The Genetic Architecture of Leukoaraiosis and Cognitive Function

by

Jennifer Ann Smith

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Epidemiological Science)
in The University of Michigan
2011

Doctoral Committee:

Professor Sharon Reilly Kardia, Chair
Professor Patricia A. Peyser
Assistant Professor Sebastian K. Zoellner
Assistant Research Scientist Monica Uddin

DNA Is Life.

The rest is just translation.

www.thednastore.com

To anyone who has ever said to someone else,
"You can do this. I believe in you."
It matters.

For me, there were many of you.
I am incredibly lucky, truly blessed, and forever grateful.

Acknowledgements

I offer my most sincere gratitude to my mentor and friend, Sharon Kardia. Sharon, you have been a pleasure to work with, a constant source of inspiration and encouragement, and a truly insightful scientific role model throughout my graduate career. From you, I have learned how to be an accomplished scientist, but more importantly, I have learned how to be an accomplished human being. Thank you for so generously giving your time and attention not only to my scientific endeavors, but also to helping me navigate through the complicated journeys of life. I am also grateful to Pat Peyser for always making my scientific growth a priority in her busy schedule, for including me in fun and exciting projects, and especially for the amazing lunches at Zingerman's! I would also like to thank the other members of my doctoral committee, Sebastian Zoellner and Monica Uddin, for their helpful comments and encouragement through the dissertation process. I am looking forward to many more years of working with and learning from all of you.

I would also like to thank the members of the Kardia Research Group, past and present: Yan Sun, Tracy Fuller, Reagan Kelly, Kristin Meyers, Jian Chu, Wei Zhao, Erin Payne, Alicia Lazarus, Erin Bakshis, Koji Yanagisawa, Todd Greene, Joyce Wu, Ting Lu, and Maureen Roche. You have all been a true joy to work with. Thank you all for your patience, especially during the past year! A special thank you goes to Linda Feldkamp for all of the years that she has helped our research group stay organized and on top of

things, for her intuitive understanding that scientists sometimes lack common sense, and for her unfailing patience and support.

I would also like to acknowledge the GENOA and ARIC study participants and our collaborators that helped make this research possible: Steve Turner (Mayo Clinic), Tom Mosley (University of Mississippi), and Myriam Fornage (University of Texas). Data collection and statistical analysis for this dissertation was supported by the National Institutes of Health research grants U10 HL54457, U01 HL054481, R01 NS041558, R01 HL087660, and R01 HL075794, National Human Genome Research Institute training grant T32 HG00040, and the Horace Rackham Predoctoral Fellowship.

A huge thank you goes to R.C. for starting it all, and to K.G., J.K., J.R., K.B., J.S., L.C., C.A., C.M., C.H., V.C., R.P, L.R., and countless others who have helped me learn to live a joyous life one moment at a time. A special thank you goes to C.F., M.S., and H.P. for being there with me every step of the way, and for always being honest.

Last but most certainly not least, I would like to offer a huge thank you to the family and friends who have seen me through many years of graduate work and helped me celebrate all of the milestones. My family has been incredible, and although they often question why I am still in school after all these years, they continue to love and support me anyway. To Mom, Jeff, Jon, Leeza, Dad, Sue, Ashley, Grama June, and all of my aunts and uncles and cousins who have offered encouragement over the years, I really appreciate you. A special shout out to my sister for generously sharing her hugs and

laughs, for cooking and cleaning when I got overwhelmed, and for always being a true friend to me. Grampa, even though you didn't get a chance to see me begin my college career, I know you have been with me every step of the way. Grama and Grampa, I hope you know that every accomplishment I have is in due, in part, to you. I would also like to thank my adopted parents, Marty and Mike Kelly, for generously treating me as one of their own, and for going out of their way to make me smile and feel loved. I am lucky to have a family that has always has been, and always will be, proud of me.

I am forever grateful to Kristen, Monica, Honey-Mau, Nugget, Sunshine, Kristin, and Reagan for seeing the real me, loving me fiercely, and always being willing to listen. Sunshine, thank you for being a source of calm in the storm of dissertating and for always being ready with a well-timed "You Rock!!!". I hope you know that you totally rock, too. Kristin, you are my ultimate partner in crime. Thank you for bringing me ice cream instead of gelato, for teaching me the joys of backyard chickens, and for your limitless strength, humor, and commiseration. Reagan, this dissertation truly wouldn't have been possible without you. Thank you for being the best friend and companion that I could ever hope to have, for late nights spent deep in programming languages and manuscripts, for never giving up on me, and most of all, for the laughter. You mean more to me than I could ever express, and I am grateful for you every single day. And finally to Alison, for loving me unconditionally through all of it, for being brave enough to grow with me, for never missing goodnites, and for slowly but faithfully letting me into her uniquely beautiful umwelt. But most importantly, thank you to Alison for letting me take a picture of her wearing a tiara, for the sole purpose of making me smile.

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations

AD	Alzheimer's disease
ANOVA	analysis of variance
APOE	apolipoprotein E
ARIC	Atherosclerosis Risk in Communities
BMI	body mass index
BP	blood pressure
CAC	coronary artery calcification
CHD	coronary heart disease
CNV	copy number variation
CV	cross-validation
DAG	directed acyclic graph
DBP	diastolic blood pressure
DNA	deoxyribonucleic acid
DSM	Diagnostic and Statistical Manual of Mental Disorders
ENCODE	Encyclopedia of DNA Elements
FBPP	Family Blood Pressure Program
FDR	false discovery rate
FLAIR	fluid-attenuated inversion recovery
GENOA	Genetic Epidemiology Network of Arteriopathy

GMBI	Genetics of Microangiopathic Brain Injury
GO	Gene Ontology
GWAS	genome-wide association study
HDL	high density lipoprotein
LD	linkage disequilibrium
LDL	low density lipoprotein
LME	linear mixed effects model
LOD	logarithm of odds
MaCH	Markov Chain Haplotyper
MAF	minor allele frequency
MMSE	Mini-Mental State Examination
MRI	magnetic resonance imaging
NHLBI	National Heart Lung and Blood Institute
PCA	principal component analysis
QTL	quantitative trait locus
QTN	quantitative trait nucleotide
RNA	ribonucleic acid
SBP	systolic blood pressure
SNP	single nucleotide polymorphism
SOLAR	Sequential Oligogenic Linkage Analysis Routine
TG	triglycerides
TIV	total intracranial volume
VaD	vascular dementia

Abstract

Uncontrolled hypertension gives rise to arteriosclerotic target organ damage of the brain, heart, and kidneys. Ischemic brain injury due to hypertension is associated with clinical endpoints such as stroke and dementia. Subclinical measures of hypertension-related brain injury, such as leukoaraiosis (white matter hyperintensity on magnetic resonance imaging), are powerful predictors of stroke and dementia that aggregate in families and are likely to be the consequence of complex interactions between many genetic and environmental factors. Understanding the genetic architecture of leukoaraiosis may provide new insights into the etiology of the disease process and may help identify individuals that are at increased risk of developing stroke and dementia.

In this dissertation, both a candidate gene association study and a genome-wide association study were used to investigate the genetic architecture of leukoaraiosis in the white and African American cohorts of the Genetic Epidemiology Network of Arteriopathy (GENOA) study. Since the genetic component of inter-individual variation in leukoaraiosis is likely to involve multiple loci that act alone or through interactions with other genetic or environmental factors, we also explored interactions between pairs of single nucleotide polymorphisms (SNPs) and between SNPs and traditional risk factors. Finally, given that stroke and dementia are both associated with hypertension-related brain injury, we investigated the extent of pleiotropy between leukoaraiosis and seven measures of cognitive function using both biometrical and measured genetic

approaches. A greater understanding of the underlying genetic architecture of leukoaraiosis has the potential to provide insight into the etiological processes of stroke and dementia and to assist in earlier identification of individuals at increased risk for disease, the development of more efficacious treatments, and the tailoring of particular treatments to people most likely to respond positively.

Chapter 1

Introduction

Introduction

Hypertension affects approximately 1 in 3 American adults (76.4 million people), and accounts for $43.5 billion in yearly direct and indirect costs in the United States (Roger, 2011). Uncontrolled hypertension gives rise to target organ damage of the brain, heart, and kidney (Turner, 2000). This damage is due to arteriosclerosis (i.e., atherosclerosis and arteriolosclerosis) of the arteries that deliver blood to these organ systems, which results in clinical endpoints such as stroke and dementia, heart attack and heart failure, and chronic kidney disease (Turner, 2000).

Stroke causes considerable morbidity, mortality, and economic burden in the US, accounting for 1 in 18 deaths each year (Roger, 2011). Approximately 7 million Americans are currently living with the consequences of stroke, which include substantial cognitive and physical disabilities (Roger, 2011). An estimated 4% of direct health care costs in the US, well over $40 billion per year, is due to stroke (Donnan, 2008). Increased blood pressure is a powerful risk factor for stroke (Roger, 2011; Kannel, 1995), contributing to 70% of all strokes (Cubrilo-Turek, 2004) and accounting for a population attributable fraction of about 32%-38% of ischemic stroke (Hajat, 2004; Ohira, 2006; Goldstein, 2001; Roger, 2011). In adults over 55, having blood pressure greater than

140/90 mm Hg doubles the lifetime risk of stroke compared to having blood pressure less than 120/80 mm Hg (Seshadri, 2007b). Other risk factors for stroke include smoking, diabetes, and low physical activity (Roger, 2011).

Subclinical measures of brain injury allow the study of the pathophysiological processes that lead to stroke while also giving clinicians a method to identify individuals at increased risk of stroke (O'Sullivan, 2008; Markus, 2008) and dementia. Magnetic resonance imaging (MRI) of the brain provides a powerful tool to investigate these subclinical measures of brain injury, including cerebral white matter hyperintensity (leukoaraiosis) and asymptomatic small vessel (lacunar) strokes. Leukoaraiosis is a manifestation of hypertension-related arteriosclerosis of the brain and has been repeatedly documented as a powerful predictor of stroke (Kuller, 2005; Fu, 2005; Salerno, 1992). It is also a predictor of dementia (Kuller, 2005; Pantoni, 1997; Inzitari, 2003), another important public health concern in the aging U.S. population. Though leukoaraiosis is associated with elevated blood pressure and lack of blood pressure control (van Dijk, 2004; Liao, 1996), there is a substantial amount of inter-individual variation in leukoaraiosis volume among subjects with similar duration and severity of hypertension (Szolnoki, 2006; Schmidt, 2004).

It is poorly understood why a particular individual with hypertension develops target organ damage of one type but not another. Investigation of genetic factors underlying the susceptibility to various types of target organ damage provides a means to better understand these inter-individual differences. In order to identify the genetic risk factors

underlying hypertension and its different types of target organ damage, the Genetic Epidemiology Network of Arteriopathy (GENOA) study was initiated in 1995 (Daniels, 2004). This dissertation focuses on the genetic susceptibility to leukoaraiosis, a subclinical measure of hypertension-related brain injury, among participants in the GENOA study. Other GENOA investigators are focusing on measures of target organ damage such as coronary artery calcification, a measure of coronary artery atherosclerosis, and glomerular filtration rate, a measure of kidney disease. Ultimately, studies of the genetic correlations and pleiotropic genetic effects across target organ damage phenotypes will help elucidate the shared genetic component of risk for better prediction of at-risk subgroups in the population at large.

Background and Public Health Significance

Hypertension

Hypertension is commonly defined as having a systolic blood pressure ≥ 140 mm Hg or a diastolic ≥ 90 mm Hg, taking antihypertensive medication, or having been told at least twice by a physician or other health professional that one has high blood pressure (Roger, 2011). A substantial proportion of the American population are not aware of their hypertension (20.4%), are not under current treatment (29.1%), or do not have it adequately controlled (52.2%) (Roger, 2011). A similar proportion of men and women are affected by hypertension, and risk factors include increasing age, ethnicity, family history of hypertension, lower education and socioeconomic status, greater weight, tobacco use, dietary factors, and lower physical activity (Roger, 2011). However, in

general, these risk factors do not explain the majority of variation in the risk of developing hypertension.

Hypertension has a higher prevalence in US blacks than among US whites, though it is an increasing public health concern among both groups. Between the periods of 1988-1994 and 1999-2002, age-adjusted prevalence of hypertension increased from 24.3% to 28.1% among US whites and from 35.8% to 41.4% among US blacks (Hertz, 2005). Blacks also develop hypertension at earlier ages than whites, have higher average blood pressures, and are at increased risk for hypertension-related clinical endpoints related to hypertension (Roger, 2011).

Uncontrolled hypertension is the most prevalent risk factor for diseases in a number of organ systems, including the brain, heart, and kidney (Turner, 2000). Hypertension-related target organ damage occurs primarily due to chronic ischemia resulting from atherosclerosis (hyperplasia and remodeling of the intimal cells) of the larger elastic conduit arteries that deliver blood to the organ and arteriolosclerosis (thickening of the vessel media due to remodeling, hypertrophy, and/or hyperplasia of smooth muscle) of the smaller arteries within the organ itself. Figure 1.1 depicts the role of genetic and environmental factors in influencing arteriosclerosis (i.e., atherosclerosis and arteriolosclerosis) and its impact on target organ systems.

Hypertension is associated with manifestations of coronary heart disease (CHD), such as angina pectoris, myocardial infarction, and sudden cardiac death (MacMahon, 1990).

Subclinical measures of arteriosclerosis of the heart, such as coronary artery calcification and left ventricular hypertrophy are also associated with hypertension and are predictive of clinical CHD endpoints (Turner, 2002; Bielak, 2004; Bielak, 2000; Arnett, 2004; Koren, 1991). In the kidney, arteriosclerotic damage due to hypertension strongly contributes to chronic kidney disease and end-stage renal disease (Klag, 1996; Tozawa, 2003). Subclinical measures such as glomerular filtration rate and albumin/creatinine ratio are indicative of risk for these clinical endpoints (National Kidney Foundation, 2002).

The brain is also susceptible to both atherosclerotic and arteriolosclerotic processes related to hypertension, each accounting for one of the two recognized subtypes of ischemic cerebrovascular disease that increases risk of later-life stroke and dementia. The atherosclerotic process in the brain mirrors the atherosclerotic process that occurs in CHD. It tends to affect the larger vessels in the brain and is most strongly related to cardiovascular risk factors such as cholesterol. Ischemic brain injury resulting from cerebral arteriolosclerosis, in contrast, is more strongly related to hypertension and affects the smallest vessels in the brain, leading to small, often undetected areas of ischemic damage. Little is known about how various physiological and genetic pathways may lead to target organ damage of one type, but not another, in hypertensive individuals.

Stroke

Stroke is the second most common cause of death and disability-adjusted life-years in industrialized countries (Lopez, 2001; Murray, 1997) and the third most common cause

of death in the US, accounting for approximately 1 in 18 deaths in 2007 (Roger, 2011). New or recurrent stroke occurs in 795,000 people annually, and about 610,000 of these are first attacks (Roger, 2011), though the stroke death rate has decreased substantially over the past several decades due in part to successful management of hypertension (Bonita, 1986; Bornstein, 2006). Approximately 7 million Americans in the US are currently living with the consequences of stroke (Roger, 2011), and an estimated 4% of direct health care costs in the United States are due to stroke (approximately $40.9 billion per year in 1997 dollars) (Donnan, 2008). Prevalent stroke was estimated to be 3.8% among white US adults and 4.0% among US black adults (Roger, 2011). The risk of first-ever stroke is almost twice as high for blacks than whites, and is higher for men than women (age-adjusted incidence among people ages 45-84 is 6.6, 4.9, 3.6, and 2.3 per 1,000 in black men, black women, white men, and white women, respectively) (Roger, 2011).

Ischemic stroke accounts for 87% of all stroke in the U.S., and hemorrhagic stroke accounts for the remainder (Roger, 2011). Ischemic stroke can be further classified into categories according to the mechanism of vessel occlusion including cardioembolism, large-artery embolism, or small vessel occlusion (lacunar) (Adams, 1993). Leukoaraiosis is associated with small-vessel (lacunar) stroke (Markus, 2008), one of the most heritable subtypes, which accounts for approximately 20% of ischemic strokes (O'Sullivan, 2008; Schulz, 2004).

Dementia

Dementia consists of an array of conditions that result in measurable cognitive decline that impairs physical, social, and intellectual function (Kukull, 2002; Haan, 2004). Dementia is an important public health burden in the U.S. and abroad (Kukull, 2002; Haan, 2004), and the World Health Organization predicts that there will be approximately 29 million people affected by all forms of dementia by the year 2020 (Essink-Bot, 2002). Alzheimer's disease (AD) and other dementias affect over 5.2 million Americans, including between 200,000 and 500,000 people under the age of 65. Dementias place a heavy economic burden on the health care system, with each Medicare patient with dementia accounting for more than three times as much spending than the average beneficiary (Alzheimer's Association, 2008). The aging population of the U.S. is expected to dramatically increase the prevalence of dementia, which is thought to affect 3%-11% of people older than 65 and 25%-47% of people older than 85 (Boustani, 2003). Older African Americans are about twice as likely to develop AD and other dementias as older whites, which may in part be due to the higher prevalences of hypertension and diabetes and lower average socioeconomic status of this group (Alzheimer's Association, 2010).

AD accounts for approximately 53%-80% of all cases of dementia in populations of European origin, and vascular dementia (VaD) is thought to account for a substantial portion of the remainder (approximately 15%) (Kukull, 2002; Lobo, 2000). Exact percentages of VaD are often difficult to estimate because differentiating between VaD

and AD is difficult in many cases, and it is becoming increasingly apparent that vascular pathology can contribute to dementia in people with AD (Kukull, 2002; Breteler, 2000). Leukoaraiosis and multiple lacunar (small vessel) strokes, both caused by cerebral small vessel disease, are the primary markers of VaD (Geldmacher, 1997) and are thought to be contributors to cognitive impairment in individuals who have not yet progressed to dementia (Pantoni, 2007; De Groot, 2002; Schmidt, 2007).

Leukoaraiosis

Leukoaraiosis physiology and pathology

The human brain is composed of gray matter (the cerebral cortex) that is responsible for consciousness, movement, and cognition and white matter that consists of nerve fibers that transmit impulses among cerebral areas and to the central nervous system (see Figure 1.2) (Marieb, 1998). Leukoaraiosis is visible as bright spots in the white matter on T2-weighted MRIs, as shown in Figure 1.3 (O'Sullivan, 2008; Markus, 2008). Leukoaraiosis ranges in severity from small, distinct areas of white matter hyperintensity (punctuate lesions) to large regions of white matter hyperintensity (early confluent or confluent lesions) (O'Sullivan, 2008). Leukoaraiosis is thought to be a marker of cerebral small vessel disease in the long, narrow penetrating arterioles that supply the white matter with blood (Markus, 2008). This type of small vessel disease is defined by areas of diffuse arteriolosclerosis with deposits of a proteinaceous substance that includes fibrin, amyloid, and collagen, which results in thickening of the vessel and chronic ischemia that leads to demyelination, axonal loss, and gliosis (O'Sullivan, 2008; Markus, 2008; Pantoni, 1997). It occurs in regions of the brain that have low perfusion pressure, such as the deep white

matter, and results in chronic ischemia and multiple diffuse infarctions due to small vessel occlusions (lacunar infarctions), both of which are visible as leukoaraiosis on MRIs (Markus, 2008). In regions of leukoaraiosis, there appears to be decreased blood flow (hypoperfusion) and impaired ability to regulate blood flow (autoregulation) (Markus, 2008).

The primary risk factors for leukoaraiosis are older age and hypertension, supporting the notion that the mechanism of pathology results from arteriolosclerosis rather than an atherosclerotic process (O'Sullivan, 2008; Markus, 2008) since atherosclerosis is additionally associated with lipid infiltration and thus predicted by plasma lipid levels (e.g. low and high density lipoprotein cholesterol levels). Recently, it has been suggested that endothelial dysfunction – i.e. the inability of endothelial cells to perform tasks such as mediation of coagulation, platelet adhesion, and immune response - may be the intermediate process between hypertension and the alterations in blood flow observed in areas of leukoaraiosis (Markus, 2008; Hassan, 2003; Szolnoki, 2007a). Circulating endothelial markers may show a pro-coagulant pattern of endothelial function (e.g. higher circulating levels of thrombomodulin (*TM*) and lower circulating levels of tissue factor pathway inhibitor (*TFPI*)) that is specific to leukoaraiosis (Hassan, 2003) and may be related to progression of leukoaraiosis (Markus, 2005). Further support for endothelial dysfunction comes from the strong association between leukoaraiosis and elevated homocysteine level, which is hypothesized to mediate its effect through endothelial damage (Hassan, 2004b).

Leukoaraiosis epidemiology

In studies that have conducted MRIs on older adults without stroke or dementia, a common finding is that a small amount of leukoaraiosis is detected in the majority of subjects. However, prevalence depends greatly on the population and rating scales utilized (de Leeuw, 2001). Schmidt (2007) estimated that leukoaraiosis was present in more than half of elderly patients in routine scans (Schmidt, 2007), and two studies of stroke- and dementia-free adults ages 50-75 reported that less than a third of participants had no leukoaraiosis at all (Markus, 2005; Mosley, 2005). In the Framingham Heart Study, about half of the subjects (mean age 62.6 years of age) had a leukoaraiosis volume that was less than 1% of their total cranial volume (DeStefano, 2006). After accounting for age, blacks tend to have a slightly larger amount of leukoaraiosis and related downstream consequences such as silent brain infarction, which may be due to the increased prevalence of hypertension in this group (Prabhakaran, 2008). Though older age and hypertension are the primary risk factors for leukoaraiosis, others include previous stroke, ischemic heart disease, and possibly diabetes (Fu, 2005; Pantoni, 1997).

A strong predictive relationship has been found between leukoaraiosis and stroke. Risk of stroke increases as leukoaraiosis severity increases, independent of traditional stroke risk factors such as hypertension, diabetes, and history of myocardial infarction (Markus, 2005). Leukoaraiosis is also a strong predictor of recurrent stroke in patients with previous history of stroke, with patients having severe leukoaraiosis at four times the risk of those with no leukoaraiosis after adjustment for other risk factors (Kuller, 2004).

Leukoaraiosis is also associated with cognitive decline and dementia in multiple epidemiological studies. It is detected in approximately 60%-70% of persons with VaD and AD (O'Sullivan, 2008; Jeerakathil, 2004), though it is typically more severe in VaD (Lehericy, 2007). Several studies have shown that leukoaraiosis is predictive of incident VaD (Kuller, 2005; Prins, 2004). In a review of leukoaraiosis and cognition, Pantoni et al. (2007) conclude that despite different study characteristics, there is almost invariably evidence of an effect of leukoaraiosis on cognition. In particular, leukoaraiosis is more strongly associated with decreasing cognitive performance than memory and is also associated with a decline in motor performances such as gait disturbances (Pantoni, 2007; Schmidt, 2007). The rate of progression of leukoaraiosis over time is also related to cognitive decline (De Groot, 2002; Schmidt, 2007), and the severity of leukoaraiosis at baseline is a significant predictor of progression (Schmidt, 2007). It is also important to keep in mind, however, that other factors may affect the association between leukoaraiosis and cognitive decline such as lacunar infarcts and brain atrophy (Pantoni, 2007; Schmidt, 2007).

Leukoaraiosis and cognitive function

Dementia is a heterogeneous group of disorders with variable etiology that involves impairment in cognitive domains such as memory, executive function, and language as well as specific physical impairments such as gait abnormalities that cause significant impairment in social or occupational function and represent a decline from a previous level of functioning (American Psychiatric Association, 2000). The Mini-Mental State

Examination (MMSE), developed in 1975 as a brief, standardized instrument to screen for impairment in a limited number of cognitive functions (Lezak, 1995) is now a commonly-used screening tool for classifying the extent of individual cognitive impairment and assessing change in cognitive function level over time (Crum, 1993; Tombaugh, 1992). Formal diagnosis of dementia and differentiating between different types of the major forms of dementia involves neuropsychological tests in combination with visual imaging of the brain (Pohjasvaara, 2000).

The differential diagnosis of VaD incorporates the underlying vascular cause (evidence of excessive leukoaraiosis on brain MRI or evidence of ischemic stroke) as well as the cognitive and physical symptomology (Pohjasvaara, 2000). One of the criteria for differential diagnosis of VaD in the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) is "focal neurological signs and symptoms or laboratory evidence indicative of cerebrovascular disease (multiple infarctions involving cortex and underlying white matter) that are judged to be etiologically related to the disturbance)" (American Psychiatric Association, 2000). Therefore, leukoaraiosis is not only a risk factor for VaD, but also is part of the diagnostic criteria. The pathological manifestation of leukoaraiosis in VaD is highly variable but typically involves multifocal or diffuse leukoaraiosis in the basal gangli, thalamus, and white matter that affects various functionally important areas of the brain including the frontal and limbic cortical structures (Jellinger, 2008).

Several studies have demonstrated an association between hypertension in midlife and cognitive dysfunction in later life (Elias, 1993; Launer, 2000), and it has been hypothesized that this is due to the cumulative effects of sublinical damage due to cerebrovascular disease (Knopman, 2001; Swan, 1998). Knopman et al. (2001) showed that cognitive decline in midlife is also associated with hypertension, lending credence to the claim that later cognitive dysfunction and dementia are the clinical manifestations of a disease process in the brain that is cumulative throughout the lifespan. Leukoaraoisis, a measure of the extent of subclinical damage due to hypertension-associated small vessel disease in the brain, is likely to be involved in one of the major mechanistic pathways between hypertension in midlife and cognitive decline (Sierra, 2006).

Strategies for Studying the Genetic Architecture of Complex Traits

The genetic architecture of a trait is defined by the genes that affect the trait, the genetic variations within those genes, the frequency distribution of those variations, and the effects of those genetic variations on the trait mean levels, variability (e.g. plasticity) and covariability (e.g. pleiotropy) when they are considered alone or when interacting with other genes and measures of environment (both internal and external). Recent evidence from association studies demonstrates that many genetic variations with small effect sizes may act in conjunction to influence the development of complex traits (Scott, 2007; Zeggini, 2008), and it has long been known that quantitative traits are likely to be influenced by many genes acting with a wide range of effect sizes (Falconer, 1996; Wang, 2005). It has also been demonstrated that the effect of any particular genetic variation is likely to depend on its environmental context (gene-environment interaction)

or its genetic context (gene-gene interaction or epistasis) (Kardia, 2000; Kraft, 2007; Sing, 2004; Mackay, 2009; Mackay, 2010; Zhou, 2009). It is clear that the genetic architecture of complex traits such as leukoaraiosis will be best elucidated by examining both single nucleotide polymorphism (SNP) main effects as well as their context-dependent effects. A brief summary of the multiple strategies used to study the genetic architecture of complex traits is presented below, with special emphasis on the methods used in this dissertation.

Heritability

Before undertaking research to identify specific genetic factors that contribute to a trait of interest, a typical first step is to examine how much of the observed variation in the trait is attributable to genetic factors and environmental factors using family studies (e.g. twins, sibships, nuclear families, etc.). Quantitative genetic theory of polygenic inheritance provides a means of decomposing the variation in a trait to estimate "narrow sense" heritability, which is defined as the proportion of variance in a trait due to variability in additive genetic factors (Falconer, 1996). Heritabilities are estimated biometrically with observed phenotypic information from related individuals, requiring no directly measured genetic information. Additive genetic effects are unobservable and can only be modeled using the variance-covariance matrix of the trait, expressed as a function of identity-by-descent relationships. The expected covariance of a trait between a pair of individuals is modeled as a function of the variance parameters and the expected correlation between the individuals (Sing, 1987; Kempthorne, 1954).

Heritability of a trait is a population-specific parameter that can change over age, environments, and genetic backgrounds (Visscher, 2008b). Population parameters that influence heritability estimates include allele frequencies and differences in demographic histories that affect population genetic and environmental variation. Study-specific factors that influence heritability measures include the size and structure of the pedigrees, measurement error, bias due to assortative mating and/or selection, and the amount of variation among individuals and families in environmental factors. Traits with higher heritability are prime candidates for linkage and association studies. Estimating heritability is particularly important to evaluate new phenotypes that emerge as technologies advance (such as MRI) since it can be used to test the fundamental null hypothesis that no gene variations are involved in explaining inter-individual variation in the population. Recently, much attention has been focused on the "missing heritability" since the relative impact of measured genes identified by the recent success in genome-wide association studies (GWAS) have explained very modest amounts of trait variability (Manolio, 2009; Eichler, 2010).

Linkage studies

Linkage and association techniques are the two basic analysis methods that have been used for human gene mapping and localizing the specific genetic regions that affect disease-related traits. Parametric linkage analysis is a family-based method for scanning a set of genome-wide markers for co-segregation with a trait of interest by counting offspring with recombinant or non-recombinant allele combinations (Hoh, 2001). Maximum likelihood methods are used to identify genomic regions associated with a trait

assessed by the log of the odds (LOD) scores – i.e. the odds of the likelihood of linkage versus the likelihood of no linkage (Morton, 1955). An advantage of linkage studies is that they require a much less dense set of genome-wide markers than other genetic study techniques such as association studies, because the goal of linkage studies is to identify broad genomic regions rather than specific genetic variants that have high likelihood of affecting a trait. Linkage studies have been relatively successful at mapping traits that are inherited in a Mendelian fashion and even in mapping rarer forms of complex diseases that have strong single gene effects in specific families, such as breast cancer, type I diabetes, and AD (Risch, 2000). However, most common chronic diseases aggregate but do not segregate in families and the success of linkage, even with advances in non-parametric linkage, has been limited. This limited success in not surprising in light of the mounting evidence that many quantitative human traits and common diseases are highly genetically heterogeneous, context dependent, and probably also additionally reflect many genes with small effects that follow a polygenic rather than a Mendelian model of inheritance (Risch, 2000; Terwilliger, 2009; Weiss, 2000; Terwilliger, 1998; Van Heyningen, 2004).

