J. and Peiper,S.C. (1997). From malaria to chemokine receptor: the emerging physiologic role of the Duffy blood group antigen. Blood *89*, 3077-3091.

Hammer,M.F., Woerner,A.E., Mendez,F.L., Watkins,J.C., and Wall,J.D. (2011). Genetic evidence for archaic admixture in Africa. Proc. Natl. Acad. Sci. U. S. A *108*, 15123-15128.

Hamsten,A., Iselius,L., de,F.U., and Blomback,M. (1987). Genetic and cultural inheritance of plasma fibrinogen concentration. Lancet *2*, 988-991.

Hanson,B., Tuna,N., Bouchard,T., Heston,L., Eckert,E., Lykken,D., Segal,N., and Rich,S. (1989). Genetic factors in the electrocardiogram and heart rate of twins reared apart and together. Am. J Cardiol. *63*, 606-609.

Havlik,R.J., Garrison,R.J., Fabsitz,R., and Feinleib,M. (1980). Variability of heart rate, P-R, QRS and Q-T durations in twins. J Electrocardiol. *13*, 45-48.

Heinrich,J., Balleisen,L., Schulte,H., Assmann,G., and van de,L.J. (1994). Fibrinogen and factor VII in the prediction of coronary risk. Results from the PROCAM study in healthy men. Arterioscler. Thromb. *14*, 54-59.

Helgadottir,A., Manolescu,A., Helgason,A., Thorleifsson,G., Thorsteinsdottir,U., Gudbjartsson,D.F., Gretarsdottir,S., Magnusson,K.P., Gudmundsson,G., Hicks,A., Jonsson,T., Grant,S.F., Sainz,J., O'Brien,S.J., Sveinbjornsdottir,S., Valdimarsson,E.M., Matthiasson,S.E., Levey,A.I., Abramson,J.L., Reilly,M.P., Vaccarino,V., Wolfe,M.L.,

Gudnason,V., Quyyumi,A.A., Topol,E.J., Rader,D.J., Thorgeirsson,G., Gulcher,J.R., Hakonarson,H., Kong,A., and Stefansson,K. (2006). A variant of the gene encoding leukotriene A4 hydrolase confers ethnicity-specific risk of myocardial infarction. Nat. Genet. *38*, 68-74.

Hermans,J. and McDonagh,J. (1982). Fibrin: structure and interactions. Semin. Thromb. Hemost. *8*, 11-24.

Hindorff,L.A., Sethupathy,P., Junkins,H.A., Ramos,E.M., Mehta,J.P., Collins,F.S., and Manolio,T.A. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc. Natl. Acad. Sci. U. S. A *106*, 9362-9367.

Holm,H., Gudbjartsson,D.F., Arnar,D.O., Thorleifsson,G., Thorgeirsson,G., Stefansdottir,H., Gudjonsson,S.A., Jonasdottir,A., Mathiesen,E.B., Njolstad,I., Nyrnes,A., Wilsgaard,T., Hald,E.M., Hveem,K., Stoltenberg,C., Lochen,M.L., Kong,A., Thorsteinsdottir,U., and Stefansson,K. (2010). Several common variants modulate heart rate, PR interval and QRS duration. Nat. Genet. *42*, 117-122.

Hooper,W.C., Lally,C., Austin,H., Benson,J., Dilley,A., Wenger,N.K., Whitsett,C., Rawlins,P., and Evatt,B.L. (1999). The relationship between polymorphisms in the endothelial cell nitric oxide synthase gene and the platelet GPIIIa gene with myocardial infarction and venous thromboembolism in African Americans. Chest *116*, 880-886.

Hsieh,M.M., Everhart,J.E., Byrd-Holt,D.D., Tisdale,J.F., and Rodgers,G.P. (2007). Prevalence of neutropenia in the U.S. population: age, sex, smoking status, and ethnic differences. Ann. Intern. Med. *146*, 486-492.

Hunt,M.S., Katzmarzyk,P.T., Perusse,L., Rice,T., Rao,D.C., and Bouchard,C. (2002). Familial resemblance of 7-year changes in body mass and adiposity. Obes. Res. *10*, 507-517.

Hunt,S.C., Hasstedt,S.J., Kuida,H., Stults,B.M., Hopkins,P.N., and Williams,R.R. (1989). Genetic heritability and common environmental components of resting and stressed blood pressures, lipids, and body mass index in Utah pedigrees and twins. Am. J Epidemiol. *129*, 625-638.

Ioannidis,J.P., Ntzani,E.E., and Trikalinos,T.A. (2004). 'Racial' differences in genetic effects for complex diseases. Nat. Genet. *36*, 1312-1318.

Jeff,J., Brown-Gentry,K., and Crawford,D.C. (2011a). Replication and characterization of genetic variants in the fibriongen gene cluster with fibrinogen levels and hematological traits in the Third National Health and Nutrition Examination Survey. Thromb Haemost. (In Press)

Jeff,J.M., Brown-Gentry,K., Buxbaum,S.G., Sarpong,D.F., Taylor,H.A., George,A.L., Jr., Roden,D.M., and Crawford,D.C. (2011b). SCN5A Variation is Associated with Electrocardiographic Traits in the Jackson Heart Study. Circ. Cardiovasc. Genet.

Johnson,A.D., Yanek,L.R., Chen,M.H., Faraday,N., Larson,M.G., Tofler,G., Lin,S.J., Kraja,A.T., Province,M.A., Yang,Q., Becker,D.M., O'Donnell,C.J., and Becker,L.C. (2010). Genome-wide meta-analyses identifies seven loci associated with platelet aggregation in response to agonists. Nat. Genet. *42*, 608-613.

Jouven,X., Desnos,M., Guerot,C., and Ducimetiere,P. (1999). Predicting sudden death in the population: the Paris Prospective Study I. Circulation *99*, 1978-1983.

Kannel,W.B. and McGee,D.L. (1979). Diabetes and glucose tolerance as risk factors for cardiovascular disease: the Framingham study. Diabetes Care *2*, 120-126.

Kannel,W.B., Wolf,P.A., Castelli,W.P., and D'Agostino,R.B. (1987). Fibrinogen and risk of cardiovascular disease. The Framingham Study. JAMA *258*, 1183-1186.

Kao,W.H., Arking,D.E., Post,W., Rea,T.D., Sotoodehnia,N., Prineas,R.J., Bishe,B., Doan,B.Q., Boerwinkle,E., Psaty,B.M., Tomaselli,G.F., Coresh,J., Siscovick,D.S., Marban,E., Spooner,P.M., Burke,G.L., and Chakravarti,A. (2009). Genetic Variations in Nitric Oxide Synthase 1 Adaptor Protein Are Associated With Sudden Cardiac Death in US White Community-Based Populations. Circulation.