Association studies

Association studies compare allele or genotype frequencies among individuals with different trait values instead of tracing allelic inheritance in families (Altshuler, 2008). The advantages of association studies over linkage studies are that they can be performed in unrelated individuals, may have more power to detect common alleles (minor allele frequency > 0.05) with modest effect, and localize the genetic effect to specific

polymorphisms rather than broad genomic regions (Hirschhorn, 2005). Initially, candidate gene studies were used to test a small set of selected genes and a small set of selected variations. Recently, GWAS have become the standard method to identify gene regions associated with complex human traits. The main disadvantages of association studies are that they may be susceptible to population stratification and they can only be used to detect common genetic variants.

Candidate gene studies

Until the past decade, genetic association studies for complex traits have been conducted using a relatively small number of SNPs (<500) from candidate genes, which are chosen based on gene function ("biological candidate genes" that are known or hypothesized to be involved in disease pathophysiology) or genomic location ("positional candidate genes" that are under "linkage peaks" identified in family studies) (Hirschhorn, 2005). Candidate gene studies have the advantage of being grounded in prior knowledge and have been used successfully to identify genetic variations that have replicable effects across studies and populations for complex traits such as leukoaraiosis (Gormley, 2007; Hassan, 2002). The disadvantage of candidate gene studies is that they typically examine only a small fraction of the variations in the gene and hence rely heavily on linkage disequilibrium (LD) with unknown causal mutations to identify the key genes and variants. In addition, influential genes that are outside the pathways known to affect the trait are absent from investigations.

Though there have been multiple successful replications of SNP associations in candidate gene studies, the majority of published candidate gene-disease associations have failed to replicate (Hirschhorn, 2002). One reason for this failure to replicate was the study designs and analytical strategies that were not in place to safeguard against false positives (e.g. small studies with no replication cohorts). Although replication of genetic associations within other study samples is currently the gold standard for reducing false positive reporting (NCI-NHGRI Working Group on Replication in Association Studies, 2007), replication studies are sometimes not possible with candidate gene studies (compared to the more recent GWAS) because other samples typically lacked the genotype information for the same SNPs of interest or do not have the same state-of-the-art measurements of traits, such as leukoaraiosis. When replication samples are unavailable, innovative combinations of statistical techniques may be necessary to reduce the rate of false positive associations. Combining several statistical methods such as cross-validation, internal replication, and use of the false discovery rate (FDR) provides a means for helping to distinguish between true and false positives, and a filtering approach based on these techniques has been successfully been used in practice (Smith, 2008; Smith, 2009; Kardia, 2008).

Genome-wide association studies (GWAS)

Technological advances and a vast reduction in genotyping cost have recently resulted in a significant expansion of the number of genetic markers that can be typed for a given subject, allowing progression beyond candidate gene studies. GWAS, which measure genetic variation at a high density within the genome (>100,000 SNPs), have several

advantages over candidate gene studies. Primarily, scanning hundreds of thousands of genetic markers across the genome provides an agnostic, unbiased approach to finding common genetic variations that play a role in human trait variation because no assumptions are made about the genomic location of causal variants (Hirschhorn, 2005).

GWAS take advantage of the naturally occurring LD structure of the human genome to identify genomic regions likely to harbor disease-associated deoxyribonucleic acid (DNA) variations. Most genetic variants are strongly correlated with their physically neighboring SNPs, giving rise to sections known as LD blocks, in which highly correlated SNPs can be used as proxies for one another. Thus, causal variants can be detected either by direct genotyping or by being highly correlated to a directly genotyped SNP (Hirschhorn, 2005). The existence of LD in the human genome also facilitates the imputation of genotypes at millions of unmeasured SNP markers with the use of phased data from the International HapMap Project (International HapMap Consortium, 2005) through imputation software such as Markov Chain Haplotyper (MaCH) (Li, 2006). Imputation of unmeasured genotypes allows study samples that have been genotyped on different platforms to be used as replication datasets.

One disadvantage of GWAS is that multiple testing issues accompany the examination of millions of genetic markers. Replication in independent samples provides the strongest evidence of true positive association, and rigorous standards have been proposed for conducting, replicating, and publishing GWAS (NCI-NHGRI Working Group on Replication in Association Studies, 2007). The need to replicate in independent samples

has resulted in huge collaborative agreements among studies with thousands of participants. The focus on these collaborative agreements has been on detecting single SNPs with small to moderate main effects on dichotomous disease outcomes (typically using case-control studies) or on quantitative traits (typically using population-based samples), but an interest in examining gene-environment and gene-gene interactions is now emerging. However, sample size requirements for detecting gene-gene and gene-enviornment interactions tend to be very large.

Studies of pleiotropy

Pleiotropy is most simply defined as the effect of variation in a single gene on variation in multiple traits (Hodgkin, 1998). A more stringent definition of pleiotropy is "the phenomenon in which a single gene controls several distinct, seemingly unrelated, phenotypic effects" (Zou, 2008), implying that pleiotropic genes by nature perform multiple biological functions. Pleiotropic genetic variation has been studied extensively in model organisms using quantitative genetic methods, primarily linkage, in order to inform breeding programs, examine evolutionary mechanisms, and elucidate molecular pathways (Hodgkin, 1998; Mackay, 2009; Cheverud, 2004). Studies of quantitative trait loci in model organisms and preliminary studies of pleiotropic effects in humans using bivariate linkage and association techniques have set the stage for the development of systems-level genetics approaches to examining the pleiotropic mechanisms that affect human health and contribute to variation in common chronic diseases.

A classic example of pleiotropy in humans is the ε4 allele of the apolipoprotein E (*APOE*) gene, which is associated with risk of atherosclerosis as well as dementia (Dickstein, 2010). GWAS studies have also revealed some surprising pleiotropic findings at the gene level in humans. Several genic regions have been implicated in cancers in addition to at least one other human disease, as have individual SNPs. For example, at least three SNPs associated with type 2 diabetes have also been found to be associated with both prostate and colon cancers (Winckler, 2007; Gudmundsson, 2007; Zeggini, 2008; Thomas, 2008; Slattery, 2008). Methods for studying pleiotropy using measured genotype data in humans, such as those that derive metrics of pleiotropy from GWAS results, are beginning to emerge (Johnson, 2009; Karasik, 2010).

The study of pleiotropy in model organisms and humans serves several functions. First, it serves to further the understanding and elucidation of the complex biological pathways that regulate the development of phenotypes, providing information about normal cellular function, normal development and function at the organismal level, connections between previously unrecognized related health and diseases processes, and increased predictive ability in breeding programs and health management (Hodgkin, 1998). Second, it provides insight into the mechanisms of evolution, as effects on multiple phenotypes due to a single genetic variant may pose severe evolutionary constraints (Cheverud, 2004). Understanding the mechanisms of pleiotropy will lead to increased ability to identify connections between genetic variation and disease in humans, increase success of animal breeding programs, and shed light on the mechanism of natural selection in the study of evolution.

Missing Heritability

Successes of GWAS

To date, GWAS studies have had notable success in identifying hundreds of replicated SNP associations with various human traits and diseases, and much has been learned about the complexity of the genetic architecture that underlies the variation in human traits (Hindorff, 2009; Donnelly, 2008). However, progress toward identifying the genes and SNPs that impact most traits has been more incremental and difficult than anticipated, and even the most elegantly designed studies and large consortia have been unable to identify sets of SNPs that explain more than a small portion of trait variation (Manolio, 2009). Even for the identified SNPs with the strongest associations with disease susceptibility, it is remains unclear the utility that these SNPs will have in predicting disease risk (Donnelly, 2008).

GWAS variants leave much of the trait variability unexplained

Many GWAS have been performed on traits that exhibit a high heritability, yet identifying genetic variants that contribute substantially to inter-individual variation either alone or in combination has proven difficult even with enormous study populations. A classically cited example of a highly heritable quantitative trait that has been intensely studied via GWAS is human height. GWAS studies have been relatively successful in identifying variants associated with height, as three large studies with a combined sample size of over 60,000 individuals identified over 50 loci reproducibly shown to affect height (Visscher, 2008a). Though heritability of height is approximately

80%, the identified variants explained less than 5% of the phenotypic variance of this trait (Manolio, 2009; Visscher, 2008a). Similarly, more than 30 loci have been associated with Crohn's disease, but these account for less than 10% of the phenotypic variance (Barrett, 2008). This discrepancy between the large biometrically estimated contribution of additive genetic effects to trait variance and the small proportion of trait variance explained by the SNPs detected in GWAS studies has been termed "missing heritability" (Manolio, 2009; Eichler, 2010).

Possible explanations for missing heritability

Many explanations have been posed and discussed as the reasons why much of the heritability of human traits remains unexplained by GWAS findings. The leading arguments are that an additional proportion of the heritability may be explained by rare and structural variants such as copy number variants (CNVs) (Eichler, 2010; Frazer, 2009), epigenetic and parent-of-origin effects (Eichler, 2010; Kong, 2009), low power to detect gene-gene interaction, and failure to properly account for shared environment in family studies leading to inflated heritability estimates (Manolio, 2009). While it is possible that heritabilities may be overestimated in some cases, animal studies tend to agree with heritability estimates in humans (Manolio, 2009), and methods that use observed identity-by-descent sharing of markers in siblings have also estimated consistently high heritabilities for traits such as height (Visscher, 2006).

Missing heritability may be due to low LD and stringent significance criteria

Assuming that heritabilities have not been consistently overestimated, it is possible that the discrepancy between heritability estimates and the genetic variance explained by GWAS findings may be due to several factors. One possibility described by Yang et al. (2010) is that causal variants do not reach the threshold of statistical significance currently applied in GWAS studies, and a second possibility is that genotyped SNPs are not in complete LD with the causal variants and thus do not fully capture the strength of their association. Yang and colleagues elegantly examined these possible explanations for missing heritability by estimating the proportion of variance for height explained by measured genotypes. Their findings indicated that 45% of the variance of height could be explained by incorporating the effects of all measured SNPs simultaneously, showing that much of the missing heritability may be due to causal variants not meeting the stringent thresholds of statistical significance applied to GWAS findings. Next, they showed that low LD between the genotyped SNPs and the causal variants, particularly those with lower allele frequencies such as rare variants, could explain the remaining 35% of the missing heritability.

Missing heritability may be due to rare variants

GWAS represent a powerful tool for detecting common variants that contribute to disease, but there is emerging evidence that other types of genetic variants such as rare variants and CNVs also play a role in the genetic architecture of at least some human diseases and traits. Rare variants are defined as those with MAF less than about 1%, and these go undetected in current GWAS because they have been excluded from most

genotyping platforms. Several studies have already demonstrated that rare variants in candidate genes are associated with traits such as blood pressure (Ji, 2008), colorectal cancer (Fearnhead, 2005), triglyceride levels (Cohen, 2005; Wang, 2007), and body mass (Ahituv, 2007) in both European Americans and African Americans. It has been proposed that rare variants may have much larger effect sizes than common variants (Bodmer, 2008), although this has yet to be shown. Rare variants are generally not tagged well through the LD structure that GWAS techniques rely upon (Bodmer, 2008), but there have been a few instances in which GWAS have detected associations between highly penetrant variants that have a MAF<3% and lipid profiles (Pollin, 2008; Sabatti, 2009). More complete imputation methods that increase the number of imputed SNPs from ~3 million using the HapMap Project to ~17 million using the 1,000 Genomes Project (Kaiser, 2008) will perhaps increase our ability to study rare variants through GWAS techniques. However, the full impact of rare variants are only likely to emerge through new technologies such as deep resequencing of candidate genes, the entire exome, and the whole genome (Out, 2009; Choi, 2009).

Missing heritability may be due to copy number variants (CNVs)

Recent research has also revealed that the genetic structural variation in the human genome is much richer than single nucleotide polymorphism and large-scale chromosomal anomalies (Wain, 2009). Submicroscopic rearrangement, duplication, insertion, or deletion of segments ranging from less than a kilobase to as large as megabases, collectively termed CNVs have been found to be surprisingly common in humans and animals (Redon, 2006; Henrichsen, 2009). These variations likely alter the

dosage of expressed genes, depending on the number of copies a person carries (Freeman, 2006). Associations between CNVs and several human diseases and traits have been documented, including AD (Rovelet-Lecrux, 2006), schizophrenia (Stefansson, 2008), and autism (Sebat, 2007; International Schizophrenia Consortium, 2008). While some CNVs can be detected through "footprints" that they leave on the results of GWAS studies (patterns of null alleles, Hardy-Weinberg disequilibrium, or Mendelian inconsistencies) (McCarroll, 2006), other technologies such as array-based comparative genomic hybridization and fosmid compared end sequence comparison are needed to fully capture the variation of CNVs in humans (Freeman, 2006). To this end, the Copy Number Variation Project is currently utilizing a variety of technologies to examine CNVs in the HapMap samples (http://www.sanger.ac.uk/humgen/cnv). The Affymetrix® Genome-Wide Human SNP Array 6.0 genotyping platform used in the genome-wide studies in this dissertation also has 202,000 probes for CNV regions from the Toronto Database of Genomic Variants as well as 744,000 CNV probes evenly spaced across the genome (Affymetrix, 2007).

The relationship between context-dependent effects and missing heritability

The discussion on missing heritability refers to "narrow sense" heritability, the proportion of total phenotypic variance due to additive genetic variance. It has been argued that context-dependent effects do not contribute to the additive genetic variance and thus are irrelevant to the issue of missing heritability (Yang, 2010). However, Cheverud and Routman have demonstrated mathematically that context-dependent effects in the form of gene-gene interaction (epistasis) can make important contributions to the average effects

of alleles and thus to additive genetic (heritable) variance (Cheverud, 1995). In a commentary on missing heritability, Moore proposed that it is likely not single variants that account for a substantial portion of the missing heritability, but rather rare combinations of interacting common variants (Eichler, 2010). Methods for exploring epistasis and gene-environment interaction in GWAS are in the early stages of development and currently consist primarily of stratified analyses and tests of homogeneity across strata to detect important environmental and genetic contexts. To this end, more sophisticated statistical methodologies to detect epistasis using a systems-biology approach have been proposed as a critical step in exploring the missing heritability, in addition to a "philosophical and analytical retooling for a complex genetic architecture" (Eichler, 2010).

Genetics of Leukoaraiosis and Cognitive Phenotypes

Genetic studies of leukoaraiosis

The manner in which genes influence the development of leukoaraiosis and its relationship to blood pressure is complex (Turner, 2000; Schmidt, 2004). Turner and Boerwinkle outline three different mechanisms of gene action on leukoaraiosis including 1) genes that influence blood pressure directly, which in turn affects leukoaraiosis, 2) genes that influence leukoaraiosis directly, independent of blood pressure (but that may be exaggerated by increased blood pressure), and 3) genes that contribute to leukoaraiosis both through pathways mediated by blood pressure and those independent of blood pressure, as shown in Figure 1.4 (Turner, 2000). The authors further suggest that involved pathways may include those responsible for vessel remodeling and growth.

Heritability of white matter hyperintensities on MRI was estimated to be 0.71 in study of male twins after adjustment for age and head size (Carmelli, 1998) and 0.55 in the Framingham Heart Study after adjustment for sex, age, age^2, and total cranial volume (Atwood, 2004). Turner et al. (2009) estimated the heritability of the logarithm of leukoaraiosis in the GENOA study participants as 0.49 in whites and 0.45 in African Americans after adjustment for age, sex, and brain volume. In an earlier publication, Turner et al. (2004) showed that leukoaraiosis has a consistently high heritability even after adjustment for blood pressure.

Turner et al. demonstrated evidence for linkage between microsattelite markers and the logarithm of leukoaraiosis adjusted for age and sex on chromosome 5 (LOD=1.91, p-value=0.0015) (Turner, 2005) and the logarithm of leukoaraiosis adjusted for age, sex, and brain volume on chromosome 11 (LOD=2.21, p-value=0.0007) and chromosome 21 (LOD=1.75, p-value=0.002) in GENOA whites (Turner, 2009). The strongest evidence for linkage in GENOA African Americans was on chromosome 22 (LOD=2.02, p-value=0.001) and on chromosome 21 at a different location than in whites (LOD=1.99, p-value=0.001) (Turner, 2009). A linkage study done by Framingham Heart Study investigators identified linkage peaks for leukoaraiosis in whites adjusted for age and the ratio of leukoaraiosis volume to brain volume on chromosome 4 (LOD=3.69) and chromosome 17 (LOD=1.78) (DeStefano, 2006).

Several candidate gene association studies have been conducted for leukoaraiosis quantity and/or progression (see Appendix 1.1 for examples). These studies have primarily concentrated on genes in pathways known to be involved in hypertension, vasculature, and endothelial damage. Polymorphisms in genes for angiotensin-converting enzyme (*ACE*) (Gormley, 2007; Schmidt, 2001) and angiotensinogen (*AGT*) (Gormley, 2007; Hassan, 2002) appear to have the strongest evidence of association, and there is preliminary evidence for matrix metalloproteinase (*MMP*) -3 and -9 (Fornage, 2007), fibrinogen (van Oijen, 2008), interleuken-6 (IL-6) (Fornage, 2008), kinesin light chain 1 (*KNS1*) (Szolnoki, 2007b), and paraoxonase (*PON1*) (Schmidt, 2000). There is also evidence for gene-gene interaction between *ACE*, methylene tetrahydrofolate reductase (*MTHFR*), and *APOE* that is associated with leukoaraiosis (Szolnoki, 2004). Although candidate gene studies have offered some encouraging initial findings, a recent systematic review and meta-analysis of 19 candidate gene polymorphisms in four of the most promising genes (*APOE*, *MTHFR*, *ACE*, and *AGT*) concluded that there is no overwhelmingly convincing evidence of association between any specific polymorphism and leukoaraiosis at this time, though the most evidence exists for the *ACE* insertion/deletion (Paternoster, 2009).

To date, only two GWAS have been conducted using MRI measures of brain aging in stroke-and dementia-free subjects. Seshadri et al. (2007a) evaluated the association between SNPs on the Affymetrix® 100K GeneChip Human Mapping 100K Set (Affymetrix, 2007) in 705 related white participants, with white matter hyperintensity volume measured as a z-score within 10-year age- and sex-specific categories of the

logarithm of leukoaraiosis using linear generalized estimating equations, family-based association tests, and linkage analysis.

The second study, a meta-analysis of GWAS in six cohorts (N=9,401 whites) conducted by the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium explored the association between genome-wide SNPs and absence or presence of infarcts visible as white matter hyperintensities on MRI that are greater than 3-4mm. This dichotomous measure differs from the quantitative leukoaraiosis trait explored in this dissertation because it excludes diffuse white matter hyperintensity that is included as part of the leukoaraiosis measurement in the GENOA analysis. However, just like diffuse leukoaraiosis, MRI infarct is a downstream consequence of the development of small vessel disease and a known risk factor for clinically relevant ischemic stroke. This study identified an intronic SNP in the *MACRO domain containing 2* (*MACROD2*) gene in the downstream region of the *fibronectin leucine-rich transmembrane protein 3* (*FLTR3*) gene. This SNP, however, did not replicate in an independent sample of 1.822 whites, but four SNPs within 200kb from the original SNP did show association with the phenotype in a sample of 644 black participants (Debette, 2010).

Genetic studies of cognitive traits

Age-related cognitive changes take place in several domains including episodic memory, working memory, inhibition and attention, executive function, and processing speed (Mattay, 2008), and there is substantial inter-individual variability in cognitive functioning and brain structure. Deary and colleagues (2004) attest that genetic variability

accounts for a significant portion of the variation in cognitive decline. The heritability of general cognitive functioning seems to be relatively high, with estimates from twin studies ranging from 55% to 80% (McGue, 2002; Finkel, 1995).

Genetic studies in humans and animals are beginning to shed light on the biological pathways that may play a role in cognitive decline. The most promising candidate genes include those that are associated with hypertension, leukoaraiosis, AD, normal cognitive functioning, cardiovascular function, oxidative stress, and inflammation (Deary, 2004), and it is likely that subsets of these genes may act in a variety of pathways to give rise to different types of dementia. A recent review by Mattay and colleagues (2008) briefly outlines the evidence for candidate genes of cognitive aging, and a representative subset of the candidate genes under consideration, in addition to those associated with hypertension and leukoaraiosis, is presented in Appendix 1.2.

There have been very few genome-wide studies conducted on cognitive phenotypes, particularly in stroke-and dementia-free subjects. In the study described above conducted by Seshadri et al. (2007a) in 705 white subjects, several measures of cognition were analyzed for evidence of association with the Affymetrix® 100K GeneChip Human Mapping 100K Set. The authors used age- and sex-standardized measures of verbal memory (VM) and attention/executive function (AEF) as primary indicators of amnestic, Alzheimer-type, and vascular cognitive dementia. Using generalized estimating equations, they found evidence of association with VM on chromosome 5 (p-value $1.1x10^{-5}$) and with AEF on chromosomes 1, 21, 11, 2, and 7 (all p-values less than $1x10^{-}$

^{5}). Family-based association tests found evidence of association with VM on chromosome 13 (p-value $3.8x10^{-5}$) and with AEF on chromosome 14 (p-value $3.8x10^{-5}$). Linkage analysis found evidence of association with AEF on chromosome 8 (LOD score=2.20).

Overview of Dissertation Research

The overall goal of this dissertation is to investigate the genetic and environmental factors that influence the individual-level variation in leukoaraiosis. Heritability estimates of leukoaraiosis range from 0.55-0.80 (Carmelli, 1998; Atwood, 2004; Turner, 2004), indicating that it is influenced by genetic factors; however, candidate gene and linkage approaches have had limited success in identifying specific polymorphisms with significant and replicated effects. As with many complex traits, the genetic contribution to leukoaraiosis is likely to involve many genetic loci with small or modest effects that may be acting alone or through interactions with environmental factors. To develop a more comprehensive understanding of the genetic architecture this trait, three different approaches have been used to examine the effects of genetic variants on leukaroaraiosis and cognitive function.

In the first study (Chapter 2), we utilize a candidate gene approach to understand the contribution of single gene effects on mean levels of leukoaraiosis as well as gene-risk factor and gene-gene interactions. In the second study (Chapter 3), we investigate genetic variation across the entire genome to identify chromosomal regions associated with additive effects on leukoaraiosis. In the final study (Chapter 4), we investigate the genetic

correlations among leukoaraiosis and seven measures of cognitive function by using both measured (GWAS) and unmeasured (biometrical) genetic approaches. Each of these approaches assesses different aspects of the genetic architecture of leukoaraiosis, and they combine to offer a broad perspective on the different ways in which genetic factors may be influencing leukoaraiosis and cognitive function.

Study population

The GENOA study

The study population for this dissertation research is comprised of whites and African Americans from the Genetic Epidemiology Network of Arteriopathy (GENOA) study. The National Heart, Lung and Blood Institute established the Family Blood Pressure Program (FBPP) in 1996 from four existing research networks that were investigating the genetics of hypertension and its sequelae (FBPP Investigators, 2002), including GENOA. GENOA recruited hypertensive sibships from Rochester, Minnesota and Jackson, Mississippi for linkage and association studies to investigate the genetic underpinnings of hypertension and target organ damage related to hypertension (Daniels, 2004).

In the initial phase of the GENOA study (Phase I: 1996-2001), all members of sibships containing ≥ 2 individuals with early-onset essential hypertension clinically diagnosed before age 60 were invited to participate, including both hypertensive and normotensive siblings (1,583 non-Hispanic whites and 1,841 African Americans). The diagnosis of essential hypertension was established based on blood pressure levels measured at the study visit (>140 mmHg average systolic blood pressure or >90 mmHg average diastolic

blood pressure) or a prior diagnosis of hypertension and current treatment with antihypertensive medications. Exclusion criteria were secondary hypertension, alcoholism or drug abuse, pregnancy, insulin-dependent diabetes mellitus, or active malignancy. In the second phase of the GENOA study (Phase II: 2000-2004), 1,241 white and 1,482 African American participants were successfully re-recruited to measure potential target organ damage due to hypertension. Phase I and II GENOA data consist of demographic information, medical history, clinical characteristics, lifestyle factors, and blood samples for genotyping and biomarker assays. Written informed consent was obtained from all subjects and approval was granted by participating institutional review boards. All reported phenotype and covariate data used for this dissertation was collected during the Phase II exam.

The GMBI ancillary study of GENOA

The Genetics of Microangiopathic Brain Injury (GMBI) study (2001-2006) is an ancillary study of GENOA undertaken to investigate susceptibility genes for ischemic brain injury. Phase II GENOA participants that had a sibling willing and eligible to participate in the GMBI study underwent a neurocognitive testing battery to assess several domains of cognitive function including learning, memory, attention, concentration, and language (967 whites and 1,010 African Americans). Ischemic brain damage to the subcortical and periventricular white matter (leukoaraiosis) was quantified by MRI in subjects who had no history of stroke or neurological disease and no implanted metal devices (916 whites and 830 African Americans).

Measurement of leukoaraiosis in the GMBI study

The MRI methods used to quantify leukoaraiosis volume in GMBI participants are described in detail in subsequent chapters. Briefly, leukoaraiosis volumes were determined from axial fluid-attenuated inversion recovery (FLAIR) images, which are T2-weighted images with the signal of the cerebrospinal fluid nulled such that brain pathology appears as the brightest intracranial tissue. A fully automated algorithm was used to segment each slice of the edited multi-slice FLAIR sequence into voxels assigned to one of three categories: brain, cerebrospinal fluid, or leukoaraiosis.

While this method represents the state of the art in technology for quantifying leukoaraiosis at the time that the MRIs were performed, the difficulties of distinguishing brain tissue affected by leukoaraiosis via imaging results in notable degree of imprecision in the measurement of leukoaraiosis volume (mean absolute error = 6.6% and mean test-retest coefficient of variation = 1.4%) (Jack, 2001). In addition, there may be areas of brain tissue affected by chronic ischemia that are undetectable by MRI because the severity of ischemia in those areas has not progressed enough to be captured by leukoaraiosis measurement techniques. However, these areas of ischemia are not considered to be areas of leukoaraiosis, since leukoaraiosis by definition is "tissue that appears as bright white spots on MRI." Thus, the volume of leukoaraiosis that a person has is defined inherently by the MRI technique that is performed, and this volume will vary with the sensitivity of the MRI technique. However, the relative amount of leukoaraiosis across individuals should remain reasonably constant when all individuals are assessed using the same standardized MRI procedure.

Conclusion

The three research projects presented in this dissertation utilize an integrative approach that combines innovative statistical techniques with methods in genetic epidemiology to investigate the diverse mechanisms of genetic effects on leukoaraiosis and the shared genetic effects among leukoaraiosis and measures of cognitive function. Only a handful of genetic studies have investigated asymptomatic subclinical brain phenotypes due to target organ damage from hypertension rather than the clinical endpoints of stroke and dementia, which are heterogeneous disease endpoints that are also etiologically associated with pathways unrelated to hypertension. However, a greater understanding of the underlying genetic architecture of leukoaraiosis itself has the potential to provide insight into the etiological processes that leads to stroke and dementia and to assist in earlier identification of individuals at increased risk for disease, the development of more efficacious treatments, and the tailoring of particular treatments to people most likely to respond positively.

The research strategies reflected in this dissertation represent the beginning arc of research in the post-Human Genome Project era. Just as candidate gene studies (Chapter 2) gave way to genome-wide studies (Chapter 3), there is now a growing desire to begin to integrate what we know from across the many different traits that have been studied (Chapter 4). Furture studies that build upon current knowledge of the genetic architecture of leukoaraiosis must begin to utilize systems-level approaches to integrate information from next-generation sequencing, epigenomic assays, and transcriptional profiles with the

candidate gene and genome-wide analyses presented here in order to gain a more thorough understanding of how genetic factors affect leukoaraiosis development and progression.

Figure 1.1. Arteriosclerosis and target organ damage

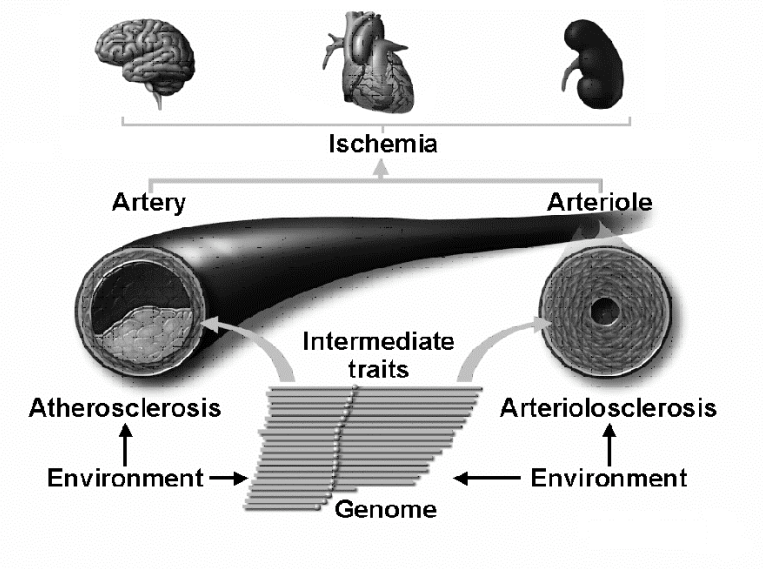
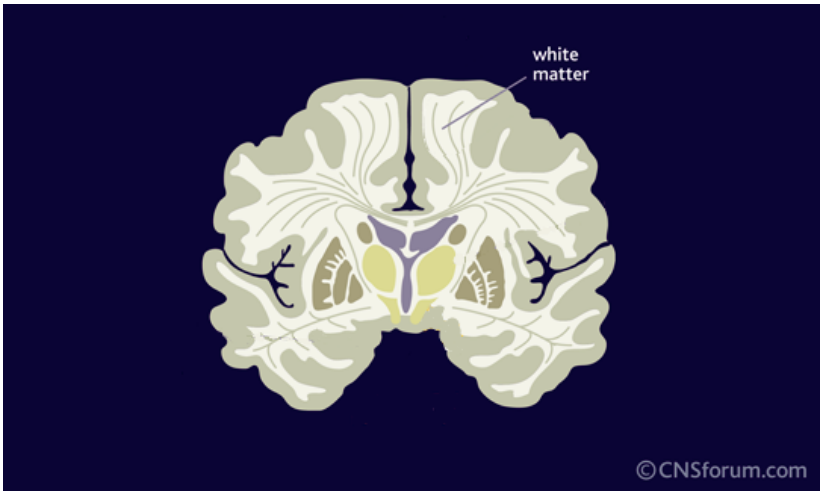
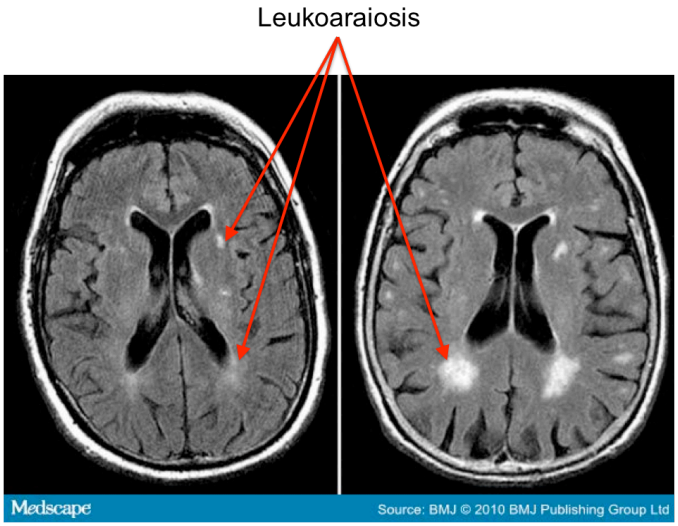


Figure 1.2. The location of white matter in the brain



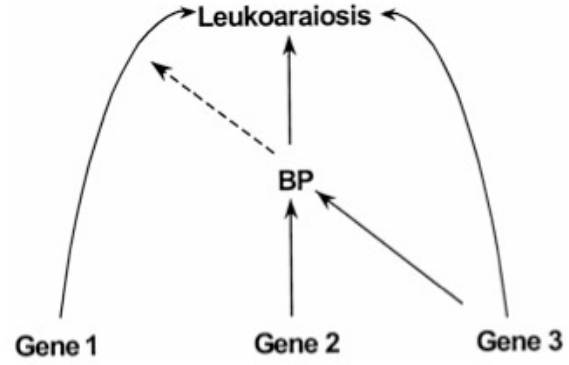
Photograph courtesy of CNSforum (www.CNSforum.com)

Figure 1.3. Leukoaraiosis on brain MRI



Adapted from Debette (2010)

Figure 1.4. Genetic contributions to leukoaraiosis



Adapted from Turner and Boerwinkle (2000)

Appendix 1.1. Summary of selected candidate-gene association studies for leukoaraiosis

Gene	Phenotype	Population	Polymorphism	Reference
Angiotensinogen (AGT)	Ischemic leukoaraiosis subtype of small vessel disease	Hypertensives from small vessel disease cases and community controls	-20 A->C	Gormley, 2007
	White matter lesion progression	Community sample 50-75	M235T	Schmidt, 2001
Angiotensinogen converting enzyme (ACE)	Ischemic leukoaraiosis subtype of small vessel disease	Small vessel disease cases and community controls	No association found	Gormley, 2007
	Presence of ischemic leukoaraiosis	Patients presenting with classic lacunar syndromes	Insertion/ Deletion	Hassan, 2002
	White matter lesions	Memory clinic patients	Insertion/ Deletion	Amar, 1998
Apolipoprotein E (APOE)	White matter hyperintensity grading and area	Subjects 50-75 without neuropsychiatric disease	No association with e4 carriers	Schmidt, 1996
	White matter lesions	Memory clinic patients	No association found	Amar, 1998
	White matter lesions	Elderly depressed cases and nondepressed elderly controls	No association with e4 carriers	Steffens, 2003
C-Reactive Protein (CRP)	White matter lesions	White subjects over age 60	No association found	Fornage, 2008
	White matter lesions	Black subjects over age 60	No association found	Fornage, 2008
	White matter lesions	Population-based subjects 60-90	No association found	Reitz, 2007
Endothelin-1 and Recptors (ET1, ETA and ETB)	Lacunar infarct with leukoaraiosis	Small vessel disease patients and community controls	No association found	Gormley, 2005
Fibrinogen Alpha/Gamma (FGA, FGG)	White matter lesions	Elderly patients undergoing MRI	Haplotype GGCGATA	van Oijen, 2008

Gene	Phenotype	Population	Polymorphism	Reference
Interleuken-6 (IL6)	White matter lesions	White subjects over age 60	-174G/C haplotype	Fornage, 2008
	White matter lesions	Black subjects over age 60	No association found	Fornage, 2008
Kinesin Light-Chain 1 (KNS2)	Leukoaraiosis	Hypertensive smokers with leukoaraiosis (cases) and neuroimaging alteration-free (controls)	G56836C	Szolnoki, 2007b
Nitric Oxide Synthetase (eNOS)	Ischemic leukoaraiosis	Small vessel disease cases and community controls	No association found	Hassan, 2004a
Paraoxonase (PON1)	White matter lesions and progression	Subjects 44-75 without neuropsychiatric disease	M54L	Schmidt, 2000

Appendix 1.2. Summary of evidence of candidate genes for cognitive aging

Gene	Region	Pathway/Disease	Finding
KLOTHO	13q12	Oxidative Stress (antioxidative function)	SNPs affect cognitive aging in mice
Prion protein gene (PRNP)	20pter-p12	Oxidative Stress (antioxidative function)	SNP association with long-term memory and cognitive impairment in humans
Insulin-like growth factor (IGF)	15q26.3	Apoptosis	Linked to longevity in humans and cognitive function in rats
Death-associated protein kinase 1 (DAPK1)	9q34.1	Apoptosis	Linked to late-onset AD in humans as well as cognitive function in rats
Interleukin-1B (IL-1B)	2q14	Inflammation	SNPs associated with verbal memory in elderly humans
Tumor necrosis factor alpha (TNF-a)	6p21.3	Inflammation	SNPs associated with processing speed in elderly humans
Interleukin-1B-converting enzyme	11q22	Inflammation	SNPs associated with executive function in elderly humans
Catecholamine genes (e.g. catechol-O-methyl transferase (COMT))	22q11.21-11.23	Cognition/Memory	SNPs associated with signal to noise ratio in information processing in humans
Serotonin genes (e.g. seratonin transporter (SLC6A4))	17q11.1-q12	Cognition/Memory	VNTRs associated with speed of cognitive decline in humans
Trophic genes (e.g. BDNF)	11p13	Cognition/Memory	SNPs associated with altered hippocampal function during memory tasks in humans
KIBRA	5q35.1	Cognition/Memory	SNPs associated with altered hippocampal function during memory tasks in humans
Glutamate receptor, metabotropic (GRM3)	7q21.1-q21.2	Cognition/Memory	SNPs associated with differences in verbal and memory scores in humans
Disrupted in Schizophrenia (DISC1)	1q42.1	Cognition/Memory	SNPs associated with differences in cognitive ability scores in humans

Adapted from Mattay (2008)

References

Adams, H.P.,Jr, Bendixen, B.H., Kappelle, L.J., Biller, J., Love, B.B., Gordon, D.L., and Marsh, E.E.,3rd. (1993). Classification of Subtype of Acute Ischemic Stroke. Definitions for use in a Multicenter Clinical Trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke* **24(1):** 35-41.

Affymetrix. (2007). Data Sheet: Affymetrix® Genome-Wide Human SNP Array 6.0.

Ahituv, N., Kavaslar, N., Schackwitz, W., Ustaszewska, A., Martin, J., Hebert, S., Doelle, H., Ersoy, B., Kryukov, G., Schmidt, S. et al. (2007). Medical Sequencing at the Extremes of Human Body Mass. *Am J Hum Genet* **80(4):** 779-791.

Altshuler, D., Daly, M.J., and Lander, E.S. (2008). Genetic Mapping in Human Disease. *Science* **322(5903):** 881-888.

Alzheimer's Association. (2008). Alzheimer's Disease Facts and Figures 2008: 10 Million U.S. Baby Boomers Will Develop Alzheimer's Disease. **4(2):** 110-133.

Alzheimer's Association. (2010). Alzheimer's Disease Facts and Figures 2010. *Alzheimer's and Dementia* **6(2):** 158-194.

Amar, K., MacGowan, S., Wilcock, G., Lewis, T., and Scott, M. (1998). Are Genetic Factors Important in the Aetiology of Leukoaraiosis? Results from a Memory Clinic Population. *Int J Geriatr Psychiatry* **13(9):** 585-590.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders.* Washington, D.C.

Arnett, D.K., de las Fuentes, L., and Broeckel, U. (2004). Genes for Left Ventricular Hypertrophy. *Curr Hypertens Rep* **6(1):** 36-41.

Atwood, L.D., Wolf, P.A., Heard-Costa, N.L., Massaro, J.M., Beiser, A., D'Agostino, R.B., and DeCarli, C. (2004). Genetic Variation in White Matter Hyperintensity Volume in the Framingham Study. *Stroke* **35(7):** 1609-1613.

Barrett, J.C., Hansoul, S., Nicolae, D.L., Cho, J.H., Duerr, R.H., Rioux, J.D., Brant, S.R., Silverberg, M.S., Taylor, K.D., Barmada, M.M. et al. (2008). Genome-Wide Association Defines More than 30 Distinct Susceptibility Loci for Crohn's Disease. *Nat Genet* **40(8):** 955-962.

Bielak, L.F., Turner, S.T., Franklin, S.S., Sheedy, P.F.,2nd, and Peyser, P.A. (2004). Age-Dependent Associations between Blood Pressure and Coronary Artery Calcification in Asymptomatic Adults. *J Hypertens* **22(4):** 719-725.

Bielak, L.F., Rumberger, J.A., Sheedy, P.F.,2nd, Schwartz, R.S., and Peyser, P.A. (2000). Probabilistic Model for Prediction of Angiographically Defined Obstructive Coronary Artery Disease using Electron Beam Computed Tomography Calcium Score Strata. *Circulation* **102(4):** 380-385.

Bodmer, W., and Bonilla, C. (2008). Common and Rare Variants in Multifactorial Susceptibility to Common Diseases. *Nat Genet* **40(6):** 695-701.

Bonita, R., and Beaglehole, R. (1986). Does Treatment of Hypertension Explain the Decline in Mortality from Stroke? *Br Med J (Clin Res Ed)* **292(6514):** 191-192.

Bornstein, N., Silvestrelli, G., Caso, V., and Parnetti, L. (2006). Arterial Hypertension and Stroke Prevention: An Update. *Clin Exp Hypertens* **28(3-4):** 317-326.

Boustani, M., Peterson, B., Hanson, L., Harris, R., Lohr, K.N., and U.S. Preventive Services Task Force. (2003). Screening for Dementia in Primary Care: A Summary of the Evidence for the U.S. Preventive Services Task Force. *Ann Intern Med* **138(11):** 927-937.

Breteler, M.M. (2000). Vascular Risk Factors for Alzheimer's Disease: An Epidemiologic Perspective. *Neurobiol Aging* **21(2):** 153-160.

Carmelli, D., DeCarli, C., Swan, G.E., Jack, L.M., Reed, T., Wolf, P.A., and Miller, B.L. (1998). Evidence for Genetic Variance in White Matter Hyperintensity Volume in Normal Elderly Male Twins. *Stroke* **29(6):** 1177-1181.

Cheverud, J.M., Ehrich, T.H., Vaughn, T.T., Koreishi, S.F., Linsey, R.B., and Pletscher, L.S. (2004). Pleiotropic Effects on Mandibular Morphology II: Differential Epistasis and Genetic Variation in Morphological Integration. *J Exp Zool B Mol Dev Evol* **302(5):** 424-435.

Cheverud, J.M., and Routman, E.J. (1995). Epistasis and its Contribution to Genetic Variance Components. *Genetics* **139(3):** 1455-1461.

Choi, M., Scholl, U.I., Ji, W., Liu, T., Tikhonova, I.R., Zumbo, P., Nayir, A., Bakkaloglu, A., Ozen, S., Sanjad, S. et al. (2009). Genetic Diagnosis by Whole Exome Capture and Massively Parallel DNA Sequencing. *Proc Natl Acad Sci U S A* **106(45):** 19096-19101.

Cohen, J., Pertsemlidis, A., Kotowski, I.K., Graham, R., Garcia, C.K., and Hobbs, H.H. (2005). Low LDL Cholesterol in Individuals of African Descent Resulting from Frequent Nonsense Mutations in PCSK9. *Nat Genet* **37(2):** 161-165.

Crum, R.M., Anthony, J.C., Bassett, S.S., and Folstein, M.F. (1993). Population-Based Norms for the Mini-Mental State Examination by Age and Educational Level. *JAMA* **269(18):** 2386-2391.

Cubrilo-Turek, M. (2004). Stroke Risk Factors: Recent Evidence and New Aspects. *Inter Congress Ser* **1262:** 466-469.

Daniels, P.R., Kardia, S.L., Hanis, C.L., Brown, C.A., Hutchinson, R., Boerwinkle, E., Turner, S.T., and Genetic Epidemiology Network of Arteriopathy study. (2004). Familial Aggregation of Hypertension Treatment and Control in the Genetic Epidemiology Network of Arteriopathy (GENOA) Study. *Am J Med* **116(10):** 676-681.

De Groot, J.C., De Leeuw, F.E., Oudkerk, M., Van Gijn, J., Hofman, A., Jolles, J., and Breteler, M.M. (2002). Periventricular Cerebral White Matter Lesions Predict Rate of Cognitive Decline. *Ann Neurol* **52(3):** 335-341.

de Leeuw, F.E., de Groot, J.C., Achten, E., Oudkerk, M., Ramos, L.M., Heijboer, R., Hofman, A., Jolles, J., van Gijn, J., and Breteler, M.M. (2001). Prevalence of Cerebral White Matter Lesions in Elderly People: A Population Based Magnetic Resonance Imaging Study. the Rotterdam Scan Study. *J Neurol Neurosurg Psychiatry* **70(1):** 9-14.

Deary, I.J., Wright, A.F., Harris, S.E., Whalley, L.J., and Starr, J.M. (2004). Searching for Genetic Influences on Normal Cognitive Ageing. *Trends Cogn Sci* **8(4):** 178-184.

Debette, S., Bis, J.C., Fornage, M., Schmidt, H., Ikram, M.A., Sigurdsson, S., Heiss, G., Struchalin, M., Smith, A.V., van der Lugt, A. et al. (2010). Genome-Wide Association Studies of MRI-Defined Brain Infarcts: Meta-Analysis from the CHARGE Consortium. *Stroke* **41(2):** 210-217.

DeStefano, A.L., Atwood, L.D., Massaro, J.M., Heard-Costa, N., Beiser, A., Au, R., Wolf, P.A., and DeCarli, C. (2006). Genome-Wide Scan for White Matter Hyperintensity: The Framingham Heart Study. *Stroke* **37(1):** 77-81.

Dickstein, D.L., Walsh, J., Brautigam, H., Stockton, S.D.,Jr, Gandy, S., and Hof, P.R. (2010). Role of Vascular Risk Factors and Vascular Dysfunction in Alzheimer's Disease. *Mt Sinai J Med* **77(1):** 82-102.

Donnan, G.A., Fisher, M., Macleod, M., and Davis, S.M. (2008). Stroke. *Lancet* **371(9624):** 1612-1623.

Donnelly, P. (2008). Progress and Challenges in Genome-Wide Association Studies in Humans. *Nature* **456(7223):** 728-731.

Eichler, E.E., Flint, J., Gibson, G., Kong, A., Leal, S.M., Moore, J.H., and Nadeau, J.H. (2010). Missing Heritability and Strategies for Finding the Underlying Causes of Complex Disease. *Nat Rev Genet* **11(6):** 446-450.

Elias, M.F., Wolf, P.A., D'Agostino, R.B., Cobb, J., and White, L.R. (1993). Untreated Blood Pressure Level is Inversely Related to Cognitive Functioning: The Framingham Study. *Am J Epidemiol* **138(6):** 353-364.

Essink-Bot, M.L., Pereira, J., Packer, C., Schwarzinger, M., and Burstrom, K. (2002). Cross-National Comparability of Burden of Disease Estimates: The European Disability Weights Project. *Bull World Health Organ* **80(8):** 644-652.

Falconer, D.S., and Mackay, T.F. (1996). *Introduction to Quantitative Genetics.* Pearson Prentice Hall: England.

Fearnhead, N.S., Winney, B., and Bodmer, W.F. (2005). Rare Variant Hypothesis for Multifactorial Inheritance: Susceptibility to Colorectal Adenomas as a Model. *Cell Cycle* **4(4):** 521-525.

FBPP Investigators. (2002). Multi-Center Genetic Study of Hypertension: The Family Blood Pressure Program (FBPP). *Hypertension* **39(1):** 3-9.

Finkel, D., Pedersen, N.L., McGue, M., and McClearn, G.E. (1995). Heritability of Cognitive Abilities in Adult Twins: Comparison of Minnesota and Swedish Data. *Behav Genet* **25(5):** 421-431.

Fornage, M., Chiang, Y.A., O'Meara, E.S., Psaty, B.M., Reiner, A.P., Siscovick, D.S., Tracy, R.P., and Longstreth, W.T.,Jr. (2008). Biomarkers of Inflammation and MRI-Defined Small Vessel Disease of the Brain: The Cardiovascular Health Study. *Stroke* **39(7):** 1952-1959.

Fornage, M., Mosley, T.H., Jack, C.R., de Andrade, M., Kardia, S.L., Boerwinkle, E., and Turner, S.T. (2007). Family-Based Association Study of Matrix Metalloproteinase-3 and -9 Haplotypes with Susceptibility to Ischemic White Matter Injury. *Hum Genet* **120(5):** 671-680.

Frazer, K.A., Murray, S.S., Schork, N.J., and Topol, E.J. (2009). Human Genetic Variation and its Contribution to Complex Traits. *Nat Rev Genet* **10(4):** 241-251.

Freeman, J.L., Perry, G.H., Feuk, L., Redon, R., McCarroll, S.A., Altshuler, D.M., Aburatani, H., Jones, K.W., Tyler-Smith, C., Hurles, M.E. et al. (2006). Copy Number Variation: New Insights in Genome Diversity. *Genome Res* **16(8):** 949-961.

Fu, J.H., Lu, C.Z., Hong, Z., Dong, Q., Luo, Y., and Wong, K.S. (2005). Extent of White Matter Lesions is Related to Acute Subcortical Infarcts and Predicts further Stroke Risk in Patients with First Ever Ischaemic Stroke. *J Neurol Neurosurg Psychiatry* **76(6):** 793-796.

Geldmacher, D.S., and Whitehouse, P.J.,Jr. (1997). Differential Diagnosis of Alzheimer's Disease. *Neurology* **48(5 Suppl 6):** S2-9.

Goldstein, L.B., Adams, R., Becker, K., Furberg, C.D., Gorelick, P.B., Hademenos, G., Hill, M., Howard, G., Howard, V.J., Jacobs, B. et al. (2001). Primary Prevention of

Ischemic Stroke: A Statement for Healthcare Professionals from the Stroke Council of the American Heart Association. *Stroke* **32(1):** 280-299.

Gormley, K., Bevan, S., and Markus, H.S. (2007). Polymorphisms in Genes of the Renin-Angiotensin System and Cerebral Small Vessel Disease. *Cerebrovasc Dis* **23(2-3):** 148-155.

Gormley, K., Bevan, S., Hassan, A., and Markus, H.S. (2005). Polymorphisms in Genes of the Endothelin System and Cerebral Small-Vessel Disease. *Stroke* **36(8):** 1656-1660.

Gudmundsson, J., Sulem, P., Steinthorsdottir, V., Bergthorsson, J.T., Thorleifsson, G., Manolescu, A., Rafnar, T., Gudbjartsson, D., Agnarsson, B.A., Baker, A. et al. (2007). Two Variants on Chromosome 17 Confer Prostate Cancer Risk, and the One in TCF2 Protects Against Type 2 Diabetes. *Nat Genet* **39(8):** 977-983.

Haan, M.N., and Wallace, R. (2004). Can Dementia be Prevented? Brain Aging in a Population-Based Context. *Annu Rev Public Health* **25:** 1-24.

Hajat, C., Tilling, K., Stewart, J.A., Lemic-Stojcevic, N., and Wolfe, C.D. (2004). Ethnic Differences in Risk Factors for Ischemic Stroke: A European Case-Control Study. *Stroke* **35(7):** 1562-1567.

Hassan, A., Gormley, K., O'Sullivan, M., Knight, J., Sham, P., Vallance, P., Bamford, J., and Markus, H. (2004a). Endothelial Nitric Oxide Gene Haplotypes and Risk of Cerebral Small-Vessel Disease. *Stroke* **35(3):** 654-659.

Hassan, A., Hunt, B.J., O'Sullivan, M., Bell, R., D'Souza, R., Jeffery, S., Bamford, J.M., and Markus, H.S. (2004b). Homocysteine is a Risk Factor for Cerebral Small Vessel Disease, Acting Via Endothelial Dysfunction. *Brain* **127(Pt 1):** 212-219.

Hassan, A., Hunt, B.J., O'Sullivan, M., Parmar, K., Bamford, J.M., Briley, D., Brown, M.M., Thomas, D.J., and Markus, H.S. (2003). Markers of Endothelial Dysfunction in Lacunar Infarction and Ischaemic Leukoaraiosis. *Brain* **126(Pt 2):** 424-432.

Hassan, A., Lansbury, A., Catto, A.J., Guthrie, A., Spencer, J., Craven, C., Grant, P.J., and Bamford, J.M. (2002). Angiotensin Converting Enzyme insertion/deletion Genotype is Associated with Leukoaraiosis in Lacunar Syndromes. *J Neurol Neurosurg Psychiatry* **72(3):** 343-346.

Henrichsen, C.N., Chaignat, E., and Reymond, A. (2009). Copy Number Variants, Diseases and Gene Expression. *Hum Mol Genet* **18(R1):** R1-R8.

Hertz, R.P., Unger, A.N., Cornell, J.A., and Saunders, E. (2005). Racial Disparities in Hypertension Prevalence, Awareness, and Management. *Arch Intern Med* **165:** 2098-2104.

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(23):** 9362-9367.

Hirschhorn, J.N., and Daly, M.J. (2005). Genome-Wide Association Studies for Common Diseases and Complex Traits. *Nat Rev Genet* **6(2):** 95-108.

Hirschhorn, J.N., Lohmueller, K., Byrne, E., and Hirschhorn, K. (2002). A Comprehensive Review of Genetic Association Studies. *Genet Med* **4(2):** 45-61.

Hodgkin, J. (1998). Seven Types of Pleiotropy. *Int J Dev Biol* **42(3):** 501-505.

Hoh, J., and Ott, J. (2001). A Train of Thoughts on Gene Mapping. *Theor Popul Biol* **60(3):** 149-153.

International HapMap Consortium. (2005). A Haplotype Map of the Human Genome. *Nature* **437(7063):** 1299-1320.

International Schizophrenia Consortium. (2008). Rare Chromosomal Deletions and Duplications Increase Risk of Schizophrenia. *Nature* **455(7210):** 237-241.

Inzitari, D. (2003). Leukoaraiosis: An Independent Risk Factor for Stroke? *Stroke* **34(8):** 2067-2071.

Jack, C.R.,Jr, O'Brien, P.C., Rettman, D.W., Shiung, M.M., Xu, Y., Muthupillai, R., Manduca, A., Avula, R., and Erickson, B.J. (2001). FLAIR Histogram Segmentation for Measurement of Leukoaraiosis Volume. *J Magn Reson Imaging* **14(6):** 668-676.

Jeerakathil, T., Wolf, P.A., Beiser, A., Massaro, J., Seshadri, S., D'Agostino, R.B., and DeCarli, C. (2004). Stroke Risk Profile Predicts White Matter Hyperintensity Volume: The Framingham Study. *Stroke* **35(8):** 1857-1861.

Jellinger, K.A. (2008). Morphologic Diagnosis of "Vascular Dementia" - a Critical Update. *J Neurol Sci* **270(1-2):** 1-12.

Ji, W., Foo, J.N., O'Roak, B.J., Zhao, H., Larson, M.G., Simon, D.B., Newton-Cheh, C., State, M.W., Levy, D., and Lifton, R.P. (2008). Rare Independent Mutations in Renal Salt Handling Genes Contribute to Blood Pressure Variation. *Nat Genet* **40(5):** 592-599.

Johnson, A.D., and O'Donnell, C.J. (2009). An Open Access Database of Genome-Wide Association Results. *BMC Med Genet* **10:** 6.

Kaiser, J. (2008). DNA Sequencing. A Plan to Capture Human Diversity in 1000 Genomes. *Science* **319(5862):** 395.

Kannel, W.B. (1995). Framingham Study Insights into Hypertensive Risk of Cardiovascular Disease. *Hypertens Res* **18(3):** 181-196.

Karasik, D., Hsu, Y.H., Zhou, Y., Cupples, L.A., Kiel, D.P., and Demissie, S. (2010). Genome-Wide Pleiotropy of Osteoporosis-Related Phenotypes: The Framingham Study. *J Bone Miner Res* **25(7):** 1555-1563.

Kardia, S.L., Greene, M.T., Boerwinkle, E., Turner, S.T., and Kullo, I.J. (2008). Investigating the Complex Genetic Architecture of Ankle-Brachial Index, a Measure of Peripheral Arterial Disease, in Non-Hispanic Whites. *BMC Med Genomics* **1:** 16.

Kardia, S.L. (2000). Context-Dependent Genetic Effects in Hypertension. *Curr Hypertens Rep* **2(1):** 32-38.

Kempthorne, O. (1954). The Correlation between Relatives in a Random Mating Population. *Proceedings of the Royal Society, B* **143:** 103-113.

Klag, M.J., Whelton, P.K., Randall, B.L., Neaton, J.D., Brancati, F.L., Ford, C.E., Shulman, N.B., and Stamler, J. (1996). Blood Pressure and End-Stage Renal Disease in Men. *N Engl J Med* **334(1):** 13-18.

Knopman, D., Boland, L.L., Mosley, T., Howard, G., Liao, D., Szklo, M., McGovern, P., Folsom, A.R., and Atherosclerosis Risk in Communities (ARIC) Study Investigators. (2001). Cardiovascular Risk Factors and Cognitive Decline in Middle-Aged Adults. *Neurology* **56(1):** 42-48.

Kong, A., Steinthorsdottir, V., Masson, G., Thorleifsson, G., Sulem, P., Besenbacher, S., Jonasdottir, A., Sigurdsson, A., Kristinsson, K.T., Jonasdottir, A. et al. (2009). Parental Origin of Sequence Variants Associated with Complex Diseases. *Nature* **462(7275):** 868-874.

Koren, M.J., Devereux, R.B., Casale, P.N., Savage, D.D., and Laragh, J.H. (1991). Relation of Left Ventricular Mass and Geometry to Morbidity and Mortality in Uncomplicated Essential Hypertension. *Ann Intern Med* **114(5):** 345-352.

Kraft, P., Yen, Y.C., Stram, D.O., Morrison, J., and Gauderman, W.J. (2007). Exploiting Gene-Environment Interaction to Detect Genetic Associations. *Hum Hered* **63(2):** 111-119.

Kukull, W.A., and Bowen, J.D. (2002). Dementia Epidemiology. *Med Clin North Am* **86(3):** 573-590.

Kuller, L.H., Lopez, O.L., Jagust, W.J., Becker, J.T., DeKosky, S.T., Lyketsos, C., Kawas, C., Breitner, J.C., Fitzpatrick, A., and Dulberg, C. (2005). Determinants of Vascular Dementia in the Cardiovascular Health Cognition Study. *Neurology* **64(9):** 1548-1552.