Keating,B.J., Tischfield,S., Murray,S.S., Bhangale,T., Price,T.S., Glessner,J.T., Galver,L., Barrett,J.C., Grant,S.F., Farlow,D.N., Chandrupatla,H.R., Hansen,M., Ajmal,S., Papanicolaou,G.J., Guo,Y., Li,M., Derohannessian,S., de Bakker,P.I., Bailey,S.D., Montpetit,A., Edmondson,A.C., Taylor,K., Gai,X., Wang,S.S., Fornage,M., Shaikh,T., Groop,L., Boehnke,M., Hall,A.S., Hattersley,A.T., Frackelton,E., Patterson,N., Chiang,C.W., Kim,C.E., Fabsitz,R.R., Ouwehand,W., Price,A.L., Munroe,P., Caulfield,M., Drake,T., Boerwinkle,E., Reich,D., Whitehead,A.S., Cappola,T.P., Samani,N.J., Lusis,A.J., Schadt,E., Wilson,J.G., Koenig,W., McCarthy,M.I., Kathiresan,S., Gabriel,S.B., Hakonarson,H., Anand,S.S., Reilly,M., Engert,J.C., Nickerson,D.A., Rader,D.J., Hirschhorn,J.N., and Fitzgerald,G.A. (2008). Concept, design and implementation of a cardiovascular gene-centric 50 k SNP array for large-scale genomic association studies. PLoS. One. *3*, e3583.

Koch,W., Hoppmann,P., Biele,J., Mueller,J.C., Schomig,A., and Kastrati,A. (2008). Fibrinogen genes and myocardial infarction: a haplotype analysis. Arterioscler Thromb Vasc Biol *28*, 758-763.

Kottke-Marchant K (2011). Genetic polymorphisms associated with venous and arterial thrombosis: an overview. Archives of Pathology and Laboratory Medcine *126*, 295-304.

Kupper,N.H., Willemsen,G., van den,B.M., de,B.D., Posthuma,D., Boomsma,D.I., and de Geus,E.J. (2004). Heritability of ambulatory heart rate variability. Circulation *110*, 2792-2796.

Lai,L.P., Deng,C.L., Moss,A.J., Kass,R.S., and Liang,C.S. (1994). Polymorphism of the gene encoding a human minimal potassium ion channel (minK). Gene *151*, 339-340.

Lane,D.A. and Grant,P.J. (2000). Role of hemostatic gene polymorphisms in venous and arterial thrombotic disease. Blood *95*, 1517-1532.

Law,A.J., Kleinman,J.E., Weinberger,D.R., and Weickert,C.S. (2007). Disease-associated intronic variants in the ErbB4 gene are related to altered ErbB4 splice-variant expression in the brain in schizophrenia. Hum. Mol. Genet. *16*, 129-141.

Lee,A.J., Lowe,G.D., Woodward,M., and Tunstall-Pedoe,H. (1993). Fibrinogen in relation to personal history of prevalent hypertension, diabetes, stroke, intermittent claudication, coronary heart disease, and family history: the Scottish Heart Health Study. Br. Heart J. *69*, 338-342.

Liao,Y. and Cooper,R.S. (1995). Continued adverse trends in coronary heart disease mortality among blacks, 1980-91. Public Health Rep. *110*, 572-579.

Lind,L., Sundstrom,J., and Lithell,H. (1997). Cardiovascular risk factors and electrographical characteristics and abnormalities in middle-aged males. J Intern. Med. *241*, 109-113.

Lloyd-Jones,D., Adams,R., Carnethon,M., De,S.G., Ferguson,T.B., Flegal,K., Ford,E., Furie,K., Go,A., Greenlund,K., Haase,N., Hailpern,S., Ho,M., Howard,V., Kissela,B., Kittner,S., Lackland,D., Lisabeth,L., Marelli,A., McDermott,M., Meigs,J., Mozaffarian,D., Nichol,G., O'Donnell,C., Roger,V., Rosamond,W., Sacco,R., Sorlie,P., Stafford,R., Steinberger,J., Thom,T., Wasserthiel-Smoller,S., Wong,N., Wylie-Rosett,J., and Hong,Y. (2009). Heart disease and stroke statistics--2009 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Circulation *119*, e21-181.

Lo,K.S., Wilson,J.G., Lange,L.A., Folsom,A.R., Galarneau,G., Ganesh,S.K., Grant,S.F., Keating,B.J., McCarroll,S.A., Mohler,E.R., III, O'Donnell,C.J., Palmas,W., Tang,W., Tracy,R.P., Reiner,A.P., and Lettre,G. (2011). Genetic association analysis highlights new loci that modulate hematological trait variation in Caucasians and African Americans. Hum. Genet. *129*, 307-317.

Lohmueller,K.E., Indap,A.R., Schmidt,S., Boyko,A.R., Hernandez,R.D., Hubisz,M.J., Sninsky,J.J., White,T.J., Sunyaev,S.R., Nielsen,R., Clark,A.G., and Bustamante,C.D. (2008). Proportionally more deleterious genetic variation in European than in African populations. Nature *451*, 994-997.

London,B., Michalec,M., Mehdi,H., Zhu,X., Kerchner,L., Sanyal,S., Viswanathan,P.C., Pfahnl,A.E., Shang,L.L., Madhusudanan,M., Baty,C.J., Lagana,S., Aleong,R., Gutmann,R., Ackerman,M.J., McNamara,D.M., Weiss,R., and Dudley,S.C., Jr. (2007). Mutation in glycerol-3-phosphate dehydrogenase 1 like gene (GPD1-L) decreases cardiac Na+ current and causes inherited arrhythmias. Circulation *116*, 2260-2268.

Lowe,G.D. (2008). Common risk factors for both arterial and venous thrombosis. Br. J Haematol. *140*, 488-495.

Mackie,I.J., Kitchen,S., Machin,S.J., and Lowe,G.D. (2003). Guidelines on fibriongen assays. Br. J Haematol. *121*, 396-404.

Mackman,N. (2008). Triggers, targets and treatments for thrombosis. Nature *451*, 914-918.

Makielski,J.C., Ye,B., Valdivia,C.R., Pagel,M.D., Pu,J., Tester,D.J., and Ackerman,M.J. (2003). A ubiquitous splice variant and a common polymorphism affect heterologous expression of recombinant human SCN5A heart sodium channels. Circ. Res. *93*, 821-828.

Manolio,T.A., Collins,F.S., Cox,N.J., Goldstein,D.B., Hindorff,L.A., Hunter,D.J., McCarthy,M.I., Ramos,E.M., Cardon,L.R., Chakravarti,A., Cho,J.H., Guttmacher,A.E., Kong,A., Kruglyak,L., Mardis,E., Rotimi,C.N., Slatkin,M., Valle,D., Whittemore,A.S., Boehnke,M., Clark,A.G., Eichler,E.E., Gibson,G., Haines,J.L., Mackay,T.F., McCarroll,S.A., and Visscher,P.M. (2009). Finding the missing heritability of complex diseases. Nature *461*, 747-753.

Margaglione,M. and Grandone,E. (2011). Population genetics of venous thromboembolism. A narrative review. Thromb Haemost. *105*, 221-231.

Marroni,F., Pfeufer,A., Aulchenko,Y.S., Franklin,C.S., Isaacs,A., Pichler,I., Wild,S.H., Oostra,B.A., Wright,A.F., Campbell,H., Witteman,J.C., Kaab,S., Hicks,A.A., Gyllensten,U., Rudan,I., Meitinger,T., Pattaro,C., van Duijn,C.M., Wilson,J.F., and Pramstaller,P.P. (2009). A genome-wide association scan of RR and QT interval duration in 3 European genetically isolated populations: the EUROSPAN project. Circ. Cardiovasc. Genet. *2*, 322-328.

Mason,K.D., Carpinelli,M.R., Fletcher,J.I., Collinge,J.E., Hilton,A.A., Ellis,S., Kelly,P.N., Ekert,P.G., Metcalf,D., Roberts,A.W., Huang,D.C., and Kile,B.T. (2007). Programmed anuclear cell death delimits platelet life span. Cell *128*, 1173-1186.

MATHERS,J.A., OSBORNE,R.H., and DEGEORGE,F.V. (1961). Studies of blood pressure, heart rate, and the electrocardiogram in adult twins. Am. Heart J *62*, 634-642.

Matise, T., Ambite, J. L., Buyske, S., Cole, S. A, Crawford, D. C., Haiman, C., Heiss, H., Kooperberg, C., Le Marchand, L., and Haines, J. L. (2011) The next PAGE in understanding complex traits: study design for analysis of Population Architecture using Genomics and Epidemiology. 174(7); 849-859

McCarty,C.A., Chisholm,R.L., Chute,C.G., Kullo,I.J., Jarvik,G.P., Larson,E.B., Li,R., Masys,D.R., Ritchie,M.D., Roden,D.M., Struewing,J.P., and Wolf,W.A. (2011). The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. BMC. Med. Genomics *4*, 13.

McGlennen,R.C. and Key,N.S. (2002). Clinical and laboratory management of the prothrombin G20210A mutation. Arch. Pathol. Lab Med. *126*, 1319-1325.

Meade,T.W., Mellows,S., Brozovic,M., Miller,G.J., Chakrabarti,R.R., North,W.R., Haines,A.P., Stirling,Y., Imeson,J.D., and Thompson,S.G. (1986). Haemostatic function and ischaemic heart disease: principal results of the Northwick Park Heart Study. Lancet *2*, 533-537.

Meisinger,C., Prokisch,H., Gieger,C., Soranzo,N., Mehta,D., Rosskopf,D., Lichtner,P., Klopp,N., Stephens,J., Watkins,N.A., Deloukas,P., Greinacher,A., Koenig,W., Nauck,M., Rimmbach,C., Volzke,H., Peters,A., Illig,T., Ouwehand,W.H., Meitinger,T., Wichmann,H.E., and Doring,A. (2009). A genome-wide association study identifies three loci associated with mean platelet volume. Am. J Hum. Genet. *84*, 66-71.

Mendis, S. Report for the Pathobiological Determinants of Atherosclerosis in Youth (PBDAY) Research Group:Atherosclerosis in children and young adults: An overview of the World Health Organization and International Society and Federation of Cardiology Study on Pathobiological Determinants of Atherosclerosis in Youth Study (1985–1995). 1, 3-15. 2005. Prevention and Control

Mendis, S., Puska, P., Norrving, B., Alwan, A., Armstrong, T., Bettcher, D., Boerma, T., Branca, F., C.Y.Ho, J., Mathers, C., Martinez, R., Poznyak, V., Roglic, G., Riley, L., d'Espaignet, E., Stevens, G., Taubert, K., and Xuereb, G. Global Atlas on cardiovascular disease and prevention and control: 2011 World Health Organization Report. Mendis, S. 2011. Geneva, Switzerland, World Health Organization.

Mensah,G.A., Mokdad,A.H., Ford,E.S., Greenlund,K.J., and Croft,J.B. (2005). State of disparities in cardiovascular health in the United States. Circulation *111*, 1233-1241.

Micheal A Chizner (2004). Electrocardiogram. In Clinical Cardiology, (Miami: MedMaster), pp. 79-99.

Middeldorp,S. (2005). Oral contraceptives and the risk of venous thromboembolism. Gend. Med. *2 Suppl A*, S3-S9.

Miura,K., Nakagawa,H., Ueshima,H., Okayama,A., Saitoh,S., Curb,J.D., Rodriguez,B.L., Sakata,K., Okuda,N., Yoshita,K., and Stamler,J. (2006). Dietary factors related to higher plasma fibrinogen levels of Japanese-americans in hawaii compared with Japanese in Japan. Arterioscler. Thromb. Vasc. Biol. *26*, 1674-1679.

Mjos,O.D., Thelle,D.S., Forde,O.H., and Vik-Mo,H. (1977). Family study of high density lipoprotein cholesterol and the relation to age and sex. The Tromso Heart Study. Acta Med. Scand. *201*, 323-329.

Moller,P., Heiberg,A., and Berg,K. (1982). The atrioventricular conduction time - a heritable trait? III. Twin studies. Clin. Genet. *21*, 181-183.

Mosesson,M.W. (1998). Fibrinogen structure and fibrin clot assembly. Semin. Thromb. Hemost. *24*, 169-174.

Mueller,J.C., Lohmussaar,E., Magi,R., Remm,M., Bettecken,T., Lichtner,P., Biskup,S., Illig,T., Pfeufer,A., Luedemann,J., Schreiber,S., Pramstaller,P., Pichler,I., Romeo,G., Gaddi,A., Testa,A., Wichmann,H.E., Metspalu,A., and Meitinger,T. (2005). Linkage disequilibrium patterns and tagSNP transferability among European populations. Am. J Hum. Genet. *76*, 387-398.