Kuller, L.H., Longstreth, W.T.,Jr, Arnold, A.M., Bernick, C., Bryan, R.N., Beauchamp, N.J.,Jr, and Cardiovascular Health Study Collaborative Research Group. (2004). White Matter Hyperintensity on Cranial Magnetic Resonance Imaging: A Predictor of Stroke. *Stroke* **35(8):** 1821-1825.

Launer, L.J., Ross, G.W., Petrovitch, H., Masaki, K., Foley, D., White, L.R., and Havlik, R.J. (2000). Midlife Blood Pressure and Dementia: The Honolulu-Asia Aging Study. *Neurobiol Aging* **21(1):** 49-55.

Lehericy, S., Marjanska, M., Mesrob, L., Sarazin, M., and Kinkingnehun, S. (2007). Magnetic Resonance Imaging of Alzheimer's Disease. *Eur Radiol* **17(2):** 347-362.

Lezak, M. (1995). *Neuropsychological assessment.* Oxford University Press: New York.

Li, Y., and Abecasis, G.R. (2006). Mach 1.0: Rapid Haplotype Reconstruction and Missing Genotype Inference. *American Journal of Human Genetics* **(S79):** 2290.

Liao, D., Cooper, L., Cai, J., Toole, J.F., Bryan, N.R., Hutchinson, R.G., and Tyroler, H.A. (1996). Presence and Severity of Cerebral White Matter Lesions and Hypertension, its Treatment, and its Control. the ARIC Study. Atherosclerosis Risk in Communities Study. *Stroke* **27(12):** 2262-2270.

Lobo, A., Launer, L.J., Fratiglioni, L., Andersen, K., Di Carlo, A., Breteler, M.M., Copeland, J.R., Dartigues, J.F., Jagger, C., Martinez-Lage, J. et al. (2000). Prevalence of Dementia and Major Subtypes in Europe: A Collaborative Study of Population-Based Cohorts. Neurologic Diseases in the Elderly Research Group. *Neurology* **54(11 Suppl 5):** S4-9.

Lopez, A.D., Mathers, C.D., Ezzati, M., Jamison, D.T., and Murray, C.J. (2001). Global and Regional Burden of Disease and Risk Factors, 2001: Systematic Analysis of Population Health Data. *The Lancet,* **367(9524):** 1747-1757.

Mackay, T.F. (2010). Mutations and Quantitative Genetic Variation: Lessons from Drosophila. *Philos Trans R Soc Lond B Biol Sci* **365(1544):** 1229-1239.

Mackay, T.F. (2009). The Genetic Architecture of Complex Behaviors: Lessons from Drosophila. *Genetica* **136(2):** 295-302.

MacMahon, S., Peto, R., Cutler, J., Collins, R., Sorlie, P., Neaton, J., Abbott, R., Godwin, J., Dyer, A., and Stamler, J. (1990). Blood Pressure, Stroke, and Coronary Heart Disease. Part 1, Prolonged Differences in Blood Pressure: Prospective Observational Studies Corrected for the Regression Dilution Bias. *Lancet* **335(8692):** 765-774.

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. et al. (2009). Finding the Missing Heritability of Complex Diseases. *Nature* **461(7265):** 747-753.

Marieb, E.N. (1998). *Human Anatomy and Physiology.* Benjamin/Cummings Publishing Company: Menlo Park, CA.

Markus, H.S. (2008). Genes, Endothelial Function and Cerebral Small Vessel Disease in Man. *Exp Physiol* **93(1):** 121-127.

Markus, H.S., Hunt, B., Palmer, K., Enzinger, C., Schmidt, H., and Schmidt, R. (2005). Markers of Endothelial and Hemostatic Activation and Progression of Cerebral White Matter Hyperintensities: Longitudinal Results of the Austrian Stroke Prevention Study. *Stroke* **36(7):** 1410-1414.

Mattay, V.S., Goldberg, T.E., Sambataro, F., and Weinberger, D.R. (2008). Neurobiology of Cognitive Aging: Insights from Imaging Genetics. *Biol Psychol* **79(1):** 9-22.

McCarroll, S.A., Hadnott, T.N., Perry, G.H., Sabeti, P.C., Zody, M.C., Barrett, J.C., Dallaire, S., Gabriel, S.B., Lee, C., Daly, M.J. et al. (2006). Common Deletion Polymorphisms in the Human Genome. *Nat Genet* **38(1):** 86-92.

McGue, M., and Christensen, K. (2002). The Heritability of Level and Rate-of-Change in Cognitive Functioning in Danish Twins Aged 70 Years and Older. *Exp Aging Res* **28(4):** 435-451.

Morton, N.E. (1955). Sequential Tests for the Detection of Linkage. *Am J Hum Genet* **7(3):** 277-318.

Mosley, T.H.,Jr, Knopman, D.S., Catellier, D.J., Bryan, N., Hutchinson, R.G., Grothues, C.A., Folsom, A.R., Cooper, L.S., Burke, G.L., Liao, D. et al. (2005). Cerebral MRI Findings and Cognitive Functioning: The Atherosclerosis Risk in Communities Study. *Neurology* **64(12):** 2056-2062.

Murray, C.J., and Lopez, A.D. (1997). Mortality by Cause for Eight Regions of the World: Global Burden of Disease Study. *The Lancet,* **349(9061):** 1269-1276.

National Kidney Foundation. (2002). K/DOQI Clinical Practice Guidelines for Chronic Kidney Disease: Evaluation, Classification, and Stratification. *Am J Kidney Dis* **39:** S1-S266.

NCI-NHGRI Working Group on Replication in Association Studies, Chanock, S.J., Manolio, T., Boehnke, M., Boerwinkle, E., Hunter, D.J., Thomas, G., Hirschhorn, J.N., Abecasis, G., Altshuler, D. et al. (2007). Replicating Genotype-Phenotype Associations. *Nature* **447(7145):** 655-660.

Ohira, T., Shahar, E., Chambless, L.E., Rosamond, W.D., Mosley, T.H.,Jr, and Folsom, A.R. (2006). Risk Factors for Ischemic Stroke Subtypes: The Atherosclerosis Risk in Communities Study. *Stroke* **37(10):** 2493-2498.

Out, A.A., van Minderhout, I.J., Goeman, J.J., Ariyurek, Y., Ossowski, S., Schneeberger, K., Weigel, D., van Galen, M., Taschner, P.E., Tops, C.M. et al. (2009). Deep Sequencing to Reveal New Variants in Pooled DNA Samples. *Hum Mutat* **30(12):** 1703-1712.

O'Sullivan, M. (2008). Leukoaraiosis. *Pract Neurol* **8(1):** 26-38.

Pantoni, L., Poggesi, A., and Inzitari, D. (2007). The Relation between White-Matter Lesions and Cognition. *Curr Opin Neurol* **20(4):** 390-397.

Pantoni, L., and Garcia, J.H. (1997a). Pathogenesis of Leukoaraiosis: A Review. *Stroke* **28(3):** 652-659.

Paternoster, L., Chen, W., and Sudlow, C.L. (2009). Genetic Determinants of White Matter Hyperintensities on Brain Scans: A Systematic Assessment of 19 Candidate Gene Polymorphisms in 46 Studies in 19,000 Subjects. *Stroke* **40(6):** 2020-2026.

Pohjasvaara, T., Mantyla, R., Ylikoski, R., Kaste, M., and Erkinjuntti, T. (2000). Comparison of Different Clinical Criteria (DSM-III, ADDTC, ICD-10, NINDS-AIREN, DSM-IV) for the Diagnosis of Vascular Dementia. National Institute of Neurological Disorders and Stroke-Association Internationale Pour La Recherche Et l'Enseignement En Neurosciences. *Stroke* **31(12):** 2952-2957.

Pollin, T.I., Damcott, C.M., Shen, H., Ott, S.H., Shelton, J., Horenstein, R.B., Post, W., McLenithan, J.C., Bielak, L.F., Peyser, P.A. et al. (2008). A Null Mutation in Human APOC3 Confers a Favorable Plasma Lipid Profile and Apparent Cardioprotection. *Science* **322(5908):** 1702-1705.

Prabhakaran, S., Wright, C.B., Yoshita, M., Delapaz, R., Brown, T., DeCarli, C., Sacco, R.L. (2008). Prevalence and Determinants of Subclinical Brain Infarction: The Northern Manhattan Study. *Neurology* **70:** 425-430.

Prins, N.D., van Dijk, E.J., den Heijer, T., Vermeer, S.E., Koudstaal, P.J., Oudkerk, M., Hofman, A., and Breteler, M.M. (2004). Cerebral White Matter Lesions and the Risk of Dementia. *Arch Neurol* **61(10):** 1531-1534.

Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W. et al. (2006). Global Variation in Copy Number in the Human Genome. *Nature* **444(7118):** 444-454.

Reitz, C., Berger, K., de Maat, M.P., Stoll, M., Friedrichs, F., Kardys, I., Witteman, J.C., and Breteler, M.M. (2007). CRP Gene Haplotypes, Serum CRP, and Cerebral Small-

Vessel Disease: The Rotterdam Scan Study and the MEMO Study. *Stroke* **38(8):** 2356-2359.

Risch, N.J. (2000). Searching for Genetic Determinants in the New Millennium. *Nature* **405(6788):** 847-856.

Roger, V.L., Go, A.S., Lloyd-Jones, D.M., Adams, R.J., Berry, J.D., Brown, T.M., Carnethon, M.R., Dai, S., de Simone, G., Ford, E.S., et al. (2011). Heart Disease and Stroke Statistics 2011 Update: A Report From the American Heart Association. *Circulation* **123**: e000-e000 (epub ahead of print).

Rovelet-Lecrux, A., Hannequin, D., Raux, G., Le Meur, N., Laquerriere, A., Vital, A., Dumanchin, C., Feuillette, S., Brice, A., Vercelletto, M. et al. (2006). APP Locus Duplication Causes Autosomal Dominant Early-Onset Alzheimer Disease with Cerebral Amyloid Angiopathy. *Nat Genet* **38(1):** 24-26.

Sabatti, C., Service, S.K., Hartikainen, A.L., Pouta, A., Ripatti, S., Brodsky, J., Jones, C.G., Zaitlen, N.A., Varilo, T., Kaakinen, M. et al. (2009). Genome-Wide Association Analysis of Metabolic Traits in a Birth Cohort from a Founder Population. *Nat Genet* **41(1):** 35-46.

Salerno, J.A., Murphy, D.G., Horwitz, B., DeCarli, C., Haxby, J.V., Rapoport, S.I., and Schapiro, M.B. (1992). Brain Atrophy in Hypertension. A Volumetric Magnetic Resonance Imaging Study. *Hypertension* **20(3):** 340-348.

Schmidt, H., Schmidt, R., Fazekas, F., Semmler, J., Kapeller, P., Reinhart, B., and Kostner, G.M. (1996). Apolipoprotein E e4 Allele in the Normal Elderly: Neuropsychologic and Brain MRI Correlates. *Clin Genet* **50(5):** 293-299.

Schmidt, R., Petrovic, K., Ropele, S., Enzinger, C., and Fazekas, F. (2007). Progression of Leukoaraiosis and Cognition. *Stroke* **38(9):** 2619-2625.

Schmidt, R., Scheltens, P., Erkinjuntti, T., Pantoni, L., Markus, H.S., Wallin, A., Barkhof, F., and Fazekas, F. (2004). White Matter Lesion Progression: A Surrogate Endpoint for Trials in Cerebral Small-Vessel Disease. *Neurology* **63(1):** 139-144.

Schmidt, R., Schmidt, H., Fazekas, F., Launer, L.J., Niederkorn, K., Kapeller, P., Lechner, A., and Kostner, G.M. (2001). Angiotensinogen Polymorphism M235T, Carotid Atherosclerosis, and Small-Vessel Disease-Related Cerebral Abnormalities. *Hypertension* **38(1):** 110-115.

Schmidt, R., Schmidt, H., Fazekas, F., Kapeller, P., Roob, G., Lechner, A., Kostner, G.M., and Hartung, H.P. (2000). MRI Cerebral White Matter Lesions and Paraoxonase PON1 Polymorphisms : Three-Year Follow-Up of the Austrian Stroke Prevention Study. *Arterioscler Thromb Vasc Biol* **20(7):** 1811-1816.

Schulz, U.G., Flossmann, E., and Rothwell, P.M. (2004). Heritability of Ischemic Stroke in Relation to Age, Vascular Risk Factors, and Subtypes of Incident Stroke in Population-Based Studies. *Stroke* **35(4):** 819-824.

Scott, L.J., Mohlke, K.L., Bonnycastle, L.L., Willer, C.J., Li, Y., Duren, W.L., Erdos, M.R., Stringham, H.M., Chines, P.S., Jackson, A.U. et al. (2007). A Genome-Wide Association Study of Type 2 Diabetes in Finns Detects Multiple Susceptibility Variants. *Science* **316(5829):** 1341-1345.

Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., Yamrom, B., Yoon, S., Krasnitz, A., Kendall, J. et al. (2007). Strong Association of De Novo Copy Number Mutations with Autism. *Science* **316(5823):** 445-449.

Seshadri, S., DeStefano, A.L., Au, R., Massaro, J.M., Beiser, A.S., Kelly-Hayes, M., Kase, C.S., D'Agostino RB, S., Decarli, C., Atwood, L.D. et al. (2007a). Genetic Correlates of Brain Aging on MRI and Cognitive Test Measures: A Genome-Wide Association and Linkage Analysis in the Framingham Study. *BMC Med Genet* **8(Suppl 1):** S15.

Seshadri, S., and Wolf, P.A. (2007b). Lifetime Risk of Stroke and Dementia: Current Concepts, and Estimates from the Framingham Study. *Lancet Neurol* **6(12):** 1106-1114.

Sierra, C., and Coca, A. (2006). White Matter Lesions and Cognitive Impairment as Silent Cerebral Disease in Hypertension. *ScientificWorldJournal* **6:** 494-501.

Sing, C.F., Boerwinkle, E., Moll, P.P., and Templeton, A.R. (1987). *Characterization of Genes affecting Quantitative Traits in Humans.* In *Proceedings of the Second International Conference on Quantitative Genetics,* B. Weir, E.J. Eisen, M.M. Goodman, and G. Namkoong, eds. Sinauer Associates, Inc.: Raliegh, NC, pp. 250-269.

Sing, C.F., Stengard, J.H., and Kardia, S.L. (2004). Dynamic Relationships between the Genome and Exposures to Environments as Causes of Common Human Diseases. *World Rev Nutr Diet* **93:** 77-91.

Slattery, M.L., Folsom, A.R., Wolff, R., Herrick, J., Caan, B.J., and Potter, J.D. (2008). Transcription Factor 7-Like 2 Polymorphism and Colon Cancer. *Cancer Epidemiol Biomarkers Prev* **17(4):** 978-982.

Smith, J.A., Turner, S.T., Sun, Y.V., Fornage, M., Kelly, R.J., Mosley, T.H., Jack, C.R., Kullo, I.J., and Kardia, S.L. (2009). Complexity in the Genetic Architecture of Leukoaraiosis in Hypertensive Sibships from the GENOA Study. *BMC Med Genomics* **2:** 16.

Smith, J.A., Arnett, D.K., Kelly, R.J., Ordovas, J.M., Sun, Y.V., Hopkins, P.N., Hixson, J.E., Straka, R.J., Peacock, J.M., and Kardia, S.L. (2008). The Genetic Architecture of

Fasting Plasma Triglyceride Response to Fenofibrate Treatment. *Eur J Hum Genet* **16(5):** 603-613.

Stefansson, H., Rujescu, D., Cichon, S., Pietilainen, O.P., Ingason, A., Steinberg, S., Fossdal, R., Sigurdsson, E., Sigmundsson, T., Buizer-Voskamp, J.E. et al. (2008). Large Recurrent Microdeletions Associated with Schizophrenia. *Nature* **455(7210):** 232-236.

Steffens, D.C., Trost, W.T., Payne, M.E., Hybels, C.F., and MacFall, J.R. (2003). Apolipoprotein E Genotype and Subcortical Vascular Lesions in Older Depressed Patients and Control Subjects. *Biol Psychiatry* **54(7):** 674-681.

Swan, G.E., DeCarli, C., Miller, B.L., Reed, T., Wolf, P.A., Jack, L.M., and Carmelli, D. (1998). Association of Midlife Blood Pressure to Late-Life Cognitive Decline and Brain Morphology. *Neurology* **51(4):** 986-993.

Szolnoki, Z. (2007a). Chemical Events Behind Leukoaraiosis: Medicinal Chemistry Offers New Insight into a Specific Microcirculation Disturbance in the Brain (a Chemical Approach to a Frequent Cerebral Phenotype). *Curr Med Chem* **14(9):** 1027-1036.

Szolnoki, Z., Kondacs, A., Mandi, Y., and Somogyvari, F. (2007b). A Genetic Variant in Cytoskeleton Motors Amplifies Susceptibility to Leukoaraiosis in Hypertensive Smokers: Gene-Environmental Interactions Behind Vascular White Matter Demyelinization. *J Mol Neurosci* **33(2):** 173-179.

Szolnoki, Z., and Melegh, B. (2006). Gene-Gene and Gene-Environment Interplay Represent Specific Susceptibility for Different Types of Ischaemic Stroke and Leukoaraiosis. *Curr Med Chem* **13(14):** 1627-1634.

Szolnoki, Z., Somogyvari, F., Kondacs, A., Szabo, M., Fodor, L., Bene, J., and Melegh, B. (2004). Specific APO E Genotypes in Combination with the ACE D/D Or MTHFR 677TT Mutation Yield an Independent Genetic Risk of Leukoaraiosis. *Acta Neurol Scand* **109(3):** 222-227.

Terwilliger, J.D., and Goring, H.H. (2009). Gene Mapping in the 20th and 21st Centuries: Statistical Methods, Data Analysis, and Experimental Design. 2000. *Hum Biol* **81(5-6):** 663-728.

Terwilliger, J.D., and Weiss, K.M. (1998). Linkage Disequilibrium Mapping of Complex Disease: Fantasy Or Reality? *Curr Opin Biotechnol* **9(6):** 578-594.

Thomas, G., Jacobs, K.B., Yeager, M., Kraft, P., Wacholder, S., Orr, N., Yu, K., Chatterjee, N., Welch, R., Hutchinson, A. et al. (2008). Multiple Loci Identified in a Genome-Wide Association Study of Prostate Cancer. *Nat Genet* **40(3):** 310-315.

Tombaugh, T.N., and McIntyre, N.J. (1992). The Mini-Mental State Examination: A Comprehensive Review. *J Am Geriatr Soc* **40(9):** 922-935.

Tozawa, M., Iseki, K., Iseki, C., Kinjo, K., Ikemiya, Y., and Takishita, S. (2003). Blood Pressure Predicts Risk of Developing End-Stage Renal Disease in Men and Women. *Hypertension* **41(6):** 1341-1345.

Turner, S.T., Fornage, M., Jack, C.R.,Jr, Mosley, T.H., Knopman, D.S., Kardia, S.L., Boerwinkle, E., and de Andrade, M. (2009). Genomic Susceptibility Loci for Brain Atrophy, Ventricular Volume, and Leukoaraiosis in Hypertensive Sibships. *Arch Neurol* **66(7):** 847-857.

Turner, S.T., Fornage, M., Jack, C.R.,Jr, Mosley, T.H., Kardia, S.L., Boerwinkle, E., and de Andrade, M. (2005). Genomic Susceptibility Loci for Brain Atrophy in Hypertensive Sibships from the GENOA Study. *Hypertension* **45(4):** 793-798.

Turner, S.T., Jack, C.R., Fornage, M., Mosley, T.H., Boerwinkle, E., and de Andrade, M. (2004). Heritability of Leukoaraiosis in Hypertensive Sibships. *Hypertension* **43(2):** 483-487.

Turner, S.T., Bielak, L.F., Narayana, A.K., Sheedy, P.F.,2nd, Schwartz, G.L., and Peyser, P.A. (2002). Ambulatory Blood Pressure and Coronary Artery Calcification in Middle-Aged and Younger Adults. *Am J Hypertens* **15(6):** 518-524.

Turner, S.T., and Boerwinkle, E. (2000). Genetics of Hypertension, Target-Organ Complications, and Response to Therapy. *Circulation* **102(20 Suppl 4):** 40-45.

van Dijk, E.J., Breteler, M.M., Schmidt, R., Berger, K., Nilsson, L.G., Oudkerk, M., Pajak, A., Sans, S., de Ridder, M., Dufouil, C. et al. (2004). The Association between Blood Pressure, Hypertension, and Cerebral White Matter Lesions: Cardiovascular Determinants of Dementia Study. *Hypertension* **44(5):** 625-630.

Van Heyningen, V., and Yeyati, P.L. (2004). Mechanisms of Non-Mendelian Inheritance in Genetic Disease. *Hum Mol Genet* **13(Supp 2):** R225-233.

van Oijen, M., Cheung, E.Y., Geluk, C.E., Hofman, A., Koudstaal, P.J., Breteler, M.M., and de Maat, M.P. (2008). Haplotypes of the Fibrinogen Gene and Cerebral Small Vessel Disease: The Rotterdam Scan Study. *J Neurol Neurosurg Psychiatry* **79(7):** 799-803.

Visscher, P.M. (2008a). Sizing Up Human Height Variation. *Nat Genet* **40(5):** 489-490.

Visscher, P.M., Hill, W.G., and Wray, N.R. (2008b). Heritability in the Genomics Era--Concepts and Misconceptions. *Nat Rev Genet* **9(4):** 255-266.

Visscher, P.M., Medland, S.E., Ferreira, M.A., Morley, K.I., Zhu, G., Cornes, B.K., Montgomery, G.W., and Martin, N.G. (2006). Assumption-Free Estimation of Heritability from Genome-Wide Identity-by-Descent Sharing between Full Siblings. *PLoS Genet* **2(3):** e41.

Wain, L.V., Armour, J.A., and Tobin, M.D. (2009). Genomic Copy Number Variation, Human Health, and Disease. *Lancet* **374(9686):** 340-350.

Wang, J., Cao, H., Ban, M.R., Kennedy, B.A., Zhu, S., Anand, S., Yusuf, S., Pollex, R.L., and Hegele, R.A. (2007). Resequencing Genomic DNA of Patients with Severe Hypertriglyceridemia (MIM 144650). *Arterioscler Thromb Vasc Biol* **27(11):** 2450-2455.

Wang, W.Y., Barratt, B.J., Clayton, D.G., and Todd, J.A. (2005). Genome-Wide Association Studies: Theoretical and Practical Concerns. *Nat Rev Genet* **6(2):** 109-118.

Weiss, K.M., and Terwilliger, J.D. (2000). How Many Diseases does it Take to Map a Gene with SNPs? *Nat Genet* **26(2):** 151-157.

Winckler, W., Weedon, M.N., Graham, R.R., McCarroll, S.A., Purcell, S., Almgren, P., Tuomi, T., Gaudet, D., Bostrom, K.B., Walker, M. et al. (2007). Evaluation of Common Variants in the Six Known Maturity-Onset Diabetes of the Young (MODY) Genes for Association with Type 2 Diabetes. *Diabetes* **56(3):** 685-693.

Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W. et al. (2010). Common SNPs Explain a Large Proportion of the Heritability for Human Height. *Nat Genet* **42(7):** 565-569.

Zeggini, E., Scott, L.J., Saxena, R., Voight, B.F., Marchini, J.L., Hu, T., de Bakker, P.I., Abecasis, G.R., Almgren, P., Andersen, G. et al. (2008). Meta-Analysis of Genome-Wide Association Data and Large-Scale Replication Identifies Additional Susceptibility Loci for Type 2 Diabetes. *Nat Genet* **40(5):** 638-645.

Zhou, S., Stone, E.A., Mackay, T.F., and Anholt, R.R. (2009). Plasticity of the Chemoreceptor Repertoire in Drosophila Melanogaster. *PLoS Genet* **5(10):** e1000681.

Zou, L., Sriswasdi, S., Ross, B., Missiuro, P.V., Liu, J., and Ge, H. (2008). Systematic Analysis of Pleiotropy in C. Elegans Early Embryogenesis. *PLoS Comput Biol* **4(2):** e1000003.

Chapter 2

Complexity in the Genetic Architecture of Leukoaraiosis in Hypertensive Sibships*

Abstract

Subcortical white matter hyperintensity on magnetic resonance imaging (MRI) of the brain, referred to as leukoaraiosis, is associated with increased risk of stroke and dementia. Hypertension may contribute to leukoaraiosis by accelerating the process of arteriosclerosis involving penetrating small arteries and arterioles in the brain. Leukoaraiosis volume is highly heritable but shows significant inter-individual variability that is not predicted well by any clinical covariates (except for age) or by single nucleotide polymorphisms (SNPs). As part of the Genetics of Microangiopathic Brain Injury (GMBI) Study, 777 individuals (74% hypertensive) underwent brain MRI and were genotyped for 1649 SNPs from genes known or hypothesized to be involved in arteriosclerosis and related pathways. We examined SNP main effects, epistatic (gene-gene) interactions, and context-dependent (gene-environment) interactions between these SNPs and covariates (including conventional and novel risk factors for arteriosclerosis) for association with leukoaraiosis volume. Three methods were used to reduce the chance of false positive associations: 1) false discovery rate (FDR) adjustment for multiple testing, 2) an internal replication design, and 3) a ten-iteration four-fold cross-validation

*This work was previously published as Smith JA, Turner ST, Sun YV, Fornage M, Kelly RJ, Mosley T, Jack CR, Kullo IJ, Kardia SLR. (2009). Complexity in the Genetic Architecture of Leukoaraiosis in Hypertensive Sibships from the GENOA Study. BMC Medical Genomics 2(1): 16.

scheme. Four SNP main effects (in *F3*, *KITLG*, *CAPN10*, and *MMP2*), 12 SNP-covariate interactions (including interactions between *KITLG* and homocysteine, and between *TGFB3* and both physical activity and C reactive protein), and 173 SNP-SNP interactions were significant, replicated, and cross-validated. While a model containing the top single SNPs with main effects predicted only 3.87% variation in leukoaraiosis in independent test samples, a multiple variable model that included the four most highly predictive SNP-SNP and SNP-covariate interactions predicted 11.83%. These results indicate that the genetic architecture of leukoaraiosis is complex, yet predictive, when the contributions of SNP main effects are considered in combination with effects of SNP interactions with other genes and covariates.

Introduction

Stroke and dementia are age-related neurological disorders that cause considerable morbidity and financial burden in the US. The lifetime risk for developing one or both of these disorders is greater than 1 in 3 (Seshadri, 2007). Risk factors for stroke and dementia overlap in part with those for cardiovascular disease (including age, sex, tobacco use, hypertension, diabetes mellitus, and low physical activity), but it has been established that both disorders have a significant genetic component that operates independently of these risk factors (Flossmann, 2004). Although several rare genetic variations have been identified that are associated with significantly elevated risk of stroke or dementia, the vast majority of genes that influence risk for these disorders remain unidentified.

In an effort to increase the statistical power for detecting genetic variants that have small effects on the development of late-life endpoints, such as stroke and dementia, quantitative subclinical phenotypes are often used as an indicator for risk of future disease. Subclinical phenotypes may be closer in the biological hierarchy to the underlying genetic processes, including the influence of gene-environment interactions, and thus may have a larger genetic component than clinical disease endpoints.

Ischemic damage to the subcortical white matter that manifests as white matter hyperintensity on magnetic resonance imaging (MRI) of the brain, referred to as leukoaraiosis, is associated with increased risk of stroke and dementia (Kuller, 2004; Salerno, 1992; Pantoni, 1997). One of the strongest predictors of leukoaraiosis is elevated blood pressure (van Dijk, 2004), in particular, inadequate blood pressure control in persons with hypertension (Liao, 1996). Hypertension is thought to contribute to the pathology of leukoaraiosis through accelerating the age-related process of arteriosclerosis resulting in ischemic damage to small penetrating arterioles in the subcortical white matter of the brain (Schwartz, 2007). This connection between hypertension and leukoaraiosis motivated the measurement of this subclinical phenotype in a subsample of the Genetic Epidemiology Network of Arteriopathy (GENOA) study of hypertensive sibships (FBPP Investigators, 2002).

In the GENOA cohort, the heritability of the logarithm transformed measure of leukoaraiosis volume was 0.80, which decreased to 0.68 after adjustment for sex, age, systolic blood pressure, and brain volume (Turner, 2004). In other studies, the heritability

of white matter hyperintensities on MRI was estimated to be 0.73 in a study of male twins (Carmelli, 1998) and 0.55 in the Framingham Heart Study (Atwood, 2004).

To begin to explore the genetic architecture of this trait, we identified single nucleotide polymorphisms (SNPs) in 268 genes that have been previously identified as playing a role in processes related to arteriosclerosis including blood pressure regulation, vascular wall biology, oxidative stress, inflammation, obesity, diabetes, and lipoprotein metabolism. The goal of the present study was to investigate the contributions, covariation, and interaction among the many hypothesized genetic and environmental factors that may influence inter-individual variation in leukoaraiosis. Using a systematic approach that simultaneously investigates the contributions of these factors (as main effects or as part of interactions) and their underlying covariation, this study is a first step toward understanding the complexity of the genetic architecture of leukoaraiosis in order to begin to build multivariable models that can predict levels of structural brain injury that may result from a person's unique combination of risk factors.