Nalls,M.A., Couper,D.J., Tanaka,T., van Rooij,F.J., Chen,M.H., Smith,A.V., Toniolo,D., Zakai,N.A., Yang,Q., Greinacher,A., Wood,A.R., Garcia,M., Gasparini,P., Liu,Y., Lumley,T., Folsom,A.R., Reiner,A.P., Gieger,C., Lagou,V., Felix,J.F., Volzke,H., Gouskova,N.A., Biffi,A., Doring,A., Volker,U., Chong,S., Wiggins,K.L., Rendon,A., Dehghan,A., Moore,M., Taylor,K., Wilson,J.G., Lettre,G., Hofman,A., Bis,J.C., Pirastu,N., Fox,C.S., Meisinger,C., Sambrook,J., Arepalli,S., Nauck,M., Prokisch,H., Stephens,J., Glazer,N.L., Cupples,L.A., Okada,Y., Takahashi,A., Kamatani,Y., Matsuda,K., Tsunoda,T., Tanaka,T., Kubo,M., Nakamura,Y., Yamamoto,K., Kamatani,N., Stumvoll,M., Tonjes,A., Prokopenko,I., Illig,T., Patel,K.V., Garner,S.F., Kuhnel,B., Mangino,M., Oostra,B.A., Thein,S.L., Coresh,J., Wichmann,H.E., Menzel,S., Lin,J., Pistis,G., Uitterlinden,A.G., Spector,T.D., Teumer,A., Eiriksdottir,G., Gudnason,V., Bandinelli,S., Frayling,T.M., Chakravarti,A., van Duijn,C.M., Melzer,D., Ouwehand,W.H., Levy,D., Boerwinkle,E., Singleton,A.B., Hernandez,D.G., Longo,D.L., Soranzo,N., Witteman,J.C., Psaty,B.M., Ferrucci,L., Harris,T.B., O'Donnell,C.J., and Ganesh,S.K. (2011). Multiple loci are associated with white blood cell phenotypes. PLoS. Genet. *7*, e1002113.

Nelson,T.L., Vogler,G.P., Pedersen,N.L., Hong,Y., and Miles,T.P. (2000). Genetic and environmental influences on body fat distribution, fasting insulin levels and CVD: are the influences shared? Twin. Res. *3*, 43-50.

Newton-Cheh,C., Eijgelsheim,M., Rice,K.M., de Bakker,P.I., Yin,X., Estrada,K., Bis,J.C., Marciante,K., Rivadeneira,F., Noseworthy,P.A., Sotoodehnia,N., Smith,N.L., Rotter,J.I., Kors,J.A., Witteman,J.C., Hofman,A., Heckbert,S.R., O'Donnell,C.J., Uitterlinden,A.G., Psaty,B.M., Lumley,T., Larson,M.G., and Stricker,B.H. (2009). Common variants at ten loci influence QT interval duration in the QTGEN Study. Nat. Genet. *41*, 399-406.

Newton-Cheh,C., Guo,C.Y., Larson,M.G., Musone,S.L., Surti,A., Camargo,A.L., Drake,J.A., Benjamin,E.J., Levy,D., D'Agostino,R.B., Sr., Hirschhorn,J.N., and O'Donnell,C.J. (2007a). Common genetic variation in KCNH2 is associated with QT interval duration: the Framingham Heart Study. Circulation *116*, 1128-1136.

Newton-Cheh,C., Guo,C.Y., Wang,T.J., O'Donnell,C.J., Levy,D., and Larson,M.G. (2007b). Genome-wide association study of electrocardiographic and heart rate variability traits: the Framingham Heart Study. BMC. Med. Genet. *8 Suppl 1*, S7.

Newton-Cheh,C. and Shah,R. (2007). Genetic determinants of QT interval variation and sudden cardiac death. Curr. Opin. Genet. Dev. *17*, 213-221.

Norris,J.M., Langefeld,C.D., Scherzinger,A.L., Rich,S.S., Bookman,E., Beck,S.R., Saad,M.F., Haffner,S.M., Bergman,R.N., Bowden,D.W., and Wagenknecht,L.E. (2005). Quantitative trait loci for abdominal fat and BMI in Hispanic-Americans and African-Americans: the IRAS Family study. Int. J Obes. (Lond) *29*, 67-77.

Noseworthy,P.A. and Newton-Cheh,C. (2008). Genetic determinants of sudden cardiac death. Circulation *118*, 1854-1863.

Patterson,N., Hattangadi,N., Lane,B., Lohmueller,K.E., Hafler,D.A., Oksenberg,J.R., Hauser,S.L., Smith,M.W., O'Brien,S.J., Altshuler,D., Daly,M.J., and Reich,D. (2004a). Methods for high-density admixture mapping of disease genes. Am. J Hum. Genet. *74*, 979-1000.

Patterson,N., Hattangadi,N., Lane,B., Lohmueller,K.E., Hafler,D.A., Oksenberg,J.R., Hauser,S.L., Smith,M.W., O'Brien,S.J., Altshuler,D., Daly,M.J., and Reich,D. (2004b). Methods for high-density admixture mapping of disease genes. Am. J Hum. Genet. *74*, 979-1000.

Pearson,T.A., Mensah,G.A., Alexander,R.W., Anderson,J.L., Cannon,R.O., III, Criqui,M., Fadl,Y.Y., Fortmann,S.P., Hong,Y., Myers,G.L., Rifai,N., Smith,S.C., Jr., Taubert,K., Tracy,R.P., and Vinicor,F. (2003). Markers of inflammation and cardiovascular disease: application to clinical and public health practice: A statement for healthcare professionals from the Centers for Disease Control and Prevention and the American Heart Association. Circulation *107*, 499-511.

Pendergrass,S.A., Brown-Gentry,K., Dudek,S.M., Torstenson,E.S., Ambite,J.L., Avery,C.L., Buyske,S., Cai,C., Fesinmeyer,M.D., Haiman,C., Heiss,G., Hindorff,L.A., Hsu,C.N., Jackson,R.D., Kooperberg,C., Le,M.L., Lin,Y., Matise,T.C., Moreland,L., Monroe,K., Reiner,A.P., Wallace,R., Wilkens,L.R., Crawford,D.C., and Ritchie,M.D. (2011). The use of phenome-wide association studies (PheWAS) for exploration of novel genotype-phenotype relationships and pleiotropy discovery. Genet. Epidemiol. *35*, 410-422.

Pendergrass, S. A., Dudek, S., Crawford, D. C. C., and Ritchie, M. (2010) Synthesis-View: visualization and interpretation of SNP association results for multi-cohort, multi-phenotype data and meta-analysis. BioData Mining . 3:10

Pennington,J., Garner,S.F., Sutherland,J., and Williamson,L.M. (2001). Residual subset population analysis in WBC-reduced blood components using real-time PCR quantitation of specific mRNA. Transfusion *41*, 1591-1600.

Perez,G.L. and Keith,L. (1978). Electrocardiograms of twins. Prog. Clin. Biol Res. *24 Pt C*, 39-47.

Pfeufer,A., van,N.C., Marciante,K.D., Arking,D.E., Larson,M.G., Smith,A.V., Tarasov,K.V., Muller,M., Sotoodehnia,N., Sinner,M.F., Verwoert,G.C., Li,M., Kao,W.H., Kottgen,A., Coresh,J., Bis,J.C., Psaty,B.M., Rice,K., Rotter,J.I., Rivadeneira,F., Hofman,A., Kors,J.A., Stricker,B.H., Uitterlinden,A.G., van Duijn,C.M., Beckmann,B.M., Sauter,W., Gieger,C., Lubitz,S.A., Newton-Cheh,C., Wang,T.J., Magnani,J.W., Schnabel,R.B., Chung,M.K., Barnard,J., Smith,J.D., Van Wagoner,D.R., Vasan,R.S., Aspelund,T., Eiriksdottir,G., Harris,T.B., Launer,L.J., Najjar,S.S., Lakatta,E., Schlessinger,D., Uda,M., Abecasis,G.R., Muller-Myhsok,B., Ehret,G.B., Boerwinkle,E., Chakravarti,A., Soliman,E.Z., Lunetta,K.L., Perz,S., Wichmann,H.E., Meitinger,T., Levy,D., Gudnason,V., Ellinor,P.T., Sanna,S., Kaab,S., Witteman,J.C., Alonso,A., Benjamin,E.J., and Heckbert,S.R. (2010). Genome-wide association study of PR interval. Nat. Genet. *42*, 153-159.

Plant,L.D., Bowers,P.N., Liu,Q., Morgan,T., Zhang,T., State MW, Chen,W., Kittles,R.A., and Goldstein,S.A. (2006). A common cardiac sodium channel variant associated with sudden infant death in African Americans, SCN5A S1103Y. J Clin. Invest *116*, 430-435.

Post,W., Shen,H., Damcott,C., Arking,D.E., Kao,W.H., Sack,P.A., Ryan,K.A., Chakravarti,A., Mitchell,B.D., and Shuldiner,A.R. (2007). Associations between genetic variants in the NOS1AP (CAPON) gene and cardiac repolarization in the old order Amish. Hum. Hered. *64*, 214-219.

Poulsen,P., Kyvik,K.O., Vaag,A., and Beck-Nielsen,H. (1999). Heritability of type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance--a population-based twin study. Diabetologia *42*, 139-145.

Probst,V., Wilde,A.A.M., Barc,J., Sacher,F., Babuty,D., Mabo,P., Mansourati,J., Le Scouarnec,S., Kyndt,F., Le Caignec,C., Guicheney,P., Gouas,L., Albuisson,J., Meregalli,P.G., Le Marec,H., Tan,H.L., and Schott,J.J. (2009). SCN5A Mutations and the Role of Genetic Background in the Pathophysiology of Brugada Syndrome. Circ Cardiovasc Genet *2*, 552-557.

Purcell,S., Neale,B., Todd-Brown,K., Thomas,L., Ferreira,M.A., Bender,D., Maller,J., Sklar,P., de Bakker,P.I., Daly,M.J., and Sham,P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J Hum. Genet. *81*, 559-575.

Ramirez,A.H., Schildcrout,J.S., Blakemore,D.L., Masys,D.R., Pulley,J.M., Basford,M.A., Roden,D.M., and Denny,J.C. (2011). Modulators of normal electrocardiographic intervals identified in a large electronic medical record. Heart Rhythm. *8*, 271-277.

Ramsuran,V., Kulkarni,H., He,W., Mlisana,K., Wright,E.J., Werner,L., Castiblanco,J., Dhanda,R., Le,T., Dolan,M.J., Guan,W., Weiss,R.A., Clark,R.A., Karim,S.S., Ahuja,S.K., and Ndung'u,T. (2011). Duffy-null-associated low neutrophil counts influence HIV-1 susceptibility in high-risk South African black women. Clin. Infect. Dis. *52*, 1248-1256.

Reich,D.E., Cargill,M., Bolk,S., Ireland,J., Sabeti,P.C., Richter,D.J., Lavery,T., Kouyoumjian,R., Farhadian,S.F., Ward,R., and Lander,E.S. (2001a). Linkage disequilibrium in the human genome. Nature *411*, 199-204.

Reiner,A.P., Carty,C.L., Carlson,C.S., Wan,J.Y., Rieder,M.J., Smith,J.D., Rice,K., Fornage,M., Jaquish,C.E., Williams,O.D., Tracy,R.P., Lewis,C.E., Siscovick,D.S., Boerwinkle,E., and Nickerson,D.A. (2006a). Association between patterns of nucleotide variation across the three fibrinogen genes and plasma fibrinogen levels: the Coronary Artery Risk Development in Young Adults (CARDIA) study. J Thromb Haemost. *4*, 1279-1287.

Reiner,A.P., Lettre,G., Nalls,M.A., Ganesh,S.K., Mathias,R., Austin,M.A., Dean,E., Arepalli,S., Britton,A., Chen,Z., Couper,D., Curb,J.D., Eaton,C.B., Fornage,M., Grant,S.F., Harris,T.B., Hernandez,D., Kamatini,N., Keating,B.J., Kubo,M., LaCroix,A., Lange,L.A., Liu,S., Lohman,K., Meng,Y., Mohler,E.R., III, Musani,S., Nakamura,Y., O'Donnell,C.J., Okada,Y., Palmer,C.D., Papanicolaou,G.J., Patel,K.V., Singleton,A.B., Takahashi,A., Tang,H., Taylor,H.A., Jr., Taylor,K., Thomson,C., Yanek,L.R., Yang,L., Ziv,E., Zonderman,A.B., Folsom,A.R., Evans,M.K., Liu,Y., Becker,D.M., Snively,B.M., and Wilson,J.G. (2011). Genome-wide association study of white blood cell count in 16,388 African Americans: the continental origins and genetic epidemiology network (COGENT). PLoS. Genet. *7*, e1002108.

Roberts,W.C. and Day,P.J. (1985). Electrocardiographic observations in clinically isolated, pure, chronic, severe aortic regurgitation: analysis of 30 necropsy patients aged 19 to 65 years. Am. J Cardiol. *55*, 432-438.

Roden,D.M., Pulley,J.M., Basford,M.A., Bernard,G.R., Clayton,E.W., Balser,J.R., and Masys,D.R. (2008). Development of a large-scale de-identified DNA biobank to enable personalized medicine. Clin. Pharmacol. Ther. *84*, 362-369.

Rook,M.B., Bezzina,A.C., Groenewegen,W.A., van,G., I, van Ginneken,A.C., Jongsma,H.J., Mannens,M.M., and Wilde,A.A. (1999). Human SCN5A gene mutations alter cardiac sodium channel kinetics and are associated with the Brugada syndrome. Cardiovasc. Res. *44*, 507-517.

Rosamond,W.D., Chambless,L.E., Folsom,A.R., Cooper,L.S., Conwill,D.E., Clegg,L., Wang,C.H., and Heiss,G. (1998). Trends in the incidence of myocardial infarction and in mortality due to coronary heart disease, 1987 to 1994. N. Engl. J Med. *339*, 861-867.