Methods

Study population

The 777 study participants consisted of non-Hispanic white adults (322 male and 455 female) from 357 sibships that were initially enrolled in the Genetic Epidemiology Network of Arteriopathy (GENOA) study, a community-based study of hypertensive sibships that aims to identify genes influencing blood pressure (BP) (FBPP Investigators, 2002; Daniels, 2004). The study was approved by the Institutional Review Board of

Mayo Clinic, Rochester MN, and written informed consent was obtained from each participant. In the initial phase of the GENOA study (9/1995 to 6/2001), sibships containing ≥ 2 individuals with essential hypertension diagnosed before age 60 years were selected for participation. Participants returned for a second phase of the study (12/2000 to 6/2004) that included a physical examination and measurement of conventional and novel risk factors.

As an ancillary study of GENOA conducted between August 2001 and May 2006, the Genetics of Microangiopathic Brain Injury (GMBI) study was undertaken to determine susceptibility genes for ischemic brain injury. Leukoaraiosis was quantified by MRI in 916 non-Hispanic white subjects who participated in the second phase of the GENOA study, had a sibling willing and eligible to participate in the GMBI study, and had no history of stroke or neurological disease and no implanted metal devices. The median time between the second GENOA examination and the GMBI brain MRI was 11.9 months. Brain MRIs were suitable for analysis in 883 of the 916 participants; in the 33 without analyzable data, the most common reasons were unsuspected prior brain infarctions, masses, metallic artifacts, and failure to complete the MRI. After removing individuals who did not have genotyping data available, the final analysis subset consisted of 777 GMBI participants.

Clinical assessments and covariate definitions

The diagnosis of hypertension was established based on BP levels measured at the study visit (>140 mmHg average systolic BP or >90 mmHg average diastolic BP) or a prior

diagnosis of hypertension and current treatment with antihypertensive medications. Height was measured by stadiometer, weight by electronic balance, and body mass index (BMI) was calculated as weight in kilograms divided by the square of height in meters. Resting systolic blood pressure (SBP) and diastolic blood pressure (DBP) were measured by a random zero sphygmomanometer, and pulse pressure was calculated as the difference between SBP and DBP. A person was considered having ever smoked if they had smoked more than 100 cigarettes in their lifetime, was considered to have coronary heart disease if they had ever experienced a myocardial infarction or surgery for a blocked artery in the heart or neck (carotid artery), and was considered obese if they had a BMI > 30 kg/m^2.

Blood was drawn by venipuncture after an overnight fast. Serum triglycerides (TG), creatinine, total cholesterol, and high-density lipoprotein (HDL) cholesterol were measured by standard enzymatic methods on a Hitachi 911 Chemistry Analyzer (Roche Diagnostics, Indianapolis IN), and low-density lipoprotein (LDL) cholesterol levels were calculated using the Friedewald formula (Friedewald, 1972). Five novel vascular risk factors including C-reactive protein, homocysteine, fibrinogen, Lp(a), and LDL particle size were also measured. C-reactive protein was measured by a highly sensitive immunoturbidimetric assay (Keevil, 1998), fibrinogen was measured by the Clauss (clotting time based) method (von Clauss, 1957), and plasma homocysteine was measured by high-pressure liquid chromatography. Lp(a) in serum was measured by an immunoturbidimetric assay using the SPQ™ Test System (Diasorin, Stillwater MN) as previously described (Kullo, 2004a), and LDL particle size was measured by

polyacrylamide gel electrophoresis (Kullo, 2004b). Level of physical activity was calculated as a continuous variable based on the self-reported average number of hours per day that the subject engaged in heavy, moderate, and sedentary activities according the following formula: 2*Heavy + Moderate – 2*Sedentary.

Leukoaraiosis volume (cm^3) was obtained via MRI in a separate clinical visit. All MRI scans were performed on identically equipped Signa 1.5 T MRI scanners (GE Medical Systems, Waukesha, WI, USA) and images were centrally processed at the Mayo Clinic. Symmetric head positioning with respect to orthogonal axes was verified by a series of short scout scans. Total intracranial volume (head size) was measured from T1-weighted spin echo sagittal images, each set consisting of 32 contiguous 5 mm thick slices with no interslice gap, field of view = 24 cm, matrix = 256 x 192, obtained with the following sequence: scan time = 2.5 min, echo time = 14 ms, repetitions = 2, replication time = 500 ms (Jack, 1989). Total brain and leukoaraiosis volumes were determined from axial fluid-attenuated inversion recovery (FLAIR) images, each set consisting of 48 contiguous 3-mm interleaved slices with no interslice gap, field of view = 22 cm, matrix = 256 x 160, obtained with the following sequence: scan time = 9 min, echo time = 144.8 ms, inversion time = 2,600 ms, repetition time = 26,002 ms, bandwidth = +/- 15.6 kHz, one signal average. A FLAIR image is a T2-weighted image with the signal of the cerebrospinal fluid nulled, such that brain pathology appears as the brightest intracranial tissue. Interactive imaging processing steps were performed by a research associate who had no knowledge of the subjects' personal or medical histories or biological relationships. A fully automated algorithm was used to segment each slice of the edited

multi-slice FLAIR sequence into voxels assigned to one of three categories: brain, cerebrospinal fluid, or leukoaraiosis. The mean absolute error of this method is 1.4% for brain volume and 6.6% for leukoaraiosis volume, and the mean test-retest coefficient of variation is 0.3% for brain volume and 1.4% for leukoaraiosis volume (Jack, 2001). White matter hyperintensities in the corona-radiata and periventricular zone, as well as central gray infarcts (ie, lacunes) were included in the global leukoaraiosis measurements. Brain scans with cortical infarctions were excluded from the analyses because of the distortion of the leukoaraiosis volume estimates that would be introduced in the automated segmentation algorithm.

Genotyping

One thousand nine hundred and fifty six SNPs from 268 genes known or hypothesized to be involved in blood pressure regulation, lipoprotein metabolism, inflammation, oxidative stress, vascular wall biology, obesity and diabetes were identified from the genetic association literature and positional candidate gene studies (Barkley, 2004). SNPs were chosen based on a number of different criteria including the published literature, non-synonymous SNPs with a minor allele frequency (MAF) > 0.02, and tag SNPs identified using public databases such as dbSNP (http://www.ncbi.nlm.nih.gov/SNP/) and the Seattle SNPs database (http://pga.mbt.washington.edu).

DNA was isolated using the PureGene DNA Isolation Kit from Gentra Systems (Minneapolis MN). Genotyping, based on polymerase chain reaction (PCR) amplification techniques, was conducted at the University of Texas-Health Sciences Center at Houston

using the TaqMan assay and ABI Prism® Sequence Detection System (Applied Biosystems, Foster City CA). Primers and probes are available from the authors upon request. Quality control measures for genotyping assays included robotic liquid handling, separate pre- and post-PCR areas, standard protocols and quality control analyses including 5% duplicates, positive and negative controls, computerized sample tracking, and data validity checks. After removal of SNPs that were monomorphic in the study sample, 1649 SNPs remained for analysis (see Appendix 2.1).

Statistical analysis

All analyses were carried out using the R statistical language, version 2.8 (R Core Development Team, 2008). Covariate correlations were estimated using Pearson's product moment correlation. Linkage disequilibrium (LD), as measured by r^2 (Lynch, 1998), was estimated using an expectation maximization (EM) algorithm. Hardy-Weinberg Equilibrium was assessed using a chi-square test or Fisher's exact test if a genotype class had less than 5 individuals (Weir, 1996). Variables that showed deviation from a normal distribution in diagnostic plots, including leukoaraiosis, were transformed by taking the natural logarithm. The outcome variable for all models was the residual value of the natural logarithm of leukoaraiosis volume (cm^3) after adjustment for age, sex, and total brain volume. Age and sex were included as adjustment covariates because both have been historically used as adjustment variables for this trait. Age is a very strong independent predictor for leukoaraiosis, and sex had a marginal association with leukoaraiosis in this sample. To account for differences in brain size, brain volume was included as a covariate in all models.

In the first stage of the analysis, we tested for association between leukoaraiosis and each of the predictor variables (SNPs and quantitative covariates) using least-squares linear regression methods (Weir, 1996; Kleinbaum, 1998). Categorial covariates were modeled using logistic regression (Kleinbaum, 1998). We also tested for association between each SNP and covariate to identify potential confounders. To determine whether interactions among predictors explained additional variation in the outcome, we tested pairwise interactions among all possible pairs of predictors (i.e. SNP-SNP, SNP-covariate, and covariate-covariate interactions) for all covariates and the 444 SNPs that had a model p-value<0.2 in the association testing described above. Associations involving interactions were assessed with a partial F test, which compares a full model that includes both the interaction terms and the main effects of the variables comprising the interaction terms to a reduced model that includes only the main effects. Models with a p-value <0.1 (for single variable models) or a partial F p-value <0.1 (for models with interaction terms) were evaluated in the next stage of analysis.

To reduce false positives we used three different approaches: 1) adjustment for multiple testing using the False Discovery Rate (FDR) <0.30 (Storey, 2002), 2) internal replication with two subsets of the data (constructed so individuals were unrelated within subset), and 3) four-fold cross-validation (repeated 10 times) (Molinaro, 2005). To create internal replication subsets, we randomly selected one sibling from each sibship without replacement to create subset 1 and then randomly selected another sibling from each sibship to create subset 2. The GMBI cohort contained a small number of singletons (ie,

subjects who had no enrolled sibling) that were equally divided between the two samples. If a sibship contained more than two siblings, the remaining unselected siblings were not included in either subset. Associations that had a p-value <0.1 in both subsets were considered internally replicated if the effect of the genotype was homogeneous among subsets (the partial F p-value >0.05 from a test of the interaction between subset designation and the predictors(s) under consideration).

Cross-validation significantly reduces false positive results by eliminating associations that lack predictive ability in independent test samples. For each association, we performed four-fold cross-validation by dividing the full sample into four equally sized groups. Three of the four groups were combined into a training dataset, and the modeling strategy outlined above was carried out to estimate model coefficients. These coefficients were then applied to the fourth group, the testing dataset, to predict the value of the outcome variable for each individual in this independent test sample. This process was repeated for each of the four testing sets. Predicted values for all individuals in the test set were then subtracted from their observed values, yielding the total residual variability (SSE), $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$. The total variability in the outcome (SST) – the difference between each individual's observed value and the mean value for the outcome – was then calculated, $\sum_{i=1}^{n}(\bar{y} - y_i)^2$. In order to estimate the proportion of variation in the outcome predicted in the independent test samples, the cross-validated R^2 (CV R^2) was calculated as follows: $CV\ R^2 = \frac{SST - SSE}{SST}$. This cross-validation method provides a more accurate

measure of the predictive ability of the genetic models and will be negative when the model's predictive ability is poor. Because random variations in the sampling of the four mutually exclusive test groups can potentially impact the estimates of CV R^2, this procedure was repeated 10 times and the CV R^2 values were averaged (Molinaro, 2005).

Univariate associations were considered cross-validated if the average percent variation predicted in independent test samples was greater than 0.5% and interactions were considered cross-validated if the difference in average percent variation predicted in independent test samples between the full model containing the interaction term and the reduced model containing only main effect terms was greater than 0.5%. Using permutation testing on the models investigated in this paper, we found that the probability of observing a CV R^2x100 greater than 0.5% by chance alone was less than 5%. That is, Pr(CV R^2x100 > 0.5%) < 0.05 under the null hypothesis of no association. Due to small cell sizes (<4 subjects in a particular class), 0.3% of the SNP-covariate interaction models and 2.3% of the SNP-SNP interaction models were unable to complete the cross-validation procedure.

All single SNP or interaction models that passed the three different approaches for reducing false positives (FDR, internal replication, and cross-validation) were modeled using linear mixed effects (LME) (Raudenbush, 2002), which accounts for the sibship structure among GMBI study participants while retaining a valid type I error rate (Cupples, 2007; Raudenbush, 2002). Associations with a p-value <0.1 in the F test (described above) but a p-value >0.1 from the likelihood ratio test of the appropriate full

and reduced mixed effects models were considered to be associations due to family structure and were removed from the results.

To visualize the genetic architecture of the outcome trait, leukoaraiosis volume, we applied a novel data visualization scheme, the KGraph, described in Kelly et al. (2007). The KGraph was developed for the visualization of genetic association results and the underlying relationships among predictors such as SNP-SNP frequency correlations (i.e. LD), SNP-covariate associations, and covariate-covariate correlations. It simultaneously displays both significant univariate associations and pairwise interactions with the outcome of interest, leukoaraiosis volume, as well as the underlying correlation structure among the predictor variables.

In the final step, multivariable linear regression models combining the most predictive SNPs, covariates, and their interactions were constructed. The top four single SNP, SNP-covariate, and SNP-SNP interaction models were chosen for multiple variable modeling based on the following criteria: 1) passed all three filters to reduce false positive associations (FDR, internal replication, and cross-validation), 2) had the highest CV R^2 values of the particular modeling strategy, and 3) didn't involve SNPs in strong LD with SNPs already included in the multiple variable model. Percent variation in leukoaraiosis volume explained by each model was assessed with the adjusted R^2 value, and predictive ability of the models was assessed by four-fold, ten-iteration cross validation (CV R^2 value).

Results

Descriptive statistics

Descriptive statistics of the clinical covariates and outcomes are shown in Table 2.1. The mean age of the participants was 59.7 years and 58.6% of participants were female. Participants had a mean BMI of 30.5 kg/m^2, waist-to-hip ratio of 0.91, SBP of 131.4 mmHg, and DBP of 74.0 mmHg. The distribution of leukoaraiosis is shown in Figure 2.1. The mean volume of leukoaraiosis was 7.80 cm^3 and the mean brain size was 1159 cm^3. Allele and genotype frequencies, rs numbers from dbSNP, SNP positions and annotations (synonymous, non-synonymous, intron, etc), and test results for Hardy-Weinberg equilibrium are reported in Appendix 2.1.

Associations

Table 2.2 shows a summary of the results from testing for SNP main effects, SNP-covariate interactions, and SNP-SNP interactions. Of the 1649 SNPs that were evaluated for their association with leukoaraiosis, 37 had FDR<0.3, 15 internally replicated, 23 cross-validated, and only four met all three criteria. In tests for SNP-covariate interaction, 1561 interactions had a FDR<0.3, 834 internally replicated, 1887 cross-validated, and only 12 met all three criteria. In tests for SNP-SNP interactions, one hundred and seventy three SNP-SNP interactions passed all three criteria, and the top 20 most predictive of these interactions are listed in Table 2.3 along with the single SNP main effects and 12 SNP-covariate interactions that met all three criteria (see Appendix 2.2 for a complete list of SNP-SNP interactions that passed all three criteria).

Figure 2.2 shows a KGraph, a visual representation of the complex associations among genetic, demographic, and biochemical factors that underlie variation in leukoaraiosis volume. Using both color and spatial relationships, the KGraph presents both associations with leukoaraiosis and the correlation structure of the predictors that underlie those associations. A key to the eight regions of the KGraph is located in the lower left corner of Figure 2.2, and Appendix 2.3 gives additional detail about reading and interpreting a KGraph. Included on the KGraph are all of the covariates that were investigated in the study, SNPs that were involved in a single SNP or SNP-covariate association that passed all three filters, and SNPs that were involved in at least one of the 20 most highly predictive SNP-SNP interactions that passed all three filters. All associations involving these SNPs and covariates are presented on the KGraph, and those that passed all three filters are indicated by a horizontal black bar.

Region 1 in Figure 2.2, shown in green, displays the association between the SNPs and covariates, one source of information about the underlying pathways. The majority of SNP-covariate associations were accounted for by three SNPs in the factor VIII (*F8*) gene that were associated with log serum creatinine, height, HDL cholesterol, waist-to-hip ratio, and weight. Region 2, shown in grey, illustrates the correlations between the covariates. The majority of the risk factors are significantly associated with one another (p-value < 0.05). Region 3, in red, shows the observed LD, estimated in a sample of 357 unrelated individuals from the study sample. As expected, significant LD ($r^2 > 0.5$) occurs only between SNPs that are within the same gene.

The remaining regions are colored blue, indicating that they represent associations with the outcome of interest, leukoaraiosis. Region 4, which displays the univariate association between the covariates and leukoaraiosis, shows that only age had an association that met all three criteria. Region 5, which illustrates univariate associations between the SNPs and leukoaraiosis, shows that four SNPs have significant, replicated and cross-validated associations (*F3_rs3917643*, *CAPN10_rs7571442*, *MMP2_rs9928731*, *KITLG_rs995029*). Region 6 displays the covariate-covariate interactions that are significantly associated with leukoaraiosis, but no interactions of this type passed all three filters. Region 7 displays the interactions between the SNPs and covariates that were associated with leukoaraiosis. Overall, we detected 12 interactions that replicated and cross validated, though two pairs of SNP-covariate interactions appear to be marking the same association, due to strong LD between the involved SNPs. Region 8 displays the epistatic (SNP-SNP) interactions significantly associated with leukoaraiosis. We detected 173 replicated and cross-validated, statistically significant pairwise interactions between SNPs. The most predictive interactions included those between SNPs in *RHAG* and *GLS*, *F8* and *MPO*, *SLC20A1* and *IL22RA*, *KITLG* and *TLR4*, *NMUR1* and *GPR55*, *ACCN4* and *TNFSF10*, and *CX3CR1* and *F2*. Interactions between two genes that appear more than once in the SNP-SNP results are almost entirely due to strong LD between involved SNPs.

Predictive modeling

To begin to assess the combined predictive ability of the top SNPs, covariates, and their interactions, we constructed multiple variable models as described in the Methods section

(Table 2.4). The four single SNPs that met all three criteria explained 5.99% of variation in leukoaraiosis (adjusted R^2) and had a CV R^2x100 value of 3.72%. A model that included the main effects and interaction terms from the top four SNP-covariate interactions explained 7.88% of the variation in leukoaraiosis (CV R^2x100 = 4.53%), while a model including only the SNP and covariate main effect terms had a negative CV R^2, indicating poor predictive performance. A model consisting of the top four SNP-SNP interactions explained 14.73% of variation in leukoaraiosis (CV R^2x100 = 9.59%), while the model containing only the SNP main effects explained only 6.12% (CV R^2x100 = 2.27%), indicating that the SNP-SNP interaction terms explained an additional 7.61% of variation (difference in CV R^2x100 = 7.32%). Finally, a model that contained both the top four SNP-covariate and the top four SNP-SNP interactions explained 19.18% of the variation in leukoaraiosis (CV R^2x100 = 11.83%), while the reduced model containing only the SNP and covariate main effects terms explained 7.18% (CV R^2x100 = 1.30%). Therefore, the combination of SNP-SNP and SNP-covariate interactions was the most predictive model, explaining an additional 12.00% variation in leukoaraiosis (difference in CV R^2x100 = 10.80%).

Discussion

Although there have been several studies of the influence of polymorphisms in candidate genes on essential hypertension, stroke, and dementia, little research has been done on the impact of specific candidate gene polymorphisms on leukoaraiosis. Our motivating hypothesis for this work was that polymorphisms in physiological pathways known to affect these factors for arteriosclerosis may influence leukoaraiosis both directly and

through interactions with environmental, demographic, and behavioral risk factors and other genetic polymorphisms.

Except for age and blood pressure, conventional risk factors do not significantly predict leukoaraiosis in our study. However, these covariates predict a large fraction (~30%) of variation in leukoaraiosis. After adjustment for age and sex, four SNPs passed all three filters to reduce false positives and significantly predicted this phenotype. These SNPs represent several distinct physiological pathways, including blood coagulation (*F3*) (Davie, 1991), endothelial and hematopoietic stem cell proliferation (*KITLG*) (Martin, 1990), protease pathways contributing to obesity and diabetes (*CAPN10*) (Horikawa, 2000), and the extracellular matrix (*MMP2*) (Liu, 2006). This result emphasizes that leukoaraiosis is a complex phenotype that is influenced by genetic variation in several underlying biological processes, in part accounting for inability to predict and individual's leukoaraiosis volume with information regarding conventional and novel risk factors for arteriosclerosis.

In addition to having a significant main effect, the KIT tyrosine kinase receptor ligand (*KITLG*) shows context-dependent effects through interaction with homocysteine and with toll-like receptor 4 (*TLR4*), a mediator of immune response. Several other interactions also suggest a role for immune response and inflammation in the development of leukoaraiosis including gene-environment interactions between *IL28RA* (class II cytokine receptor) and small dense LDL size, *IL22RA1* (class II cytokine receptor) and coronary heart disease, and both *LTA4H* (leukotriene hydroxylase) and

PCSK9 (which plays a role in LDL receptor degredation) and homocysteine. Gene-gene interactions that support a role for immunity and inflammation in the disease process include an interaction between *IL22RA1* and *SLC20A1* (a receptor for retroviruses) and several interactions between an immune factor and a platelet factor such as that between *MPO* (myeloperoxidase, responsible for microbicidal activity) and platelet factor 8 (*F8*) and between *CX3CR1* (a cytokine for leukocytes) and platelet factor 2 (*F2*). An interaction between *NMUR1* (a G-protein coupled activator that appears to be involved in regulation of food intake) and *GPR55* (a G-protein coupled receptor) also points to genetic variation in signal transduction pathways playing a role in leukoaraiosis development.

Recent work has suggested a number of new potential cellular mechanisms (e.g. endothelial dysfunction, mitochondrial energy metabolism, protein transport) that may play a role in the development of leukoaraiosis and have not been previously considered in candidate gene selection (Szolnoki, 2007a; Szolnoki, 2007b). Several unexpected context-dependent effects have also been shown to consistently impact the leukoaraiosis phenotype (Szolnoki, 2006) In addition, animal and plant studies have shown more gene-gene (epistatic) interactions than previously expected (Cheverud, 2000). Given the biological complexity of the leukoaraiosis phenotype, it is not surprising that epistatic interactions and context-dependent effects play a large role and explain a larger proportion of variation in the phenotype than single covariate or SNP effects alone in this study. In accordance with this notion, multiple variable predictive modeling that was performed with the most highly ranked single SNP associations, SNP-covariate

interactions, and SNP-SNP interactions shows that the variation explained by the SNP-covariate and SNP-SNP interactions (19.18%, CV R^2x100 = 11.83%) was much higher than that explained by the main effects of these variables alone (7.18%, R^2x100 = 1.30%).

Failure to find replicated SNP effects across studies has significantly limited the utility of genetic association results. Manly suggests that internal validation methods, such as cross-validation, can be implemented as one way to avoid false positives (Manly, 2005). Cross-validation is an established method for discriminating between true associations and false positives that is based on predictive performance in independent test cases (Stone, 1974), and it has been used in a number of fields that deal with high-dimensional "omics" data (Pohjanen, 2007; Agranoff, 2006; Wood, 2007; Mertens, 2006). Another popular method for reducing false positive associations is to control the false discovery rate, for example, using Storey's q-value (Storey, 2002). There is a relatively low level of agreement between results filtered through different methods of reducing false positives (FDR q-value < 0.3, internal replication, and cross-validation), emphasizing the need for multiple false positive reduction methods.

Our study has several limitations. The design of the study is based on the premise that susceptibility alleles for common diseases are not under strong selective pressures and are relatively abundant in the population (i.e., the "common disease, common variant" hypothesis). Since the entire allelic spectrum for genes associated with quantitative measures of leukoaraiosis has not been fully delineated, our study was limited to

candidate gene choices based on physiological and biological knowledge of leukoaraiosis. In addition, it is possible that multiple rare polymorphisms in the positional and biological candidate genes we studied also influenced the phenotype; however, this study was underpowered to detect this type of effect. Since this study was conducted in a cohort of primarily hypertensive non-Hispanic white adults, the inferences may not be generalizable to individuals who are younger, normotensive, or of other ethnicities. Despite these limitations, our approach illustrates the use of candidate genes to formulate a more realistic picture of the genetic architecture of complex traits such as leukoaraiosis.

Conclusions

The genetic architecture of complex traits such as leukoaraiosis, a marker of increased risk of stroke and dementia, is comprised of SNP and covariate main effects, gene-gene interactions, and gene-environment interactions from a variety of biological pathways. Our findings indicate that systematic investigation of the context-dependent effects of genetic variation is critical for a more thorough understanding of the multidimensional architecture of complex diseases.

Table 2.1. Descriptive statistics for study participants

	Full Sample		Subset 1		Subset 2		Association with Raw Ln(leuko) in Full Sample	Association with Adj. Ln(leuko) in Full Sample
	N	Mean (±SD)	N	Mean (±SD)	N	Mean (±SD)	β estimate	β estimate
Leukoaraiosis volume (cm^3)	777	7.80 (6.31)	316	7.90 (6.18)	316	7.95 (6.49)	NA	NA
Age, years	777	59.7 (10.1)	316	59.9 (10.4)	316	59.9 (9.7)	0.031***	NA
BMI, kg/m^2	777	30.5 (5.8)	316	30.3 (5.8)	316	30.8 (5.7)	-0.006	-0.003
Height, cm	777	168 (9.2)	316	169 (9.6)	316	169 (9.4)	-0.0007	-0.0003
Weight, kg	777	86.3 (18.6)	316	86.9 (18.8)	316	87.9 (18.5)	-0.002	-0.0009
Waist-to-hip ratio	777	0.91 (0.11)	316	0.92 (0.12)	316	0.92 (0.10)	0.899***	0.2162
SBP, mm Hg	776	131.4 (16.5)	316	131.8 (17.0)	315	131.6 (16.3)	0.009***	0.0015
DBP, mm Hg	776	74.0 (9.0)	316	73.9 (8.6)	315	74.5 (9.1)	-0.0008	0.0052**
Pulse pressure, mm Hg	776	57.4 (15.3)	316	57.9 (15.8)	315	57.1 (15.0)	0.0102***	8.1 E-06
Total cholesterol, mg/dL	777	197.4 (33.7)	316	193.4 (33.4)	316	197.4 (32.3)	-0.0009	-0.0008
Triglycerides (log), mg/dL	777	4.93 (0.52)	316	4.92 (0.54)	316	4.99 (0.48)	0.0478	0.0354
HDL cholesterol, mg/dL	777	51.9 (14.6)	316	51.4 (14.8)	316	50.0 (13.6)	-0.0018	0.0007
LDL cholesterol, mg/dL	777	120.1 (31.9)	316	116.9 (31.7)	316	121.1 (31.5)	-0.0007	-0.0011*
LDL particle size, Å	777	270.2 (5.0)	316	270.3 (5.1)	316	269.8 (4.8)	0.0061	-0.0041
Lp(a), mg/dL	777	2.71 (1.21)	316	2.76 (1.17)	316	2.63 (1.22)	0.0107	-0.0047
C-reactive protein (log), mg/L	775	-1.33 (0.97)	314	-1.41 (0.91)	316	-1.30 (0.98)	0.0096	-0.0015
Fibrinogen, mg/dL	772	319.9 (76.1)	314	316.8 (71.4)	314	323.5 (77.2)	0.0006*	9.9 E-05
Homocysteine (log), µmol/L	777	2.25 (0.25)	316	2.26 (0.25)	316	2.26 (0.25)	0.455***	0.0499

*p<0.05; ** p<0.01; *** p<0.001

BMI, body mass index; SBP, systolic blood pressure; DBP, diastolic blood pressure; HDL, high density lipoprotein; LDL, low density lipoprotein; Lp(a), lipoprotein A

The adjustment model for ln(leukoaraiosis) included age, sex, and total brain volume.