Rose,K.M., Newman,B., Mayer-Davis,E.J., and Selby,J.V. (1998). Genetic and behavioral determinants of waist-hip ratio and waist circumference in women twins. Obes. Res. *6*, 383-392.

Rosenberg,N.A., Huang,L., Jewett,E.M., Szpiech,Z.A., Jankovic,I., and Boehnke,M. (2010). Genome-wide association studies in diverse populations. Nat. Rev. Genet. *11*, 356-366.

Rosendaal,F.R., Doggen,C.J., Zivelin,A., Arruda,V.R., Aiach,M., Siscovick,D.S., Hillarp,A., Watzke,H.H., Bernardi,F., Cumming,A.M., Preston,F.E., and Reitsma,P.H. (1998). Geographic distribution of the 20210 G to A prothrombin variant. Thromb Haemost. *79*, 706-708.

Russell,M.W., Law,I., Sholinsky,P., and Fabsitz,R.R. (1998). Heritability of ECG measurements in adult male twins. J. Electrocardiol. *30 Suppl*, 64-68.

Schott,J.J., Alshinawi,C., Kyndt,F., Probst,V., Hoorntje,T.M., Hulsbeek,M., Wilde,A.A., Escande,D., Mannens,M.M., and Le Marec,H. (1999). Cardiac conduction defects associate with mutations in SCN5A. Nat. Genetics *23*, 20-21.

Segal,J.B. and Moliterno,A.R. (2006). Platelet counts differ by sex, ethnicity, and age in the United States. Ann. Epidemiol. *16*, 123-130.

Sempos,C., Cooper,R., Kovar,M.G., and McMillen,M. (1988). Divergence of the recent trends in coronary mortality for the four major race-sex groups in the United States. Am. J Public Health *78*, 1422-1427.

Sempos,C.T., Bild,D.E., and Manolio,T.A. (1999). Overview of the Jackson Heart Study: a study of cardiovascular diseases in African American men and women. Am. J Med. Sci. *317*, 142-146.

Siegerink,B., Rosendaal,F.R., and Algra,A. (2009). Genetic variation in fibrinogen; its relationship to fibrinogen levels and the risk of myocardial infarction and ischemic stroke. J Thromb Haemost. *7*, 385-390.

Smith,J.G., Lowe,J.K., Kovvali,S., Maller,J.B., Salit,J., Daly,M.J., Stoffel,M., Altshuler,D.M., Friedman,J.M., Breslow,J.L., and Newton-Cheh,C. (2009c). Genome-wide association study of electrocardiographic conduction measures in an isolated founder population: Kosrae. Heart Rhythm. *6*, 634-641.

Smith,J.G., Magnani,J.W., Palmer,C., Meng,Y.A., Soliman,E.Z., Musani,S.K., Kerr,K.F., Schnabel,R.B., Lubitz,S.A., Sotoodehnia,N., Redline,S., Pfeufer,A., Muller,M., Evans,D.S., Nalls,M.A., Liu,Y., Newman,A.B., Zonderman,A.B., Evans,M.K., Deo,R., Ellinor,P.T., Paltoo,D.N., Newton-Cheh,C., Benjamin,E.J., Mehra,R., Alonso,A., Heckbert,S.R., and Fox,E.R. (2011). Genome-wide association studies of the PR interval in African Americans. PLoS. Genet. *7*, e1001304.

Smith,M.W., Patterson,N., Lautenberger,J.A., Truelove,A.L., McDonald,G.J., Waliszewska,A., Kessing,B.D., Malasky,M.J., Scafe,C., Le,E., De Jager,P.L., Mignault,A.A., Yi,Z., De,T.G., Essex,M., Sankale,J.L., Moore,J.H., Poku,K., Phair,J.P., Goedert,J.J., Vlahov,D., Williams,S.M., Tishkoff,S.A., Winkler,C.A., De,L., V, Woodage,T., Sninsky,J.J., Hafler,D.A., Altshuler,D., Gilbert,D.A., O'Brien,S.J., and Reich,D. (2004). A high-density admixture map for disease gene discovery in african americans. Am. J Hum. Genet. *74*, 1001-1013.

Snowden,C.B., McNamara,P.M., Garrison,R.J., Feinleib,M., Kannel,W.B., and Epstein,F.H. (1982). Predicting coronary heart disease in siblings--a multivariate assessment: the Framingham Heart Study. Am. J Epidemiol. *115*, 217-222.

Soranzo,N., Spector,T.D., Mangino,M., Kuhnel,B., Rendon,A., Teumer,A., Willenborg,C., Wright,B., Chen,L., Li,M., Salo,P., Voight,B.F., Burns,P., Laskowski,R.A., Xue,Y., Menzel,S., Altshuler,D., Bradley,J.R., Bumpstead,S., Burnett,M.S., Devaney,J., Doring,A., Elosua,R., Epstein,S.E., Erber,W., Falchi,M., Garner,S.F., Ghori,M.J., Goodall,A.H., Gwilliam,R., Hakonarson,H.H., Hall,A.S., Hammond,N., Hengstenberg,C., Illig,T., Konig,I.R., Knouff,C.W., McPherson,R., Melander,O., Mooser,V., Nauck,M., Nieminen,M.S., O'Donnell,C.J., Peltonen,L., Potter,S.C., Prokisch,H., Rader,D.J., Rice,C.M., Roberts,R., Salomaa,V., Sambrook,J., Schreiber,S., Schunkert,H., Schwartz,S.M., Serbanovic-Canic,J., Sinisalo,J., Siscovick,D.S., Stark,K., Surakka,I., Stephens,J., Thompson,J.R., Volker,U., Volzke,H., Watkins,N.A., Wells,G.A., Wichmann,H.E., Van Heel,D.A., Tyler-Smith,C., Thein,S.L., Kathiresan,S., Perola,M., Reilly,M.P., Stewart,A.F., Erdmann,J., Samani,N.J., Meisinger,C., Greinacher,A., Deloukas,P., Ouwehand,W.H., and Gieger,C. (2009). A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. Nat. Genet. *41*, 1182-1190.

Sorrentino,M.J., Vielhauer,C., Eisenbart,J.D., Fless,G.M., Scanu,A.M., and Feldman,T. (1992). Plasma lipoprotein (a) protein concentration and coronary artery disease in black patients compared with white patients. Am. J Med. *93*, 658-662.