Table 2.2. Quantitative summary of genetic associations with leukoaraiosis[1] that replicated internally, cross-validated, and passed FDR criterion

	SNP Main Effects	SNP-Covariate Interactions	SNP-SNP Interactions
Number of tests in full sample	1649	10625	96053
Model p<0.10 on full sample	286	1344	12673
FDR (q<0.30) on full sample	37	71	1561
Replication (model p<0.10 in both subsets)	15	103	834
Cross-validation (CV R^2>0.005) on full sample	23	189	1887
FDR and replication	7	20	281
FDR and cross-validation	22	39	763
Replication and cross-validation	4	30	249
FDR and replication and cross-validation[2]	4	12	173

[1] For all associations, the outcome was leukoaraiosis volume transformed using the natural logarithm and adjusted for age, sex, and total brain volume.
[2] All associations remained significant (p-value<0.1) in a linear mixed effects model with family as a random effect

Table 2.3. Genetic effects that replicated internally, cross-validated, and passed FDR criterion

Main Effects (4)	SNP	Subset 1 p-value	Subset 2 p-value	Full Sample p-value	Adj R^2x100	CV R^2x100
	F3_rs3917643	0.0477	0.0270	0.0021	1.58	1.05
	KITLG_rs995029	0.0001	0.0921	0.0001	2.33	0.96
	CAPN10_rs7571442	0.0856	0.0318	0.0021	1.70	0.59
	MMP2_rs9928731	0.0161	0.0383	0.0032	1.48	0.56

SNP-Covariate Interactions (12)	SNP	Covariate	Subset 1 p-value	Subset 2 p-value	Full Sample p-value	Adj R^2x100	CV R^2x100
	KITLG_rs1492347	Log Homocysteine	0.0002	0.0640	0.0003	4.37	1.57
	ITGB3_rs3851806	Height	0.0286	0.0101	0.0010	2.54	1.51
	TGFB3_rs2284791	Log C Reactive Protein	0.0073	0.0173	0.0003	2.92	1.51
	TGFB3_rs2284791	Physical Activity	0.0885	0.0015	0.0007	2.69	1.46
	TGFB3_rs2268622	Log C Reactive Protein	0.0026	0.0360	0.0003	2.86	1.42
	KITLG_rs995029	Log Homocysteine	0.0002	0.0621	0.0003	4.42	1.40
	IL28RA_rs11587500	Small dense LDL Size	0.0026	0.0014	0.0004	2.79	1.26
	ACCN4_rs1872858	Fibrinogen	0.0494	0.0339	0.0010	2.59	1.24
	LTA4H_rs17025079	Log Homocysteine	0.0006	0.0709	0.0001	3.41	1.16
	PCSK9_rs10888896	Log Homocysteine	0.0239	0.0567	0.0009	2.33	1.13
	IL22RA1_rs3795299	CHD	0.0859	0.0138	0.0009	2.41	0.70
	SERPINE1_rs2227672	Body Mass Index	0.0113	0.0219	0.0003	2.69	0.54

SNP-SNP Interactions (20 of 173)	SNP1	SNP2	Subset 1 p-value	Subset 2 p-value	Full Sample p-value	Adj R^2x100	CV R^2x100
	RHAG_rs11759060	*GLS_rs1921913*	0.0032	0.0225	8.0 E-06	5.49	2.61
	F8 _rs7053448	*MPO_rs34704261*	0.0354	0.0279	0.0001	4.10	2.58
	RHAG_rs11759060	*GLS_rs3771316*	0.0032	0.0225	8.0 E-06	5.49	2.55
	MPO_rs34704261	*F8_rs1800291*	0.0362	0.0279	0.0001	4.10	2.44
	SLC20A1_rs10758	*IL22RA1_rs12093987*	0.0439	0.0085	4.0 E-05	5.44	2.35

SLC20A1_rs3827758	*IL22RA1_rs12093987*	0.0315	0.0915	0.0004	3.88	2.27
KITLG_rs995029	*TLR4_rs1927911*	0.0035	0.0586	5.0 E-05	5.58	2.27
SLC20A1_rs1053652	*IL22RA1_rs12093987*	0.0237	0.0045	5.0 E-05	5.07	2.22
MPO_rs34704261	*F8_rs4898399*	0.0433	0.0315	0.0002	4.18	2.11
RHAG_rs2518100	*GLS_rs1921913*	0.0029	0.0534	0.0001	4.50	2.11
MMP2_rs243834	*IL28RA_rs4330872*	0.0136	0.0856	0.0002	4.46	2.09
RHAG_rs2518100	*GLS_rs3771316*	0.0029	0.0534	0.0001	4.50	1.95
MPO_rs8077532	*F8_rs1800291*	0.0417	0.0802	0.0003	3.93	1.93
NMUR1_rs10933376	*GPR55_rs2969126*	0.0945	0.0022	4.3 E-05	5.03	1.90
ACCN4_rs3770234	*TNFSF10_rs3136596*	0.0246	0.0007	0.0001	4.65	1.84
PRKAR2B_rs257376	*PKRAR2B_rs3729877*	0.0085	0.0027	4.8 E-05	3.55	1.84
CX3CR1_rs2853712	*F2_rs3136435*	0.0153	0.0480	0.0010	3.31	1.82
KITLG_rs1492347	*TLR4_rs10116253*	0.0049	0.0577	0.0001	5.44	1.81
F8_rs7053448	*MPO_rs8077532*	0.0417	0.0802	0.0003	3.93	1.81
MPO_rs8077532	*F8_rs4898399*	0.0515	0.0822	0.0006	4.04	1.81

For all associations, the outcome was leukoaraiosis volume transformed using the natural logarithm and adjusted for age, sex, and total brain volume.

ACCN4 amiloride-sensitive cation channel neuronal 4; *CAPN10* calpain 10 (cysteine protease); *CX3CR1* chemokine, CX3C motif, receptor 1; *F2* coagulation factor 2; *F3* coagulation factor 3; *F8* coagulation factor 8; GLS glutaminase, phosphate activated; *GPR55* G protein-coupled receptor 55; *IL22RA1* interleukin 22 receptor, alpha-1; *IL28RA* interleukin 28 receptor; *ITGB3* integrin, beta-3; *KITLG* KIT tyrosine kinase receptor ligand; *LTA4H* leukotriene A4 hydroxylase; *MMP2* matrix metalloproteinase 2 (72kDa type IV collagenase); *MPO* myeloperoxidase; *NMUR1* neuromedin U receptor 1 (G protein-coupled receptor 66); *PCSK9* proprotein convertase, subtilisin/kexin-type, 9; *PRKAR2B* protein kinase, cAMP-dependent, regulatory, type II, beta*; RHAG* rhesus blood group-associated glycoprotein; *SERPINE1* plasminogen activator inhibitor 1; *SLC20A1* solute carrier family 20 (phosphate transporters), member 1*; TGFB3* transforming growth factor, beta 3; *TLR4* toll-like receptor 4; *TNFSF10* tumor necrosis factor ligand superfamily, member 10

Table 2.4. Multivariable analysis to assess combined predictive ability of the best SNPs, risk factors, and interactions

Model		Adj R^2x100	CV R^2x100
1. Top Single SNPs		5.99	3.72
2. Top 4 SNP*Covariate Interactions			
	Full Model	7.88	4.53
	Reduced Model	2.64	-0.06
	Difference*		4.53
3. Top 4 SNP*SNP Interactions			
	Full Model	14.73	9.59
	Reduced Model	6.12	2.27
	Difference		7.32
4. Top 4 SNP*Covariate Interactions + Top 4 SNP*SNP Interactions			
	Full Model	19.18	11.83
	Reduced Model	7.18	1.30
	Difference		10.80
5. Single SNPs + Top 4 SNP*Covariate Interactions + Top 4 SNP*SNP Interactions			
	Full Model	21.32	11.60
	Reduced Model	9.99	2.16
	Difference		9.34

* In calculating the difference in CV R^2 between full and reduced models, the CV R^2 of the reduced model was considered to be zero if it had a negative value.
For all associations, the outcome was leukoaraiosis volume transformed using the natural logarithm and adjusted for age, sex, and total brain volume.

1. Model: *F3_rs3917643* + *KITLG_rs995029* + *CAPN10_rs7571442* + *MMP2_rs9928731*
2. Full model: *KITLG_rs1492347**Log Homocysteine + *ITGB3_rs3851806**Height + *TGFB3_rs2284791**Log CCRP + *TGFB3_rs2284791**Physical Activity; Reduced model: *KITLG_rs1492347* + Log Homocysteine + *ITGB3_rs3851806* + Height + *TGFB3_rs2284791* + Log CCRP + Physical Activity
3. Full model: *RHAG_rs11759060*GLS_rs1921913* + *F8_rs7053448*MPO_rs34704261* + *SLC20A1_rs10758*IL22RA1_rs12093987* + *KITLG_rs995029*TLR4_rs1927911*; Reduced model: *RHAG_rs11759060* + *GLS_rs1921913* + *F8_rs7053448* + *MPO_rs34704261* + *SLC20A1_rs10758* + *IL22RA1_rs12093987* + *KITLG_rs995029* + *TLR4_rs1927911*
4. Full model: *KITLG_rs1492347**Log Homocysteine + *ITGB3_rs3851806**Height + *TGFB3_rs2284791**Log CCRP + *TGFB3_rs2284791**Physical Activity + *RHAG_rs11759060*GLS_rs1921913* + *F8_rs7053448*MPO_rs34704261* + *SLC20A1_rs10758*IL22RA1_rs12093987* + *KITLG_rs995029*TLR4_rs1927911*; Reduced model: *KITLG_rs1492347* + Log Homocysteine + *ITGB3_rs3851806* + Height + *TGFB3_rs2284791* + Log CCRP + Physical Activity + *RHAG_rs11759060* + *GLS_rs1921913* + *F8_rs7053448* + *MPO_rs34704261* + *SLC20A1_rs10758* + *IL22RA1_rs12093987* + *KITLG_rs995029* + *TLR4_rs1927911*
5. Full model: *F3_rs3917643* + *KITLG_rs995029* + *CAPN10_rs7571442* + *MMP2_rs9928731* + *KITLG_rs1492347**Log Homocysteine + *ITGB3_rs3851806**Height + *TGFB3_rs2284791**Log CCRP + *TGFB3_rs2284791**Physical Activity + *RHAG_rs11759060*GLS_rs1921913* + *F8_rs7053448*MPO_rs34704261* + *SLC20A1_rs10758*IL22RA1_rs12093987* + *KITLG_rs995029*TLR4_rs1927911*; Reduced model: *F3_rs3917643* + *KITLG_rs995029* + *CAPN10_rs7571442*+ *MMP2_rs9928731*+ *KITLG_rs1492347* + Log Homocysteine + *ITGB3_rs3851806* + Height + *TGFB3_rs2284791* + Log CCRP + Physical Activity + *RHAG_rs11759060* + *GLS_rs1921913* + *F8_rs7053448* + *MPO_rs34704261* + *SLC20A1_rs10758* + *IL22RA1_rs12093987* + *TLR4_rs1927911*

Figure 2.1. The distribution of leukoaraiosis volume in GENOA-GMBI study participants

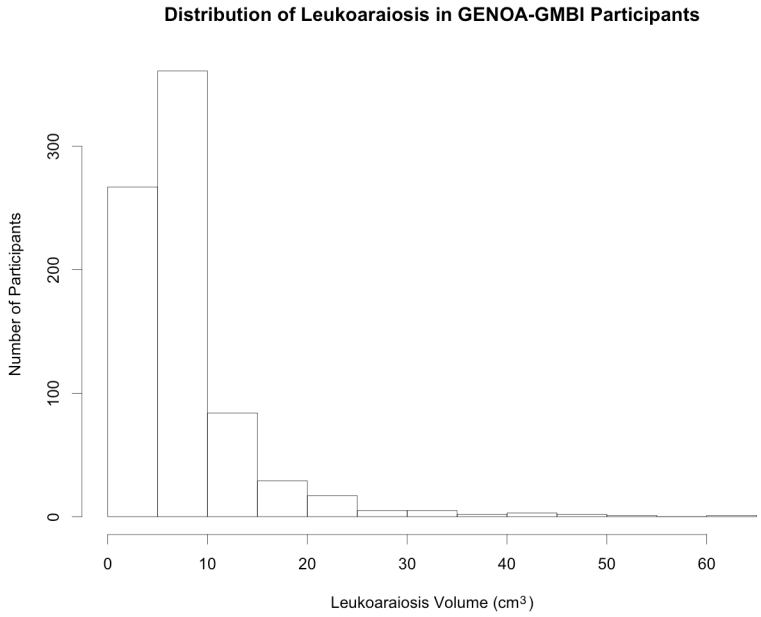
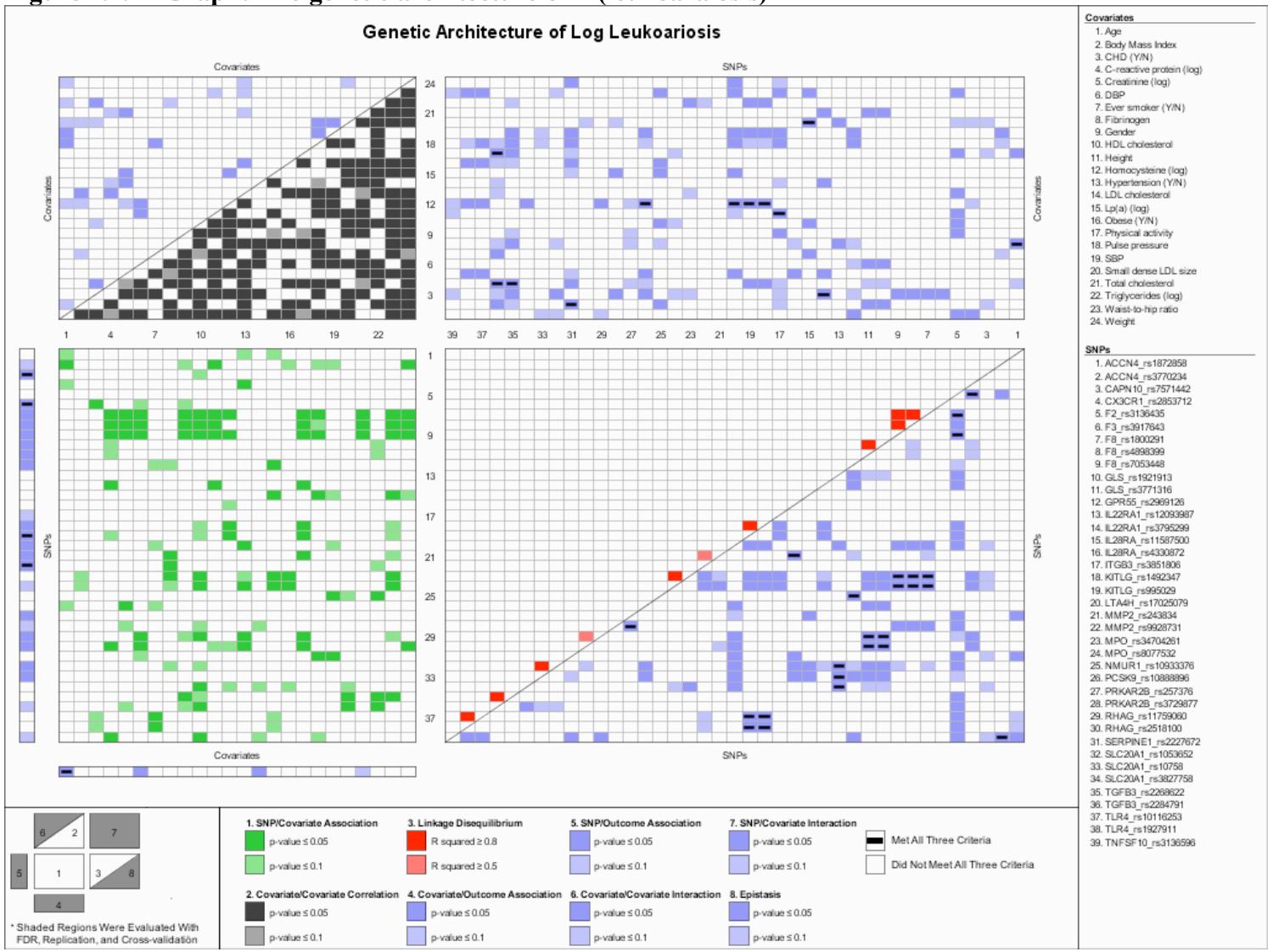


Figure 2.2. KGraph: The genetic architecture of ln(leukoaraiosis)



Appendix 2.1. Summary of SNPs included in the KGraph

SNP	Chromosome	Position	SNP Type	Call Rate	Major Allele	Minor Allele	MAF	HWE p-value
ACCN4_rs1872858	2	220205630	intron	0.942	A	G	0.397	0.731
ACCN4_rs3770234	2	220217245	intron	0.949	C	T	0.434	0.576
CAPN10_rs7571442	2	241275981	intron	0.925	G	T	0.280	0.506
CX3CR1_rs2853712	3	39293292	intron	0.997	A	G	0.440	0.458
F2_rs3136435	11	46699002	intron	0.999	G	A	0.078	0.417
F3_rs3917643	1	94774455	intron	1.000	A	G	0.051	0.366
F8_rs1800291	X	153811479	coding-nonsynonymous	0.999	C	G	0.146	0.000
F8_rs4898399	X	153740562	intron	0.999	T	A	0.131	0.000
F8_rs7053448	X	153846405	intron	1.000	A	G	0.145	0.000
GLS_rs1921913	2	191457788	intron	1.000	G	C	0.076	0.053
GLS_rs3771316	2	191457280	intron	1.000	A	C	0.076	0.053
GPR55_rs2969126	2	231603056	UTR	0.952	C	G	0.097	0.758
IL22RA1_rs12093987	1	24322215	intron	0.999	A	G	0.313	0.386
IL22RA1_rs3795299	1	24320055	coding-nonsynonymous	0.999	G	C	0.410	0.736
IL28RA_rs11587500	1	24389467	unknown	1.000	G	A	0.475	0.279
IL28RA_rs4330872	1	24348957	unknown	0.999	A	G	0.133	0.802
ITGB3_rs3851806	17	42705918	intron	0.999	C	G	0.160	0.716
KITLG_rs1492347	12	87450104	intron	1.000	G	A	0.100	0.069
KITLG_rs995029	12	87414652	untranslated	1.000	A	G	0.098	0.060
LTA4H_rs17025079	12	94943056	intron	1.000	G	A	0.023	1.000
MMP2_rs243834	16	54094188	intron	1.000	A	G	0.463	0.914
MMP2_rs9928731	16	54080512	intron	0.999	A	G	0.443	0.111
MPO_rs34704261	17	53708428	unknown	0.999	T	A	0.120	0.485
MPO_rs8077532	17	53701294	locus	0.997	A	G	0.117	0.136
NMUR1_rs10933376	2	232220024	intron	0.960	G	A	0.195	0.719
PCSK9_rs10888896	1	55281801	intron	1.000	G	C	0.228	0.886
PRKAR2B_rs257376	7	106393948	synonymous	0.970	A	G	0.446	0.590
PRKAR2B_rs3729877	7	106381016	intron	0.987	T	G	0.199	0.724
RHAG_rs11759060	6	49694375	intron	1.000	G	A	0.279	0.233
RHAG_rs2518100	6	49717093	unknown	0.996	A	C	0.379	0.735
SERPINE1_rs2227672	7	100562406	intron	1.000	C	A	0.133	0.522

SNP	Chromosome	Position	SNP Type	Call Rate	Major Allele	Minor Allele	MAF	HWE p-value
SLC20A1_rs1053652	2	113126068	intron	0.988	G	C	0.269	0.220
SLC20A1_rs10758	2	113126246	intron	0.916	A	G	0.267	0.088
SLC20A1_rs3827758	2	113131217	intron	0.861	G	A	0.424	0.015
TGFB3_rs2268622	14	75506191	intron	0.954	A	G	0.186	0.148
TGFB3_rs2284791	14	75498789	intron	0.952	C	G	0.174	0.344
TLR4_rs10116253	9	119504141	unknown	0.999	A	G	0.235	0.882
TLR4_rs1927911	9	119509875	intron	0.999	G	A	0.236	0.884
TNFSF10_rs3136596	3	173711566	intron	1.000	G	A	0.200	0.746

Appendix 2.2. SNP-SNP (epistatic) interactions that passed all filters

SNP 1	SNP 1 Locus	SNP 2	SNP 2 Locus	Subset 1 p-value	Subset 2 p-value	Full Sample p-value	R^2	CV R^2	FDR q-value
rs11759060	RHAG	rs1921913	GLS	0.0032	0.0225	0.0000	0.0549	0.0261	0.0164
rs7053448	F8	rs34704261	MPO	0.0354	0.0279	0.0001	0.0410	0.0258	0.0498
rs11759060	RHAG	rs3771316	GLS	0.0032	0.0225	0.0000	0.0549	0.0255	0.0164
rs34704261	MPO	rs1800291	F8	0.0362	0.0279	0.0001	0.0410	0.0244	0.0498
rs10758	SLC20A1	rs12093987	IL22RA1	0.0439	0.0085	0.0000	0.0544	0.0235	0.0330
rs3827758	SLC20A1	rs12093987	IL22RA1	0.0315	0.0915	0.0004	0.0388	0.0227	0.1104
rs995029	KITLG	rs1927911	TLR4	0.0035	0.0586	0.0000	0.0558	0.0227	0.0371
rs1053652	SLC20A1	rs12093987	IL22RA1	0.0237	0.0045	0.0000	0.0507	0.0222	0.0184
rs34704261	MPO	rs4898399	F8	0.0433	0.0315	0.0002	0.0418	0.0211	0.0757
rs2518100	RHAG	rs1921913	GLS	0.0029	0.0534	0.0001	0.0450	0.0211	0.0537
rs243834	MMP2	rs4330872	IL28RA	0.0136	0.0856	0.0002	0.0446	0.0209	0.0746
rs2518100	RHAG	rs3771316	GLS	0.0029	0.0534	0.0001	0.0450	0.0195	0.0537
rs8077532	MPO	rs1800291	F8	0.0417	0.0802	0.0003	0.0393	0.0193	0.0942
rs10933376	NMUR1	rs2969126	GPR55	0.0945	0.0022	0.0000	0.0503	0.0190	0.0337
rs3770234	ACCN4	rs3136596	TNFSF10	0.0246	0.0007	0.0001	0.0465	0.0184	0.0498
rs257376	PRKAR2B	rs3729877	PRKAR2B	0.0085	0.0027	0.0000	0.0355	0.0184	0.0359
rs2853712	CX3CR1	rs3136435	F2	0.0153	0.0480	0.0010	0.0331	0.0182	0.1678
rs1492347	KITLG	rs10116253	TLR4	0.0049	0.0577	0.0001	0.0544	0.0181	0.0417
rs7053448	F8	rs8077532	MPO	0.0417	0.0802	0.0003	0.0393	0.0181	0.0942
rs8077532	MPO	rs4898399	F8	0.0515	0.0822	0.0006	0.0404	0.0181	0.1377
rs699	AGT	rs5987077	F8	0.0165	0.0407	0.0007	0.0385	0.0177	0.1490
rs4076050	ADM	rs1805087	MTR	0.0837	0.0224	0.0002	0.0350	0.0177	0.0884
rs3770234	ACCN4	rs3136597	TNFSF10	0.0305	0.0008	0.0001	0.0459	0.0176	0.0643
rs2069824	IL6	rs1545970	ACCN4	0.0030	0.0050	0.0001	0.0421	0.0175	0.0515
rs2069824	IL6	rs932307	SELE	0.0251	0.0337	0.0001	0.0444	0.0174	0.0663
rs3093265	F7	rs1492347	KITLG	0.0096	0.0098	0.0000	0.0588	0.0173	0.0255
rs3770234	ACCN4	rs9859413	TNFSF10	0.0311	0.0007	0.0001	0.0460	0.0170	0.0646
rs699	AGT	rs4898398	F8	0.0062	0.0665	0.0003	0.0396	0.0167	0.0904

SNP 1	SNP 1 Locus	SNP 2	SNP 2 Locus	Subset 1 p-value	Subset 2 p-value	Full Sample p-value	R^2	CV R^2	FDR q-value
rs1006502	SLC4A5	rs31563	IL9	0.0460	0.0929	0.0012	0.0374	0.0165	0.1791
rs3917211	TGFB3	rs4927193	PCSK9	0.0324	0.0458	0.0010	0.0327	0.0164	0.1681
rs3827758	SLC20A1	rs877064	PLA2G2E	0.0444	0.0277	0.0007	0.0417	0.0159	0.1449
rs11587500	IL28RA	rs12460421	CARM1	0.0021	0.0486	0.0002	0.0385	0.0159	0.0768
rs2069824	IL6	rs11835389	HAL;LTA4H	0.0223	0.0153	0.0011	0.0247	0.0157	0.1750
rs17025079	LTA4H	rs6686007	IL22RA1	0.0439	0.0287	0.0002	0.0396	0.0155	0.0855
rs7529929	TNFSF4	rs9651118	MTHFR	0.0448	0.0683	0.0005	0.0283	0.0154	0.1283
rs3771316	GLS	rs2180723	RHAG	0.0276	0.0245	0.0003	0.0453	0.0153	0.1011
rs1921913	GLS	rs2180723	RHAG	0.0276	0.0245	0.0003	0.0453	0.0153	0.1011
rs10779958	AUP1	rs2228552	COL16A1	0.0058	0.0491	0.0000	0.0533	0.0152	0.0191
rs995029	KITLG	rs10116253	TLR4	0.0036	0.0586	0.0001	0.0554	0.0152	0.0374
rs909177	DCTN1	rs2228552	COL16A1	0.0137	0.0128	0.0007	0.0415	0.0149	0.1449
rs6048519	THBD	rs2228552	COL16A1	0.0386	0.0046	0.0005	0.0380	0.0148	0.1211
rs3770234	ACCN4	rs3136602	TNFSF10	0.0311	0.0007	0.0001	0.0460	0.0148	0.0646
rs2853712	CX3CR1	rs3136456	F2	0.0153	0.0480	0.0010	0.0331	0.0147	0.1678
rs3917211	TGFB3	rs11583680	PCSK9	0.0286	0.0562	0.0013	0.0337	0.0145	0.1857
rs2069824	IL6	rs5950587	AGTR2	0.0997	0.0244	0.0036	0.0294	0.0145	0.2696
rs4149578	TNFRSF1A	rs1959053	BDKRB2	0.0224	0.0119	0.0003	0.0379	0.0145	0.0997
rs10758	SLC20A1	rs243831	MMP2	0.0091	0.0452	0.0006	0.0436	0.0145	0.1361
rs7529929	TNFSF4	rs2071045	LEP	0.0560	0.0055	0.0002	0.0317	0.0144	0.0884
rs3755065	ACCN4	rs3136596	TNFSF10	0.0568	0.0201	0.0004	0.0448	0.0143	0.1063
rs4236625	LEP	rs1703081	KITLG	0.0000	0.0744	0.0005	0.0326	0.0143	0.1211
rs1006502	SLC4A5	rs932307	SELE	0.0731	0.0398	0.0010	0.0427	0.0141	0.1678
rs17025079	LTA4H	rs3799675	RHAG	0.0592	0.0939	0.0007	0.0438	0.0140	0.1490
rs995029	KITLG	rs1155764	MMP1	0.0046	0.0799	0.0002	0.0585	0.0138	0.0757
rs1545970	ACCN4	rs12971616	CARM1	0.0027	0.0456	0.0012	0.0404	0.0136	0.1793
rs1006502	SLC4A5	rs899653	MARS	0.0000	0.0914	0.0000	0.0552	0.0133	0.0200
rs1545970	ACCN4	rs12977506	CARM1	0.0028	0.0456	0.0013	0.0403	0.0133	0.1873
rs2069824	IL6	rs8077532	MPO	0.0262	0.0078	0.0005	0.0322	0.0133	0.1310
rs3827758	SLC20A1	rs7531863	PLA2G2E	0.0508	0.0277	0.0007	0.0416	0.0132	0.1471

SNP 1	SNP 1 Locus	SNP 2	SNP 2 Locus	Subset 1 p-value	Subset 2 p-value	Full Sample p-value	R^2	CV R^2	FDR q-value
rs13198157	RHAG	rs1921913	GLS	0.0523	0.0724	0.0017	0.0439	0.0130	0.2037
rs1799964	LTA;TNF	rs16963927	CCL5	0.0027	0.0259	0.0028	0.0296	0.0128	0.2444
rs4463047	FGB	rs3136456	F2	0.0015	0.0609	0.0010	0.0348	0.0128	0.1680
rs2853550	IL1B	rs11835389	LTA4H	0.0095	0.0505	0.0008	0.0262	0.0127	0.1600
rs2770381	AVP	rs11835389	LTA4H	0.0470	0.0889	0.0010	0.0294	0.0127	0.1683
rs1492347	KITLG	rs1927911	TLR4	0.0048	0.0577	0.0001	0.0548	0.0126	0.0410
rs3770234	ACCN4	rs1959053	BDKRB2;LOC730199	0.0867	0.0032	0.0023	0.0360	0.0125	0.2238
rs81663	GP1BA	rs11835389	HAL;LTA4H	0.0516	0.0042	0.0003	0.0313	0.0124	0.0895
rs2275905	SLC17A4	rs10012150	SPP1	0.0180	0.0559	0.0015	0.0427	0.0124	0.1955
rs2239484	ATP6V1B1	rs10933376	NMUR1	0.0278	0.0337	0.0048	0.0266	0.0123	0.2967
rs662	PON1	rs243834	MMP2	0.0578	0.0245	0.0011	0.0455	0.0123	0.1733
rs1799964	LTA;TNF	rs16971600	CCL5	0.0027	0.0259	0.0028	0.0296	0.0122	0.2444
rs10758	SLC20A1	rs3136594	TNFSF10	0.0055	0.0510	0.0046	0.0399	0.0122	0.2915
rs699	AGT	rs5987079	F8	0.0145	0.0693	0.0015	0.0374	0.0121	0.1934
rs4149579	TNFRSF1A	rs2069885	IL9	0.0580	0.0011	0.0001	0.0394	0.0121	0.0498
rs1053652	SLC20A1	rs231983	TNFSF10	0.0521	0.0220	0.0008	0.0402	0.0120	0.1631
rs877064	PLA2G2E	rs5945128	F8	0.0944	0.0109	0.0041	0.0364	0.0120	0.2794
rs3755065	ACCN4	rs9859413	TNFSF10	0.0574	0.0201	0.0011	0.0426	0.0120	0.1764
rs17833353	SLC9A2	rs6425744	FABP3	0.0984	0.0188	0.0031	0.0370	0.0119	0.2566
rs4463047	FGB	rs439401	APOE;APOC1	0.0996	0.0063	0.0021	0.0364	0.0119	0.2169
rs6919346	LPA	rs10912560	TNFSF4	0.0667	0.0107	0.0011	0.0300	0.0119	0.1733
rs3729877	PRKAR2B	rs12188950	PART1	0.0237	0.0930	0.0006	0.0385	0.0119	0.1325
rs7149602	PSMA6	rs4423707	CX3CR1	0.0933	0.0274	0.0035	0.0234	0.0118	0.2672
rs9657021	CYP11B1	rs3136594	TNFSF10	0.0049	0.0810	0.0018	0.0234	0.0118	0.2043
rs4149579	TNFRSF1A	rs17025079	LTA4H	0.0312	0.0675	0.0014	0.0350	0.0118	0.1909
rs1652507	LPA	rs3136456	F2	0.0206	0.0690	0.0001	0.0392	0.0118	0.0530
rs2740204	AVP	rs11835389	HAL;LTA4H	0.0431	0.0896	0.0009	0.0307	0.0116	0.1678
rs740387	ADD2	rs2146372	SERPINC1	0.0874	0.0319	0.0019	0.0310	0.0116	0.2067
rs699	AGT	rs2096362	F8	0.0145	0.0693	0.0015	0.0374	0.0116	0.1934
rs1492347	KITLG	rs1155764	MMP1	0.0043	0.0682	0.0002	0.0579	0.0115	0.0751