Sotoodehnia,N., Isaacs,A., de Bakker,P.I., Dorr,M., Newton-Cheh,C., Nolte,I.M., van der,H.P., Muller,M., Eijgelsheim,M., Alonso,A., Hicks,A.A., Padmanabhan,S., Hayward,C., Smith,A.V., Polasek,O., Giovannone,S., Fu,J., Magnani,J.W., Marciante,K.D., Pfeufer,A., Gharib,S.A., Teumer,A., Li,M., Bis,J.C., Rivadeneira,F., Aspelund,T., Kottgen,A., Johnson,T., Rice,K., Sie,M.P., Wang,Y.A., Klopp,N., Fuchsberger,C., Wild,S.H., Mateo,L., I, Estrada,K., Volker,U., Wright,A.F., Asselbergs,F.W., Qu,J., Chakravarti,A., Sinner,M.F., Kors,J.A., Petersmann,A., Harris,T.B., Soliman,E.Z., Munroe,P.B., Psaty,B.M., Oostra,B.A., Cupples,L.A., Perz,S., de Boer,R.A., Uitterlinden,A.G., Volzke,H., Spector,T.D., Liu,F.Y., Boerwinkle,E., Dominiczak,A.F., Rotter,J.I., van,H.G., Levy,D., Wichmann,H.E., van Gilst,W.H., Witteman,J.C., Kroemer,H.K., Kao,W.H., Heckbert,S.R., Meitinger,T., Hofman,A., Campbell,H., Folsom,A.R., van Veldhuisen,D.J., Schwienbacher,C., O'Donnell,C.J., Volpato,C.B., Caulfield,M.J., Connell,J.M., Launer,L., Lu,X., Franke,L., Fehrmann,R.S., te,M.G., Groen,H.J., Weersma,R.K., van den Berg,L.H., Wijmenga,C., Ophoff,R.A., Navis,G., Rudan,I., Snieder,H., Wilson,J.F., Pramstaller,P.P., Siscovick,D.S., Wang,T.J., Gudnason,V., van Duijn,C.M., Felix,S.B., Fishman,G.I., Jamshidi,Y., Stricker,B.H., Samani,N.J., Kaab,S., and Arking,D.E. (2010). Common variants in 22 loci are associated with QRS duration and cardiac ventricular conduction. Nat. Genet. *42*, 1068-1076.

Splawski,I., Timothy,K.W., Tateyama,M., Clancy,C.E., Malhotra,A., Beggs,A.H., Cappuccio,F.P., Sagnella,G.A., Kass,R.S., and Keating,M.T. (2002). Variant of SCN5A sodium channel implicated in risk of cardiac arrhythmia. Science *297*, 1333-1336.

Splawski,I., Tristani-Firouzi,M., Lehmann,M.H., Sanguinetti,M.C., and Keating,M.T. (1997). Mutations in the hminK gene cause long QT syndrome and suppress IKs function. Nat. Genet. *17*, 338-340.

Steinberg,K.K., Sanderlin,K.C., Ou,C.Y., Hannon,W.H., McQuillan,G.M., and Sampson,E.J. (1997). DNA banking in epidemiologic studies. Epidemiol. Rev. *19*, 156-162.

Tan,B.H., Valdivia,C.R., Rok,B.A., Ye,B., Ruwaldt,K.M., Tester,D.J., Ackerman,M.J., and Makielski,J.C. (2005). Common human SCN5A polymorphisms have altered electrophysiology when expressed in Q1077 splice variants. Heart Rhythm. *2*, 741-747.

Teo,Y.Y., Small,K.S., Fry,A.E., Wu,Y., Kwiatkowski,D.P., and Clark,T.G. (2009). Power consequences of linkage disequilibrium variation between populations. Genet. Epidemiol. *33*, 128-135.

Teo,Y.Y., Small,K.S., and Kwiatkowski,D.P. (2010). Methodological challenges of genome-wide association analysis in Africa. Nat. Rev. Genet. *11*, 149-160.

Thelle,D.S. and Forde,O.H. (1979). The cardiovascular study in Finnmark county: coronary risk factors and the occurrence of myocardial infarction in first degree relatives and in subjects of different ethnic origin. Am. J Epidemiol. *110*, 708-715.

Tholin,S., Rasmussen,F., Tynelius,P., and Karlsson,J. (2005). Genetic and environmental influences on eating behavior: the Swedish Young Male Twins Study. Am. J Clin. Nutr. *81*, 564-569.

Thom,T., Haase,N., Rosamond,W., Howard,V.J., Rumsfeld,J., Manolio,T., Zheng,Z.J., Flegal,K., O'Donnell,C., Kittner,S., Lloyd-Jones,D., Goff,D.C., Jr., Hong,Y., Adams,R., Friday,G., Furie,K., Gorelick,P., Kissela,B., Marler,J., Meigs,J., Roger,V., Sidney,S., Sorlie,P., Steinberger,J., Wasserthiel-Smoller,S., Wilson,M., and Wolf,P. (2006). Heart disease and stroke statistics--2006 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Circulation *113*, e85-151.

Thompson,S.G., Kienast,J., Pyke,S.D., Haverkate,F., and van de Loo,J.C. (1995). Hemostatic factors and the risk of myocardial infarction or sudden death in patients with angina pectoris. European Concerted Action on Thrombosis and Disabilities Angina Pectoris Study Group. N. Engl. J. Med. *332*, 635-641.

Tishkoff,S.A., Reed,F.A., Friedlaender,F.R., Ehret,C., Ranciaro,A., Froment,A., Hirbo,J.B., Awomoyi,A.A., Bodo,J.M., Doumbo,O., Ibrahim,M., Juma,A.T., Kotze,M.J., Lema,G., Moore,J.H., Mortensen,H., Nyambo,T.B., Omar,S.A., Powell,K., Pretorius,G.S., Smith,M.W., Thera,M.A., Wambebe,C., Weber,J.L., and Williams,S.M.

(2009). The genetic structure and history of Africans and African Americans. Science *324*, 1035-1044.

Tokuhiro,S., Yamada,R., Chang,X., Suzuki,A., Kochi,Y., Sawada,T., Suzuki,M., Nagasaki,M., Ohtsuki,M., Ono,M., Furukawa,H., Nagashima,M., Yoshino,S., Mabuchi,A., Sekine,A., Saito,S., Takahashi,A., Tsunoda,T., Nakamura,Y., and Yamamoto,K. (2003). An intronic SNP in a RUNX1 binding site of SLC22A4, encoding an organic cation transporter, is associated with rheumatoid arthritis. Nat. Genet. *35*, 341-348.

Tosetto,A., Prati,P., Baracchini,C., Manara,R., and Rodeghiero,F. (2011). Association of plasma fibrinogen, C-reactive protein and G-455>A polymorphism with early atherosclerosis in the VITA Project cohort. Thromb Haemost. *105*, 329-335.