SNP 1	SNP 1 Locus	SNP 2	SNP 2 Locus	Subset 1 p-value	Subset 2 p-value	Full Sample p-value	R^2	CV R^2	FDR q-value
rs1652507	LPA	rs3136435	F2	0.0206	0.0690	0.0001	0.0392	0.0115	0.0530
rs1143627	IL1B	rs12977506	CARM1	0.0519	0.0663	0.0026	0.0381	0.0114	0.2370
rs822387	ADIPOQ	rs3861950	TNFSF4	0.0885	0.0307	0.0025	0.0347	0.0114	0.2320
rs3136598	TNFSF10	rs2182833	PCSK9	0.0019	0.0737	0.0041	0.0363	0.0114	0.2794
rs1800291	F8	rs3136435	F2	0.0575	0.0225	0.0017	0.0324	0.0113	0.2039
rs4463047	FGB	rs3136435	F2	0.0015	0.0609	0.0010	0.0348	0.0110	0.1680
rs6076016	THBD	rs2228552	COL16A1	0.0213	0.0331	0.0047	0.0293	0.0109	0.2935
rs3770234	ACCN4	rs231983	TNFSF10	0.0031	0.0843	0.0031	0.0329	0.0109	0.2578
rs233998	TNFSF10	rs11835389	LTA4H	0.0469	0.0584	0.0028	0.0208	0.0106	0.2451
rs3755065	ACCN4	rs3136597	TNFSF10	0.0570	0.0202	0.0011	0.0429	0.0105	0.1741
rs13198157	RHAG	rs3771316	GLS	0.0523	0.0724	0.0017	0.0439	0.0105	0.2037
rs740387	ADD2	rs16846561	SERPINC1	0.0874	0.0405	0.0022	0.0307	0.0103	0.2183
rs2268417	HMGA1L4	rs3795300	IL22RA1	0.0170	0.0276	0.0003	0.0414	0.0103	0.1010
rs10755578	LPA	rs10912560	TNFSF4	0.0393	0.0545	0.0042	0.0311	0.0103	0.2817
rs2110981	ADD2	rs11835389	LTA4H	0.0421	0.0017	0.0035	0.0285	0.0102	0.2661
rs243834	MMP2	rs1805087	MTR	0.0327	0.0958	0.0031	0.0380	0.0102	0.2542
rs4236625	LEP	rs1798011	KITLG	0.0000	0.0744	0.0005	0.0326	0.0102	0.1211
rs2069824	IL6	rs34704261	MPO	0.0289	0.0061	0.0013	0.0282	0.0100	0.1865
rs7586970	TFPI	rs11835389	LTA4H	0.0466	0.0084	0.0021	0.0263	0.0100	0.2169
rs3755065	ACCN4	rs12971616	CARM1	0.0059	0.0721	0.0035	0.0437	0.0099	0.2661
rs5186	AGTR1	rs4795095	CCL5	0.0589	0.0303	0.0018	0.0399	0.0098	0.2046
rs2975766	CAPN10	rs697221	DDIT3	0.0436	0.0677	0.0036	0.0347	0.0098	0.2696
rs1545970	ACCN4	rs12093987	IL22RA1	0.0104	0.0568	0.0030	0.0336	0.0097	0.2535
rs243832	MMP2	rs1805087	MTR	0.0376	0.0870	0.0026	0.0394	0.0097	0.2361
rs943580	AGT	rs9987222	PLEKHA2	0.0054	0.0412	0.0026	0.0327	0.0097	0.2370
rs1472899	KITLG	rs1076669	ECE1	0.0085	0.0139	0.0007	0.0353	0.0096	0.1513
rs7531863	PLA2G2E	rs5945128	F8	0.0963	0.0109	0.0041	0.0363	0.0096	0.2794
rs2969126	GPR55	rs2192852	MMP2	0.0181	0.0834	0.0021	0.0419	0.0096	0.2151
rs1053652	SLC20A1	rs3136594	TNFSF10	0.0060	0.0975	0.0031	0.0340	0.0095	0.2542
rs17025079	LTA4H	rs2180723	RHAG	0.0325	0.0247	0.0014	0.0402	0.0095	0.1913

SNP 1	SNP 1 Locus	SNP 2	SNP 2 Locus	Subset 1 p-value	Subset 2 p-value	Full Sample p-value	R^2	CV R^2	FDR q-value
rs31563	IL9	rs11121820	AGTRAP	0.0073	0.0206	0.0003	0.0383	0.0093	0.0942
rs7760223	RHAG	rs17025079	LTA4H	0.0236	0.0388	0.0008	0.0405	0.0093	0.1592
rs7053448	F8	rs3136456	F2	0.0572	0.0225	0.0017	0.0323	0.0092	0.2039
rs11121819	AGTRAP	rs31563	IL9	0.0073	0.0206	0.0003	0.0383	0.0091	0.0942
rs1165165	SLC17A3	rs1549926	CARM1	0.0000	0.0991	0.0000	0.0571	0.0089	0.0147
rs2182833	PCSK9	rs6783667	TNFSF10	0.0019	0.0737	0.0041	0.0363	0.0088	0.2794
rs3136456	F2	rs16861205	ADIPOQ	0.0005	0.0224	0.0004	0.0328	0.0087	0.1196
rs4073489	AGTRAP	rs31563	IL9	0.0089	0.0439	0.0009	0.0347	0.0087	0.1678
rs4075034	AGTRAP	rs31563	IL9	0.0071	0.0200	0.0003	0.0385	0.0087	0.1011
rs3755065	ACCN4	rs3136602	TNFSF10	0.0574	0.0201	0.0011	0.0426	0.0087	0.1764
rs4149579	TNFRSF1A	rs6838095	SPP1	0.0911	0.0808	0.0037	0.0397	0.0086	0.2711
rs4842625	KITLG	rs1076669	ECE1	0.0230	0.0137	0.0016	0.0355	0.0085	0.1997
rs662	PON1	rs11868894	ITGB3	0.0125	0.0471	0.0019	0.0362	0.0084	0.2088
rs6437000	HTR2B	rs17025079	LTA4H	0.0459	0.0268	0.0008	0.0348	0.0084	0.1549
rs3755065	ACCN4	rs12977506	CARM1	0.0066	0.0721	0.0039	0.0436	0.0084	0.2771
rs4149579	TNFRSF1A	rs13144236	HTRA3	0.0617	0.0065	0.0018	0.0317	0.0082	0.2043
rs1076669	ECE1	rs2046971	KITLG	0.0085	0.0176	0.0009	0.0353	0.0082	0.1678
rs3795300	IL22RA1	rs13414554	GLS	0.0680	0.0016	0.0035	0.0259	0.0078	0.2688
rs1076669	ECE1	rs1000788	KITLG	0.0085	0.0176	0.0010	0.0354	0.0077	0.1678
rs2069824	IL6	rs1823227	TNFSF10	0.0868	0.0008	0.0023	0.0297	0.0077	0.2234
rs13414554	GLS	rs11868894	ITGB3	0.0812	0.0030	0.0027	0.0311	0.0077	0.2429
rs13198157	RHAG	rs17025079	LTA4H	0.0496	0.0481	0.0030	0.0403	0.0077	0.2515
rs10758	SLC20A1	rs231983	TNFSF10	0.0452	0.0813	0.0036	0.0405	0.0076	0.2695
rs3917643	F3	rs10487133	PON2	0.0344	0.0898	0.0041	0.0360	0.0076	0.2794
rs932307	SELE	rs12087657	CKS1B	0.0598	0.0161	0.0027	0.0380	0.0075	0.2415
rs2302453	PRKAR2B	rs2228552	COL16A1	0.0823	0.0202	0.0005	0.0395	0.0075	0.1310
rs2046971	KITLG	rs31563	IL9	0.0884	0.0310	0.0034	0.0307	0.0075	0.2634
rs1492347	KITLG	rs4236625	LEP	0.0000	0.0351	0.0000	0.0572	0.0075	0.0220
rs11888208	SCN7A	rs171336	GHRL	0.0452	0.0255	0.0015	0.0391	0.0074	0.1946
rs715407	PRSS25	rs2227672	SERPINE1	0.0019	0.0513	0.0000	0.0555	0.0074	0.0147

SNP 1	SNP 1 Locus	SNP 2	SNP 2 Locus	Subset 1 p-value	Subset 2 p-value	Full Sample p-value	R^2	CV R^2	FDR q-value
rs3749172	GPR35	rs10211925	IFNAR2	0.0248	0.0673	0.0043	0.0307	0.0071	0.2845
rs2284791	TGFB3	rs5987079	F8	0.0581	0.0056	0.0039	0.0397	0.0070	0.2771
rs11868894	ITGB3	rs3771310	GLS	0.0812	0.0033	0.0029	0.0309	0.0068	0.2485
rs16861205	ADIPOQ	rs3136435	F2	0.0005	0.0224	0.0004	0.0328	0.0067	0.1196
rs31563	IL9	rs17025079	LTA4H	0.0678	0.0596	0.0027	0.0291	0.0066	0.2407
rs12188950	PART1	rs243845	MMP2	0.0990	0.0376	0.0004	0.0418	0.0065	0.1159
rs10755578	LPA	rs9500	S100PBP	0.0515	0.0541	0.0018	0.0414	0.0065	0.2043
rs17025023	LTA4H	rs16861205	ADIPOQ	0.0000	0.0823	0.0000	0.0446	0.0064	0.0156
rs3093265	F7	rs995029	KITLG	0.0207	0.0188	0.0001	0.0568	0.0063	0.0503
rs909177	DCTN1	rs11835389	LTA4H	0.0762	0.0776	0.0039	0.0303	0.0063	0.2771
rs1006502	SLC4A5	rs697221	DDIT3	0.0000	0.0815	0.0002	0.0464	0.0063	0.0739
rs16861194	ADIPOQ	rs3136456	F2	0.0021	0.0502	0.0022	0.0267	0.0063	0.2183
rs3917365	IL1B	rs11835389	LTA4H	0.0104	0.0484	0.0029	0.0217	0.0061	0.2472
rs995029	KITLG	rs4236625	LEP	0.0000	0.0434	0.0000	0.0571	0.0061	0.0243
rs2069824	IL6	rs10758	SLC20A1	0.0871	0.0312	0.0048	0.0412	0.0060	0.2967
rs16861194	ADIPOQ	rs3136435	F2	0.0021	0.0502	0.0022	0.0267	0.0059	0.2183
rs1143627	IL1B	rs12971616	CARM1	0.0444	0.0663	0.0023	0.0382	0.0057	0.2255
rs877064	PLA2G2E	rs1805087	MTR	0.0087	0.0506	0.0013	0.0359	0.0055	0.1901
rs2072246	ADD2	rs4220_x	NA	0.0411	0.0989	0.0032	0.0376	0.0054	0.2602
rs2969126	GPR55	rs10751768	IL22RA1	0.0759	0.0684	0.0019	0.0348	0.0054	0.2067
rs1992188	GPR55	rs1892251	SLC17A4	0.0036	0.0701	0.0045	0.0448	0.0052	0.2897
rs2969126	GPR55	rs2146372	SERPINC1	0.0798	0.0044	0.0013	0.0379	0.0052	0.1861
rs1801253	ADRB1	rs7690786	HTRA3	0.0973	0.0694	0.0023	0.0332	0.0052	0.2234
rs7053448	F8	rs3136435	F2	0.0572	0.0225	0.0017	0.0323	0.0051	0.2039
rs31563	IL9	rs1000788	KITLG	0.0891	0.0310	0.0034	0.0307	0.0051	0.2653

Appendix 2.3. Reading a KGraph

Using 8 regions, the KGraph shows the relationships between the SNPs, covariates, and outcome by displaying the results from tests of correlation, linkage disequilibrium, association and cross-validation/FDR/Replication. The key at the bottom of the graphic shows the test criterion for each region and the colors associated with the test result. The region number key in the lower left corner shows the location of each region, and indicates whether the results in the region were assessed using FDR/Replication/Cross-validation (shaded regions). A black bar in the cell indicates that the association passed all three of these criteria. Region 1 displays the association between the SNPs and the covariates, region 2 displays the correlation between the covariates, and region 3 displays the linkage disequilibrium between the SNPs. Region 4 displays covariate association with leukoaraiosis, region 5 displays SNP association with leukoaraiosis, region 6 displays covariate-covariate interactions predicting leukoaraiosis, region 7 displays SNP-covariate interactions predicting leukoaraiosis, and region 8 displays SNP-SNP (epistatic) interactions predicting leukoaraiosis. Included on the KGraph are all of the covariates that were investigated in the study, SNPs that were involved in a single SNP or SNP-covariate association that passed all three filters, and SNPs that were involved in at least one of the 20 most predictive SNP-SNP interactions that passed all three filters.

References

Agranoff, D., Fernandez-Reyes, D., Papadopoulos, M.C., Rojas, S.A., Herbster, M., Loosemore, A., Tarelli, E., Sheldon, J., Schwenk, A., Pollok, R. et al. (2006). Identification of Diagnostic Markers for Tuberculosis by Proteomic Fingerprinting of Serum. *Lancet* **368(9540):** 1012-1021.

Atwood, L.D., Wolf, P.A., Heard-Costa, N.L., Massaro, J.M., Beiser, A., D'Agostino, R.B., and DeCarli, C. (2004). Genetic Variation in White Matter Hyperintensity Volume in the Framingham Study. *Stroke* **35(7):** 1609-1613.

Barkley, R.A., Chakravarti, A., Cooper, R.S., Ellison, R.C., Hunt, S.C., Province, M.A., Turner, S.T., Weder, A.B., Boerwinkle, E., and Family Blood Pressure Program. (2004). Positional Identification of Hypertension Susceptibility Genes on Chromosome 2. *Hypertension* **43(2):** 477-482.

Carmelli, D., DeCarli, C., Swan, G.E., Jack, L.M., Reed, T., Wolf, P.A., and Miller, B.L. (1998). Evidence for Genetic Variance in White Matter Hyperintensity Volume in Normal Elderly Male Twins. *Stroke* **29(6):** 1177-1181.

Cheverud, J.M. (2000). *Chapter 4.* In *Epistasis and the Evolutionary Process.* Oxford University Press: New York, NY, pp. 58-81.

Cupples, L.A., Arruda, H.T., Benjamin, E.J., D'Agostino RB, S., Demissie, S., DeStefano, A.L., Dupuis, J., Falls, K.M., Fox, C.S., Gottlieb, D.J. et al. (2007). The Framingham Heart Study 100K SNP Genome-Wide Association Study Resource: Overview of 17 Phenotype Working Group Reports. *BMC Med Genet* **8(Suppl 1):** S1.

Daniels, P.R., Kardia, S.L., Hanis, C.L., Brown, C.A., Hutchinson, R., Boerwinkle, E., Turner, S.T., and Genetic Epidemiology Network of Arteriopathy study. (2004). Familial Aggregation of Hypertension Treatment and Control in the Genetic Epidemiology Network of Arteriopathy (GENOA) Study. *Am J Med* **116(10):** 676-681.

Davie, E.W., Fujikawa, K., and Kisiel, W. (1991). The Coagulation Cascade: Initiation, Maintenance, and Regulation. *Biochemistry* **30(43):** 10363-10370.

FBPP Investigators. (2002). Multi-Center Genetic Study of Hypertension: The Family Blood Pressure Program (FBPP). *Hypertension* **39(1):** 3-9.

Flossmann, E., Schulz, U.G., and Rothwell, P.M. (2004). Systematic Review of Methods and Results of Studies of the Genetic Epidemiology of Ischemic Stroke. *Stroke* **35(1):** 212-227.

Friedewald, W.T., Levy, R.I., and Fredrickson, D.S. (1972). Estimation of the Concentration of Low-Density Lipoprotein Cholesterol in Plasma, without use of the Preparative Ultracentrifuge. *Clin Chem* **18(6):** 499-502.

Horikawa, Y., Oda, N., Cox, N.J., Li, X., Orho-Melander, M., Hara, M., Hinokio, Y., Lindner, T.H., Mashima, H., Schwarz, P.E. et al. (2000). Genetic Variation in the Gene Encoding Calpain-10 is Associated with Type 2 Diabetes Mellitus. *Nat Genet* **26(2):** 163-175.

Jack, C.R.,Jr, O'Brien, P.C., Rettman, D.W., Shiung, M.M., Xu, Y., Muthupillai, R., Manduca, A., Avula, R., and Erickson, B.J. (2001). FLAIR Histogram Segmentation for Measurement of Leukoaraiosis Volume. *J Magn Reson Imaging* **14(6):** 668-676.

Jack, C.R.,Jr, Twomey, C.K., Zinsmeister, A.R., Sharbrough, F.W., Petersen, R.C., and Cascino, G.D. (1989). Anterior Temporal Lobes and Hippocampal Formations: Normative Volumetric Measurements from MR Images in Young Adults. *Radiology* **172(2):** 549-554.

Keevil, B.G., Nicholls, S.P., and Kilpatrick, E.S. (1998). Evaluation of a Latex-Enhanced Immunoturbidimetric Assay for Measuring Low Concentrations of C-Reactive Protein. *Ann Clin Biochem* **35(Pt 5):** 671-673.

Kelly, R.J., Jacobsen, D.M., Sun, Y.V., Smith, J.A., and Kardia, S.L. (2007). KGraph: A System for Visualizing and Evaluating Complex Genetic Associations. *Bioinformatics* **23(2):** 249-251.

Kleinbaum, D., Kupper, L., Muller, K., and Nizam, A. (1998). *Applied Regression Analysis and Other Multivariate Methods.* Brooks/Cole Publishing Company: Pacific Grove, CA.

Kuller, L.H., Longstreth, W.T.,Jr, Arnold, A.M., Bernick, C., Bryan, R.N., Beauchamp, N.J.,Jr, and Cardiovascular Health Study Collaborative Research Group. (2004). White Matter Hyperintensity on Cranial Magnetic Resonance Imaging: A Predictor of Stroke. *Stroke* **35(8):** 1821-1825.

Kullo, I.J., Bailey, K.R., Bielak, L.F., Sheedy, P.F.,2nd, Klee, G.G., Kardia, S.L., Peyser, P.A., Boerwinkle, E., and Turner, S.T. (2004a). Lack of Association between Lipoprotein(a) and Coronary Artery Calcification in the Genetic Epidemiology Network of Arteriopathy (GENOA) Study. *Mayo Clin Proc* **79(10):** 1258-1263.

Kullo, I.J., Bailey, K.R., McConnell, J.P., Peyser, P.A., Bielak, L.F., Kardia, S.L., Sheedy, P.F.,2nd, Boerwinkle, E., and Turner, S.T. (2004b). Low-Density Lipoprotein Particle Size and Coronary Atherosclerosis in Subjects Belonging to Hypertensive Sibships. *Am J Hypertens* **17(9):** 845-851.

Liao, D., Cooper, L., Cai, J., Toole, J.F., Bryan, N.R., Hutchinson, R.G., and Tyroler, H.A. (1996). Presence and Severity of Cerebral White Matter Lesions and Hypertension, its Treatment, and its Control. the ARIC Study. Atherosclerosis Risk in Communities Study. *Stroke* **27(12):** 2262-2270.

Liu, P., Sun, M., and Sader, S. (2006). Matrix Metalloproteinases in Cardiovascular Disease. *Can J Cardiol* **22 (Suppl B):** 25B-30B.

Lynch, M., and Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits.* Sinauer Associates, Inc.: Sunderland, MA.

Manly, K.F. (2005). Reliability of Statistical Associations between Genes and Disease. *Immunogenetics* **57(8):** 549-558.

Martin, F.H., Suggs, S.V., Langley, K.E., Lu, H.S., Ting, J., Okino, K.H., Morris, C.F., McNiece, I.K., Jacobsen, F.W., and Mendiaz, E.A. (1990). Primary Structure and Functional Expression of Rat and Human Stem Cell Factor DNAs. *Cell* **63(1):** 203-211.

Mertens, B.J., De Noo, M.E., Tollenaar, R.A., and Deelder, A.M. (2006). Mass Spectrometry Proteomic Diagnosis: Enacting the Double Cross-Validatory Paradigm. *J Comput Biol* **13(9):** 1591-1605.

Molinaro, A.M., Simon, R., and Pfeiffer, R.M. (2005). Prediction Error Estimation: A Comparison of Resampling Methods. *Bioinformatics* **21(15):** 3301-3307.

Pantoni, L., and Garcia, J.H. (1997). Pathogenesis of Leukoaraiosis: A Review. *Stroke* **28(3):** 652-659.

Pohjanen, E., Thysell, E., Jonsson, P., Eklund, C., Silfver, A., Carlsson, I.B., Lundgren, K., Moritz, T., Svensson, M.B., and Antti, H. (2007). A Multivariate Screening Strategy for Investigating Metabolic Effects of Strenuous Physical Exercise in Human Serum. *J Proteome Res* **6(6):** 2113-2120.

R Core Development Team. (2008). R: A Language and Environment for Statistical Computing.

Raudenbush, S.W., and Bryk, A.S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods.* Sage Publications, Inc: Thousand Oaks, CA.

Salerno, J.A., Murphy, D.G., Horwitz, B., DeCarli, C., Haxby, J.V., Rapoport, S.I., and Schapiro, M.B. (1992). Brain Atrophy in Hypertension. A Volumetric Magnetic Resonance Imaging Study. *Hypertension* **20(3):** 340-348.

Schwartz, G.L., Bailey, K.R., Mosley, T., Knopman, D.S., Jack, C.R.,Jr, Canzanello, V.J., and Turner, S.T. (2007). Association of Ambulatory Blood Pressure with Ischemic Brain Injury. *Hypertension* **49(6):** 1228-1234.

Seshadri, S., and Wolf, P.A. (2007). Lifetime Risk of Stroke and Dementia: Current Concepts, and Estimates from the Framingham Study. *Lancet Neurol* **6(12):** 1106-1114.

Stone, M. (1974). Cross-Validatory Choice and Assessment of Statistical Predictions. *J R Stat Soc* **36:** 111-147.

Storey, J.D. (2002). A Direct Approach to False Discovery Rates. *J.R.Stat.Soc.* **Series B, 64:** 479-498.

Szolnoki, Z. (2007a). Chemical Events Behind Leukoaraiosis: Medicinal Chemistry Offers New Insight into a Specific Microcirculation Disturbance in the Brain (a Chemical Approach to a Frequent Cerebral Phenotype). *Curr Med Chem* **14(9):** 1027-1036.

Szolnoki, Z. (2007b). Pathomechanism of Leukoaraiosis: A Molecular Bridge between the Genetic, Biochemical, and Clinical Processes (a Mitochondrial Hypothesis). *Neuromolecular Med* **9(1):** 21-33.

Szolnoki, Z., and Melegh, B. (2006). Gene-Gene and Gene-Environment Interplay Represent Specific Susceptibility for Different Types of Ischaemic Stroke and Leukoaraiosis. *Curr Med Chem* **13(14):** 1627-1634.

Turner, S.T., Jack, C.R., Fornage, M., Mosley, T.H., Boerwinkle, E., and de Andrade, M. (2004). Heritability of Leukoaraiosis in Hypertensive Sibships. *Hypertension* **43(2):** 483-487.

van Dijk, E.J., Breteler, M.M., Schmidt, R., Berger, K., Nilsson, L.G., Oudkerk, M., Pajak, A., Sans, S., de Ridder, M., Dufouil, C. et al. (2004). The Association between Blood Pressure, Hypertension, and Cerebral White Matter Lesions: Cardiovascular Determinants of Dementia Study. *Hypertension* **44(5):** 625-630.

von Clauss, A. (1957). Gerinnungsphysiologische Schnellmethode Zur Bestimmung Des Fibrinogens. *Acta Haematol* **17:** 237-246.

Weir, B.S. (1996). *Genetic Data Analysis II: Methods for Discrete Population Genetic Data.* Sinauer Associates: Sunderland, MA.

Wood, I.A., Visscher, P.M., and Mengersen, K.L. (2007). Classification Based upon Gene Expression Data: Bias and Precision of Error Rates. *Bioinformatics* **23(11):** 1363-1370.

Chapter 3

Genome-wide Association Study of SNPs Associated with Leukoaraiosis in non-Hispanic Whites and African Americans

Abstract

White matter hyperintensity on magnetic resonance imaging (MRI) of the brain, referred to as leukoaraiosis, is associated with increased risk of stroke and dementia. Hypertension may contribute to leukoaraiosis by accelerating the process of arteriosclerosis in the brain. Leukoaraiosis volume is highly heritable but shows significant individual variability, and the likely numerous genetic variants underlying this trait are not known. As part of the Genetics of Microangiopathic Brain Injury (GMBI) ancillary study of the Genetic Epidemiology Network of Arteriopathy (GENOA), 759 non-Hispanic whites and 553 African Americans (73.4% and 77.4% hypertensive, respectively) underwent brain MRI. A genome-wide association study (GWAS) of leukoaraiosis volume after adjustment for age, sex, population structure, and family structure of 666,271 single nucleotide polymorphisms (SNPs) in non-Hispanic whites and 760,699 SNPs in African Americans with a minor allele frequency (MAF) ≥ 0.01 revealed several interesting findings. Five SNPs in non-Hispanic whites and nine SNPs in African Americans had a p-value less than 1.0×10^{-5}, with SNPs in *THSD7B* and *ANGPT4* showing the strongest association in whites and *CHCHD9* and *KCNMA1* showing the strongest association in

African Americans. Meta-analysis of GWAS results from African Americans in the Atherosclerosis Risk in Communities (ARIC) study was also conducted before and after removing participants with one or two copies of the *APOE* ε4, a known risk factor for cardiovascular disease and dementia. Two SNPs in an intergenic region between 44kb and 70kb upstream from the *ITGB1* gene showed the strongest association in the meta-analysis (rs9299702, p-value $3.7x10^{-8}$ and rs7898823, p-value $3.2x10^{-7}$).

Introduction

Hypertension affects approximately 1 in 3 American adults, and the majority of individuals that are diagnosed do not have their blood pressure adequately controlled (Roger, 2011). Ischemic damage to the vasculature of the brain as a result of uncontrolled or poorly controlled hypertension leads to clinical endpoints including stroke and dementia, which are significant causes of physical and cognitive disability, mortality, and economic burden (Roger, 2011; Schmidt, 2004; Bornstein, 2006; Pantoni, 2007). Areas of ischemic damage due to hypertension-related small vessel disease can be visualized on brain magnetic resonance imaging (MRI), with affected areas appearing as bright spots known as white matter hyperintensities or leukoaraiosis (O'Sullivan, 2008; Markus, 2008). The extent of leukoaraiosis is thought to be a marker of the severity of cerebral small vessel disease, ranging from small, distinct areas to large regions of diffuse white matter hyperintensity (O'Sullivan, 2008).

Leukoaraiosis is a subclinical phenotype that is rarely used in routine clinical practice before the appearance of acute symptoms leading to diagnosis of stroke or neurocognitive

evidence of cognitive decline indicative of dementia. However, asymptomatic leukoaraiosis is a strong risk factor for ischemic stroke, recurrent stroke, and vascular dementia even after adjustment for other risk factors, including hypertension (Markus, 2005; Kuller, 2004; Prins, 2004; Kuller, 2005). There is also evidence that leukoaraiosis has a detrimental effect on cognition and physical functioning prior to acute symptoms, particularly decreasing executive function and motor performance, leading to effects such as gait disturbances (Pantoni, 2007; Schmidt, 2007). These effects may be present long before clinical disease becomes apparent, and several studies have demonstrated an association between hypertension in midlife and cognitive dysfunction in midlife and later life (Elias, 1993; Launer, 2000; Knopman, 2001), likely due to the cumulative effects of subclinical damage due to cerebrovascular disease throughout the lifespan (Knopman, 2001; Swan, 1998).

Leukoaraiosis is present in a much larger proportion of the middle age and elderly population than those who will eventually exhibit clinical endpoints. In routine scans, leukoaraiosis was present in more than half of elderly patients (Schmidt, 2007), and two studies of stroke- and dementia-free adults ages 50-75 reported that less than a third of participants had no leukoaraiosis at all (Markus, 2005; Mosley, 2005). Given the public health importance of preventing stroke and dementia related to leukoaraiosis as well as the mounting evidence that the disease process begins long before the appearance of clinical endpoints, understanding the biological processes underlying leukoaraiosis and its sequelae are a key first step toward identifying and reducing disease progress in individuals at increased risk.