Turner,S.D., Berg,R.L., Linneman,J.G., Peissig,P.L., Crawford,D.C., Denny,J.C., Roden,D.M., McCarty,C.A., Ritchie,M.D., and Wilke,R.A. (2011). Knowledge-driven multi-locus analysis reveals gene-gene interactions influencing HDL cholesterol level in two independent EMR-linked biobanks. PLoS. One. *6*, e19586.

Van Norstrand,D.W., Tester,D.J., and Ackerman,M.J. (2008). Overrepresentation of the proarrhythmic, sudden death predisposing sodium channel polymorphism S1103Y in a population-based cohort of African-American sudden infant death syndrome. Heart Rhythm. *5*, 712-715.

Vasan,R.S. (2006). Biomarkers of cardiovascular disease: molecular basis and practical considerations. Circulation *113*, 2335-2362.

Vasan,R.S., Larson,M.G., Aragam,J., Wang,T.J., Mitchell,G.F., Kathiresan,S., Newton-Cheh,C., Vita,J.A., Keyes,M.J., O'Donnell,C.J., Levy,D., and Benjamin,E.J. (2007). Genome-wide association of echocardiographic dimensions, brachial artery endothelial function and treadmill exercise responses in the Framingham Heart Study. BMC. Med. Genet. *8 Suppl 1*, S2.

Voetsch,B. and Loscalzo.J (2003). Genetic Determinants of Arterial Thrombosis. Arterioscler Thromb Vasc Biol *24*, 216-229.

Wagner,G.P. and Zhang,J. (2011). The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms. Nat. Rev. Genet. *12*, 204-213.

Wang,Q., Curran,M.E., Splawski,I., Burn,T.C., Millholland,J.M., VanRaay,T.J., Shen,J., Timothy,K.W., Vincent,G.M., de,J.T., Schwartz,P.J., Toubin,J.A., Moss,A.J., Atkinson,D.L., Landes,G.M., Connors,T.D., and Keating,M.T. (1996). Positional cloning of a novel potassium channel gene: KVLQT1 mutations cause cardiac arrhythmias. Nat. Genet. *12*, 17-23.

Wang,Q., Shen,J., Li,Z., Timothy,K., Vincent,G.M., Priori,S.G., Schwartz,P.J., and Keating,M.T. (1995a). Cardiac sodium channel mutations in patients with long QT syndrome, an inherited cardiac arrhythmia. Hum. Mol. Genet. *4*, 1603-1607.

Wang,Q., Shen,J., Splawski,I., Atkinson,D., Li,Z., Robinson,J.L., Moss,A.J., Towbin,J.A., and Keating,M.T. (1995c). SCN5A mutations associated with an inherited cardiac arrhythmia, long QT syndrome. Cell *80*, 805-811.

Wassel,C.L., Lange,L.A., Keating,B.J., Taylor,K.C., Johnson,A.D., Palmer,C.D., Ho,L.A., Smith,N.L., Lange,E.M., Li,Y., Yang,Q., Delaney,J.A., Tang,W., Tofler,G., Redline,S., Taylor,H.A., Jr., Wilson,J.G., Tracy,R.P., Jacobs,D.R., Jr., Folsom,A.R., Green,D., O'Donnell,C.J., and Reiner,A.P. (2010c). Association of genomic loci from a cardiovascular gene SNP array with fibrinogen levels in European Americans and African-Americans from six cohort studies: the Candidate gene Association Resource (CARe). Blood.

Weir,B.S. and Cockerham,C.C. (1984). Estimating *F*- Statistics for the Analysis of Population Structure. Evolution *38*, 1358-1370.

Wilson,J.G., Rotimi,C.N., Ekunwe,L., Royal,C.D., Crump,M.E., Wyatt,S.B., Steffes,M.W., Adeyemo,A., Zhou,J., Taylor,H.A., Jr., and Jaquish,C. (2005). Study design for genetic analysis in the Jackson Heart Study. Ethn. Dis. *15*, S6-37.

Wilson,P.W., D'Agostino,R.B., Levy,D., Belanger,A.M., Silbershatz,H., and Kannel,W.B. (1998). Prediction of coronary heart disease using risk factor categories. Circulation *97*, 1837-1847.

Winkleby,M.A., Kraemer,H.C., Ahn,D.K., and Varady,A.N. (1998). Ethnic and socioeconomic differences in cardiovascular disease risk factors: findings for women from the Third National Health and Nutrition Examination Survey, 1988-1994. JAMA *280*, 356-362.

Winkler,C.A., Nelson,G.W., and Smith,M.W. (2010a). Admixture mapping comes of age. Annu. Rev. Genomics Hum. Genet. *11*, 65-89.

Xiao,R. and Boehnke,M. (2009). Quantifying and correcting for the winner's curse in genetic association studies. Genet. Epidemiol. *33*, 453-462.

Yusuf,H.R., Giles,W.H., Croft,J.B., Anda,R.F., and Casper,M.L. (1998). Impact of multiple risk factor profiles on determining cardiovascular disease risk. Prev. Med. *27*, 1-9.

Zhang,J.Z. and Redman,C. (1996). Fibrinogen assembly and secretion. Role of intrachain disulfide loops. J. Biol. Chem. *271*, 30083-30088.

Zhang,J.Z. and Redman,C.M. (1994). Role of interchain disulfide bonds on the assembly and secretion of human fibrinogen. J. Biol. Chem. *269*, 652-658.

Zheng,Z.J., Croft,J.B., Giles,W.H., and Mensah,G.A. (2001). Sudden cardiac death in the United States, 1989 to 1998. Circulation *104*, 2158-2163.

Zhong,H. and Prentice,R.L. (2010). Correcting "winner's curse" in odds ratios from genomewide association findings for major complex human diseases. Genet. Epidemiol. *34*, 78-91.

Zimmermann,K., Leffler,A., Babes,A., Cendan,C.M., Carr,R.W., Kobayashi,J., Nau,C., Wood,J.N., and Reeh,P.W. (2007). Sensory neuron sodium channel Nav1.8 is essential for pain at low temperatures. Nature *447*, 855-858.

Zuvich,R.L., Armstrong,L.L., Bielinski,S.J., Bradford,Y., Carlson,C.S., Crawford,D.C., Crenshaw,A.T., de,A.M., Doheny,K.F., Haines,J.L., Hayes,M.G., Jarvik,G.P., Jiang,L., Kullo,I.J., Li,R., Ling,H., Manolio,T.A., Matsumoto,M.E., McCarty,C.A., McDavid,A.N., Mirel,D.B., Olson,L.M., Paschall,J.E., Pugh,E.W., Rasmussen,L.V., Rasmussen-Torvik,L.J., Turner,S.D., Wilke,R.A., and Ritchie,M.D. (2011). Pitfalls of merging GWAS data: lessons learned in the eMERGE network and quality control procedures to maintain high data quality. Genet. Epidemiol. *35*, 887-898.