Consistently high heritability estimates for leukoaraiosis (0.45 to 0.68) in multiple cohorts of white and African American individuals indicate that this trait has a large genetic component even after adjustment for blood pressure (Turner, 2004; Atwood, 2004; Carmelli, 1998; Turner, 2009). However, candidate gene studies have revealed inconsistent findings with little unequivocal evidence of association between any specific polymorphism and leukoaraiosis (Paternoster, 2009). One potential reason for lack of consistent findings is the limited knowledge regarding the genetic and molecular mechanisms that lead to the development of elevated leukoaraiosis levels. The agnostic nature of genome-wide association studies (GWAS) may provide a powerful approach for identifying previously unrecognized pathophysiologic mechanisms that affect complex traits such as leukoaraiosis (Hirschhorn, 2005).

To identify genetic markers associated with leukoaraiosis volume, we conducted a GWAS in 759 non-Hispanic white and 553 African-American participants from the Genetics of Microangiopathic Brain Injury (GMBI) ancillary study of the Genetic Epidemiology Network of Arteriopathy (GENOA). Since the GENOA participants were recruited from families with a high prevalence of early-onset essential hypertension, they provide an ideal sample for beginning to examine the genetics of the downstream consequences of hypertension. Although few studies examine leukoaraiosis volume, we were able to identify a second sample of 428 African Americans from the Atherosclerosis Risk in Communities (ARIC) study that also had leukoaraiosis measurements, and we

performed meta-analysis of the GWAS results from the GENOA and ARIC African Americans.

There is mounting evidence that the disease processes of vascular dementia and Alzheimer's disease (AD) are not entirely independent. A substantial portion of AD patients often show increased leukoaraiosis upon closer examination (Kukull, 2002; Breteler, 2000), and there is greater recognition that many of these patients should actually be categorized as having "mixed dementia" whereby cognitive decline and other symptoms of dementia are actually due to a combination of AD and vascular pathologies such as leukoaraiosis. The *APOE* ε4 allele is a known risk factor for both AD and cardiovascular disease (Dickstein, 2010). Although there is conflicting evidence as to whether the *APOE* ε4 allele also confers risk directly to the development of leukoaraiosis (Paternoster, 2009; Szolnoki, 2004), it is possible that the AD and vascular dementia disease processes may be operating synergistically across the lifespan and in particular at the earlier stages of the biological processes that lead to dementia. In order to reduce the genetic heterogeneity of the study sample with respect to susceptibility to AD, we also performed meta-analysis on the GWAS results from the GENOA and ARIC African Americans after excluding participants with either two copies or at least one copy of the ε4 allele.

Methods

Study population

The Genetic Epidemiology Network of Arteriopathy (GENOA) study

The National Heart, Lung and Blood Institute established the Family Blood Pressure Program (FBPP) in 1996 from four existing research networks that were investigating the genetics of hypertension and its sequelae (FBPP Investigators, 2002), including GENOA. GENOA recruited hypertensive sibships from Rochester, Minnesota and Jackson, Mississippi for linkage and association studies to investigate the genetic underpinnings of hypertension and target organ damage related to hypertension (Daniels, 2004).

In the initial phase of the GENOA study (Phase I: 1996-2001), all members of sibships containing ≥ 2 individuals with essential hypertension clinically diagnosed before age 60 were invited to participate, including both hypertensive and normotensive siblings (N=1,583 non-Hispanic whites and 1,853 African Americans). The diagnosis of essential hypertension was established based on blood pressure levels measured at the study visit (>140 mmHg average systolic BP or >90 mmHg average diastolic BP) or a prior diagnosis of hypertension and current treatment with antihypertensive medications. Exclusion criteria were secondary hypertension, alcoholism or drug abuse, pregnancy, insulin-dependent diabetes mellitus, or active malignancy. In the second phase of the GENOA study (Phase II: 2000-2004), 1,239 white and 1,482 African American participants were successfully re-recruited to measure potential target organ damage due to hypertension. Phase I and II GENOA data consist of demographic information, medical history, clinical characteristics, lifestyle factors, and blood samples for

genotyping and biomarker assays. Written informed consent was obtained from all subjects and approval was granted by participating institutional review boards. All reported phenotype and covariate data used for this dissertation was collected during the Phase II exam.

The Genetics of Microangiopathic Brain Injury (GMBI) study (2001-2006) is an ancillary study of GENOA undertaken to investigate susceptibility genes for ischemic brain injury. Phase II GENOA participants that had a sibling willing and eligible to participate in the GMBI study underwent neurocognitive testing to assess several domains of cognitive function including learning, memory, attention, concentration, and language (N=967 whites and 1,010 African Americans). Ischemic brain damage to the subcortical and periventricular white matter (leukoaraiosis) was quantified by MRI in subjects who had no history of stroke or neurological disease and no implanted metal devices (N=916 whites and 830 African Americans). MRI data for 68 participants was not analyzable due to previously undetected brain infarction (N=36), artifacts related to subject or technical error (N=36), anatomic abnormalities (N=8), or failure to complete the MRI (N=7). The median time between the Phase II GENOA visit and brain MRI was 12.3 months in whites and 11.3 months in African Americans. After excluding participants without genome-wide genotype data, the sample for this analysis consisted of 759 whites from 374 sibships and 553 African Americans from 339 sibships.

The African American cohort of GENOA has a close relationship to the ARIC study. The African American sibships for the GENOA study were identified using hypertensive

subjects from the ARIC study as probands, inquiring whether another sibling in the proband's family also had hypertension, and then inviting all siblings from eligible sibships to be part of the GENOA study. Thus, each African American sibship in GENOA contains one individual who is also a participant in the ARIC study. Participants that were dually enrolled in both GENOA and ARIC MRI studies (N=52) were excluded from the GENOA sample prior to meta-analysis with ARIC, leaving a total of 501 GENOA African Americans included in the meta-analysis.

GWAS analysis and meta-analysis with the ARIC cohort was performed for the full sample as well as two subsamples created by excluding participants with 1) two *APOE* ε4 alleles or 2) at least one *APOE* ε4 allele. In the 501 GENOA African Americans, the allele frequency of the ε4 allele is 21.4%. There are 25 individuals (5.0%) that had two copies of the ε4 allele, leaving a total of 476 GENOA participants in the sample after exclusions. There are 189 individuals (37.7%) that had at least one copy of the ε4 allele, leaving a total of 312 GENOA participants in the sample after exclusions.

The Atherosclerosis Risk in Communities (ARIC) study

The ARIC study is a prospective, population-based sample of 15,972 adults aged 45-64 years recruited through probability sampling from Forsythe County, NC, Minneapolis, MN, Washington County, MD, and Jackson, MS (African Americans only) between 1987 and 1989 to study atherosclerosis and its clinical sequelae (Mosley, 2005; ARIC Investigators, 1989). In 1993 and 1994, a total of 1,949 participants aged 55 years and older from Forsythe County and Jackson underwent cerebral MRI. Participants with

reported stroke or infarction, who were not white or African American, or who were taking medications with central nervous system effects were excluded from subsequent analysis. After excluding people without genome-wide genotype data, the sample for this analysis consisted of 428 African Americans.

MRI protocol

GENOA MRI protocol

Leukoaraiosis volume (cm^3) was obtained via MRI in a separate clinical visit. All MRI scans were performed on identically equipped Signa 1.5-T MRI scanners (GE Medical Systems, Waukesha, WI, USA) and images were centrally processed at the Mayo Clinic. Symmetric head positioning with respect to orthogonal axes was verified by a series of short scout scans. Total intracranial volume (head size) was measured from T1-weighted spin echo sagittal images, each set consisting of 32 contiguous 5 mm thick slices with no interslice gap, field of view = 24 cm, matrix = 256 x 192, obtained with the following sequence: scan time = 2.5 min, echo time = 14 ms, repetitions = 2, replication time = 500 ms (Jack, 1989). Total brain and leukoaraiosis volumes were determined from axial fluid-attenuated inversion recovery (FLAIR) images, each set consisting of 48 contiguous 3-mm interleaved slices with no interslice gap, field of view = 22 cm, matrix = 256 x 160, obtained with the following sequence: scan time = 9 min, echo time = 144.8 ms, inversion time = 2,600 ms, repetition time = 26,002 ms, bandwidth = +/- 15.6 kHz, one signal average. A FLAIR image is a T2-weighted image with the signal of the cerebrospinal fluid nulled, such that brain pathology appears as the brightest intracranial tissue. Interactive imaging processing steps were performed by a research associate who

had no knowledge of the subjects' personal or medical histories or biological relationships. A fully automated algorithm was used to segment each slice of the edited multi-slice FLAIR sequence into voxels assigned to one of three categories: brain, cerebrospinal fluid, or leukoaraiosis. The mean absolute error of this method is 1.4% for brain volume and 6.6% for leukoaraiosis volume, and the mean test-retest coefficient of variation is 0.3% for brain volume and 1.4% for leukoaraiosis volume (Jack, 2001). White matter hyperintensities in the corona-radiata and periventricular zone, as well as central gray infarcts (i.e., lacunes) were included in the global leukoaraiosis measurements. Brain scans with cortical infarctions were excluded from the analyses because of the distortion of the leukoaraiosis volume estimates that would be introduced in the automated segmentation algorithm.

ARIC MRI protocol

The MRI scanning protocol and image analysis for the ARIC study is described in detail elsewhere (Mosley, 2005; Bryan, 1994). Briefly, 1.5-T MRI scanners (GE or Picker) were used to generate spin-echo, spin-density/T2-weighted, and T1-weighted images with 5-mm section thickness, 0-mm section gap, and 24-cm field of view. MRIs were read by two trained board-certified radiologists or experienced neuroimaging technicians without knowledge of participant characteristics.

Genotyping

GENOA genotyping

Subjects were genotyped on the Affymetrix® Genome-Wide Human SNP Array 6.0 array using the protocol outlined by Affymetrix (Affymetrix, 2007a) at the Mayo Clinic in Rochester, Minnesota. Briefly, 500ng genomic DNA at 50ng/ul in low EDTA-Tris buffer was digested in two separate reaction mixtures using the appropriate restriction enzyme (StyI and NspI, 250ng of DNA for each mixture). This was followed by ligation of an adaptor sequence containing a universal primer sequence. Samples were then subjected to polymerase chain reaction (PCR) (four PCR reactions per sample for the NspI mixture and three for StyI) with conditions designed to amplify 200-2,000 base pairs. The seven PCR products were then combined with Agencourt Ampure beads, passed over an E & K Scientific filter plate, and eluted with elution buffer. Agarose gel analysis of the PCR products and quantification of the amount of PCR product was performed. PCR product concentration was confirmed to be at least 5ug DNA in 1ul EB buffer. Product was then fragmented with DNase I and an agarose gel analysis of the fragmented DNA was used to confirm this step. Following fragmentation, DNA was labeled with Terminal Deoxynucleotidyl Transferase (TdT), hybridized to the appropriate GeneChip, and incubated overnight. The chip was stained and washed on the Affymetrix 450 Fluidics station and then scanned on the Affymetrix 3000 GeneChip scanner.

Preliminary SNP genotype calls were generated using the Dynamic Model (DM) algorithm (Cutler, 2001). The final SNP genotype calls were generated by Birdseed, an algorithm designed especially for the Affymetrix® Genome-Wide Human SNP Array

6.0, and based on the robust linear model with Mahalanobis distance classifier algorithm (RLMM) (Rabbee, 2006). In order to call genotypes while best accounting for experimental variability and population-specific allele frequencies, Birdseed utilizes information about variation across samples to modify pre-computed genotype calling models from Affymetrix for each SNP probe set. Birdseed has been shown to reduce the bias against heterozygous calls and boost call rates to over 99% while simultaneously increasing concordance rates (Affymetrix, 2007b).

Genome-wide genotyping of all available blood samples from Phase I and Phase II participants resulted in genotyping data for 1,386 whites and 1,263 African Americans, after removing samples with a genotype call rate < 95%. Of the approximately 900,000 Affymetrix 6.0 SNPs genotyped, 668,293 SNPs in whites and 762,766 SNPs in African Americans were available for analysis after removing SNPs that were monomorphic or had a call rate < 95% in the full sample. In order to prevent false positive associations due to a small number of people in a single genotype category, SNPs with a minor allele frequency (MAF) less than 0.01 were removed, leaving a total of 666,271 SNPs in whites and 760,699 SNPs in African Americans for association testing. For each subsample submitted for meta-analysis with the ARIC study, SNPs with a MAF less than 0.01 in the subsample were removed.

The distribution of the MAFs for the 668,293 Affymetrix 6.0 SNPs in GENOA whites and the 762,766 Affymextrix 6.0 SNPs in the GENOA African Americans are shown in Appendix 3.1. In whites, the mean MAF was 0.235 (median = 0.225). SNPs that had a

MAF < 0.01 were excluded from the analysis (n=2022 (0.3%)). There were 152,155 (22.7%) SNPs with 0.01 ≤ MAF < 0.10, 212,955 (31.9%) SNPs with 0.10 ≤ MAF < 0.25, and 301,161 (45.1%) SNPs with 0.25 ≤ MAF ≤ 0.50.

In African Americans, the mean MAF was 0.219 (median = 0.197). SNPs that had a MAF < 0.01 were excluded from the analysis (n=2067 (0.3%)). There were 193,251 (25.3%) SNPs with 0.01 ≤ MAF < 0.10, 267,866 (35.1%) SNPs with 0.10 ≤ MAF < 0.25, and 299,582 (39.3%) SNPs with 0.25 ≤ MAF ≤ 0.50. After exclusion of SNPs with MAF < 0.01, 666,271 SNPs in the white sample and 760,699 SNPs in the African American sample remained for analysis.

ARIC genotyping

ARIC subjects were genotyped on the Affymetrix® Genome-Wide Human SNP Array 6.0 array using the protocol outlined by Affymetrix (Affymetrix, 2007a) at the Broad Institute. Samples or SNPs that had a call rate <95% in the full ARIC sample were removed, as were SNPs with a MAF less than 0.01.

Statistical analysis

Descriptive statistics

Data management and statistical analyses were conducted primarily in R version 2.8.0, an open-source statistical environment for storing data, running analyses, and generating publication-quality graphics (R Core Development Team, 2008). HelixTree, a commercially available software package (http://goldenhelix.com), was used to generate

descriptive statistics of SNP allele and genotype frequencies, to obtain principal components for analysis of population substructure, and to generate linkage disequilibrium (LD) plots of selected areas of the genome. Distributional plots indicated that the measures of leukoaraiosis volume are severely right-skewed, so this variable was transformed by taking the natural log of (leukoaraiosis + 1). T-tests were conducted for the outcome measure and demographic/anthropometric covariates to test whether there were significant differences between participants that had available genotype measures and those that did not in the white and African American samples separately. T-tests were also conducted for the outcome measure and demographic/anthropometric covariates to compare the white and African American study participants that had available genotype measures.

Population substructure

Population substructure is known to be a potential source of confounding in genetic association studies, particularly in study populations that have experienced admixture (the joining of two genetically distinct "parent" populations) in recent history (Freedman, 2004). Admixture in a study population will cause confounding of the association between a SNP and the trait of interest if the three following conditions are met: 1) the admixture proportions vary among study participants, 2) the mean value of the trait varies with admixture proportions, and 3) the allele frequencies of the SNP vary with admixture proportions (Hoggart, 2003). African Americans are an admixed population because they have substantial allelic contributions from European and African ancestors. The African Americans in the GENOA study are from Jackson, MS, which is part of the Mississippi

delta area. Anthropological evidence indicates that this population is admixed from West African and Northern European populations (Jackson, 2006).

To prevent confounding from population admixture, Helix Tree was used to conduct principal components analysis using the genome-wide genotypes of the African-American sample (N=553). An additive model was assumed for the SNPs, which are standardized with a mean of 0 and variance of 1. SNPs that were missing for an individual were assumed to be zero and were not included in the estimate of the mean for that SNP. Although only one of the principal components was significantly associated with leukoaraiosis (PC9, p-value = 0.04), all the first ten principal components of the genome-wide genotypes of the Jackson sample were included as adjustment variables in all models to prevent confounding from population substructure.

Association testing in GENOA

The outcome variable for analysis was the residual value of the natural logarithm of leukoaraiosis plus one, adjusted for age at MRI, gender, total intracranial volume (TIV), and the first ten components of Affymetrix 6.0 genome-wide genotype data (for African-Americans only), as shown below.

The outcome variable for whites was the residual value (ε_j) from the following model, where *j* designates an individual:

$$\ln(\text{leukoaraiosis}+1)_j = \beta_0 + \beta_1(\text{Age at MRI}_j) + \beta_2(\text{Gender}_j) + \beta_3(\text{TIV}_j) + \varepsilon_j$$

The outcome variable for African-Americans was the residual value (ε_j) from the following model:

$$\ln(\text{leukoaraiosis}+1)_j = \beta_0 + \beta_1(\text{Age at MRI}_j) + \beta_2(\text{Gender}_j) + \beta_3(\text{TIV}_j) + \beta_4(\text{PC1}_j) + \beta_5(\text{PC2}_j) \ldots + \beta_{13}(\text{PC10}_j) + \varepsilon_j$$

Age and gender were included as adjustment covariates because both have been historically used as adjustment variables for this trait. Age is a very strong independent predictor for leukoaraiosis, and gender has been shown to have a marginal association with leukoaraiosis in some samples, including GENOA whites. To account for differences in brain size, intracranial volume was also included as an adjustment variable.

Since the GENOA cohort is composed of siblings, linear mixed effects modeling was used to test all associations between a single SNP and the outcome. Genotypes were coded additively as SNP = 0, 1, 2 for people that are homozygous for one allele of a particular SNP, heterozygous, and homozygous for the other allele, respectively (Weir, 1996). Linear mixed effects modeling retains an appropriate type I error rate in the presence of family structure (Raudenbush, 2002). A likelihood ratio test comparing a full model (which includes the SNP being tested for association) to a reduced model (in this case, a null model) was used to determine whether the SNP is significantly associated with the outcome when appropriately accounting for family structure. When testing whether a single SNP is associated with an outcome of interest, the null hypothesis is that the additive SNP effect is zero. The linear mixed effects model used for SNP association testing in the GENOA cohort is shown below. Since the ARIC cohort is composed of

unrelated individuals, linear least squares regression modeling with SNPs coded additively was used for SNP association testing in this cohort.

Full model: $\ln(\text{leukoaraiosis}+1)\ \text{residual}_{ij} = \beta_{oi} + \beta_1(SNP_{ij}) + \varepsilon_{ij}$,

Reduced model: $\ln(\text{leukoaraiosis}+1)\ \text{residual}_{ij} = \beta_{oi} + \varepsilon_{ij}$,

$\mathbf{H_o}$: $\beta_1 = 0$, $\mathbf{H_a}$: $\beta_1 \neq 0$

LR = 2 (-log likelihood full model – (-log likelihood reduced model)) $\sim \chi^2_{(df=1)}$

In this model, β_{oi} is a random intercept that allows the intercept for the i^{th} family to vary from the fixed, population average intercept. β_{oi} is normally distributed with mean β_o and variance σ_b^2. ε_{ij} is the residual variation within the j^{th} individual from the i^{th} family. ε_{ij} is assumed to be independent from β_{oi} and is normally distributed with mean 0 and variance σ^2.

Meta-analysis

In cases in which raw data can not be shared due to lack of informed consent or other ethical or logistical concerns, meta-analysis of GWAS results can be used to increase the statistical power to detect SNP associations by combining results across studies (Zeggini, 2009; Nakaoka, 2009). The MetABEL/GenABEL R package (Aulchenko, 2007) was used to conduct a fixed-effects meta-analysis of the results from the GENOA and ARIC cohorts using inverse-variance weighting, after performing genomic control within each cohort individually (meta-analysis conducted by Myriam Fornage and colleagues at the University of Texas Health Science Center at Houston). As a collaborative group, we

made the decision to perform meta-analysis only on directly measured SNPs, not imputed SNPs, because both of our samples were measured on the Affymetrix 6.0 platform and were therefore directly comparable.

As a first step, the genomic inflation factor was calculated for each cohort. The genomic inflation factor, λ, is defined as the ratio of the median of the observed distribution of test statistics for the genome-wide associations to the expected median under a null hypothesis (Devlin, 1999). Therefore, λ quantifies the extent of inflation of the association test statistics across the genome and therefore the excess false positive rate (de Bakker, 2008). Since this meta-analysis involved GWAS using a quantitative trait, standard errors were corrected within cohort using the following formula:

$$SE_{corr} = SE \times \sqrt{\lambda}$$

For each SNP, meta-analysis βs and standard errors were calculated using the following formulas:

$$\beta_{meta} = \frac{\left(\beta_{GENOA} / SE^2_{corr,GENOA}\right) + \left(\beta_{ARIC} / SE^2_{corr,ARIC}\right)}{\left(1/SE^2_{corr,GENOA}\right) + \left(1/SE^2_{corr,ARIC}\right)}$$

$$SE_{meta} = \sqrt{\frac{1}{\left(1/SE^2_{corr,GENOA}\right) + \left(1/SE^2_{corr,ARIC}\right)}}$$

This method weights the meta-analysis β by the inverse of the squared standard errors from each cohort. In this way, the meta-analysis β is most strongly influenced by the β estimate from the cohort that has the least uncertainty in its estimation, based both on

sample size and the underlying distributions of the trait under analysis and the genotype frequencies. Fixed effects meta-analysis assumes that the effect of each SNP on the trait of interest is the same in each study (i.e. that there is no heterogeneity in the effect the SNP has on the trait in different study samples) (Zeggini, 2009). The major sources of heterogeneity of SNP effects are interactions with other SNPs and environmental covariates. Since the GENOA and ARIC sample are both composed of African Americans from Jackson, MS, and have comparable distributions of most demographic variables, the underlying distributions of allele frequencies and environmental factors should be similar. An exception to this is that the GENOA sample is composed of primarily hypertensive individuals while the ARIC sample is population-based. Despite this caveat, we feel that the assumption of fixed effects is reasonable.

Results

GENOA descriptive statistics

The sibship structure of the 759 white and 553 African American GENOA participants are presented in Tables 3.1 and 3.2. The majority of the study samples (>60%) are composed of either singletons or sib-pairs, and a very small percentage of each sample is composed of sibships with more than five people (5.4% in whites, 1.1% in African Americans). Descriptive statistics for GENOA are presented in Table 3.3. GENOA whites are 58.8% female, 73.4% hypertensive, have a mean age of 60.5 years at the time of MRI, and have a mean leukoaraiosis volume of 7.89cm^3. GENOA African Americans are 68.9% female, 77.4% hypertensive, have a mean age of 63.5 years at the time of MRI, and have a mean leukoaraiosis volume of 10.55cm^3. There are very few significant

differences in either sample between participants that have Affymetrix 6.0 genotype data available and those who do not, except that whites with genotype data have a significantly higher volume of leukoaraiosis and are slightly older than whites without genotype data (see Appendix 3.2). However, there are many significant differences between white and African American samples, most notably that the African American sample has a larger percentage of females as well as higher mean leukoaraiosis volume, age at MRI, and blood pressures.

Outcome variable and adjustment covariates

The outcome variable, leukoaraiosis, was strongly right skewed in both populations (see Appendix 3.3), but had a relatively normal distribution after taking the natural log of leukoaraiosis+1 and adjusting for age at MRI, gender, TIV, and the first ten principal components of the genome-wide genotypes (in African Americans only, described below). Table 3.4 shows the results of multivariable linear mixed models that include the adjustment variables. Age at MRI and TIV had significant associations with ln(leukoaraiosis+1), but gender was only moderately significant or not significant in the multivariable model due to the strong correlation between gender and TIV (correlation = -0.626 in whites, -0.577 in African Americans). In order to adjust for population structure in the African American sample, the first ten principal components of the Affymetrix 6.0 genotypes were calculated and are discussed in Appendix 3.4.

Association testing in GENOA

The genomic inflation factor was 1.013 for the GWAS in GENOA whites and 0.999 for GENOA African Americans, indicating that neither group showed systematic inflation of the test statistic across the genome. Quantile-quantile plots of the observed and expected p-values for the association tests between each SNP and the residual value of adjusted ln(leukoaraiosis+1) are presented in Figures 3.1 and 3.2, indicating that the p-values of the results are in accordance with a distribution of p-values that would be expected by chance alone, though there appears to be a slight excess of p-values more significant than expected by chance in the range of $1x10^{-4}$ and $1x10^{-5}$ for African Americans.

None of the SNPs in either group reached the genome-wide significance level of $5x10^{-8}$, the standard significance threshold set for GWAS by the statistical genetics community based on Bonferroni correction of a p-value<0.05 for one million SNPs (Petretto, 2007); however, several associations were suggestive of significance. Manhattan plots illustrating the distribution of SNP p-values across the genome are shown in Figures 3.3 and 3.4. Tables 3.5 and 3.6 characterize the top 20 SNPs associated with the outcome in whites and African Americans.

In whites, there were 80 SNPs with association p-values $<1x10^{-4}$ and five with p-values $<1x10^{-5}$. The strongest associations in whites were rs11686818 on chromosome 2 in an intergenic region 1052bp from *THSD7B* (p-value $2.98x10^{-6}$), rs3787569 on chromosome 20 in an intron of *ANGPT4* (p-value $6.57x10^{-6}$), and rs6537339 on chromosome 4 in an intergenic region 142kb from *OTUD4* (p-value $7.33x10^{-6}$). The top 20 strongest

association results in whites also included a SNP that results in a missense mutation in *NOC3L* and intronic SNPs in *EEFSEC* and *CTNNA3*.

In African Americans, there were 83 SNPs with association p-values $<1x10^{-4}$ and nine with p-values $<1x10^{-5}$. The strongest associations were rs4979887 on chromosome 10 in an intron of *KCNMA1* (p-value $3.0x10^{-6}$), rs12554999 on chromosome 9 in an intergenic region 35kb from *CHCHD9* (p-value $3.4x10^{-6}$), and rs4575062 on chromosome 1 in an intergenic region 170kb from *PBX1* (p-value $3.45x10^{-6}$). In the top 20 results in African Americans, SNPs showing association with leukoaraiosis were also found in the intronic regions of *PRR11*, *ADAMTS19*, *N6AMT1* as well as two open reading frames that have not been definitively identified as functional genes. One associated SNP was identified in the 3' region of *TRIM45*, 184bp from the end of the gene. No genes in the top 20 results were found in both whites and African Americans.

Meta-analysis

Descriptive statistics of the full sample included in the meta-analysis (501 GENOA African Americans and 428 ARIC African Americans, total N=929) are presented in Table 3.7. The GENOA participants included in this sample are 70.3% female and 77.0% hypertensive, have a mean age at MRI of 62.7 years, and have a mean ln(leukoaraiosis+1) value of 2.17. The ARIC participants are 62.9% female and 57.5% hypertensive, have a mean age at MRI of 71.8 years, and have a mean ln(leukoaraiosis+1) value of 2.41. The ARIC participants are almost a decade older than the GENOA participants, potentially accounting for their higher mean value of

leukoaraiosis. The other large difference between the cohorts is the larger percentage of hypertensive individuals in the GENOA sample due to the difference in sample recruitment strategies.

Tables 3.8-3.10 show the results of the meta-analyses conducted using the GENOA and ARIC African Americans, and Figures 3.5-3.7 show the Manhattan plots of the meta-analysis results. There were five results with a p-value $<1x10^{-5}$ from the meta-analysis in the full sample. The strongest association was on rs1945938 on chromosome 11, in an intergenic region near *SNX19* (p-value $5.2x10^{-7}$). Meta-analysis conducted after excluding participants with two copies of the *APOE* ε4 allele consisted of 476 GENOA participants and 407 ARIC participants (total N=883). The strongest associations from this sample included the same SNP from *SNX19* (p-value $8.2x10^{-6}$) as well as SNPs in introns of *MOSC2* (rs10863562, p-value $4.9x10^{-6}$) and *PRR11* (rs2687065, p-value $6.3x10^{-6}$). Meta-analysis conducted after excluding participants with at least one copy of the *APOE* ε4 allele consisted of 312 GENOA participants and 290 ARIC participants (total N=602). The strongest results from this analysis were two SNPs in an intergenic region between 44kb and 70kb upstream from the *ITGB1* gene (rs9299702, p-value $3.7x10^{-8}$ and rs7898823, p-value $3.2x10^{-7}$). Figure 3.8 shows the recombination rates in the Yoruban HapMap sample in this chromosomal region, along with the –log p-values from the meta-analysis. Top findings also included SNPs from intronic regions of *ARHGAP20*, *CTNNA3*, *MAP3K12*, *PARK2*, *ASTN2*, and *NRXN1*.

Discussion

Leukoaraiosis has a detrimental effect on cognition and physical functioning prior to the development of acute symptoms and can lead to serious clinical endpoints such as stroke and dementia. Since the development of leukoaraiosis begins long before the appearance of clinical endpoints, a greater understanding of the biological processes underlying leukoaraiosis may assist in identifying individuals at increased risk for disease and reducing disease progression. In this study, we performed a GWAS to identify SNPs associated with leukoarariosis in stroke- and dementia-free whites and African Americans. While this study was limited by sample size, the biological relevance of some of the top findings in this study suggests that these results may contain promising leads for the next phase of analysis in examining the genetic architecture of leukoaraiosis.

Biological relevance of GWAS results

While none of the top 20 findings overlapped between GENOA whites and African Americans at the gene level, one striking finding was that a member of the *ADAMTS* family (a disintegrin-like and metalloproteinase with trombospondin motif) was identified as a top result in the GWAS for each group. Members of the *ADAMTS* family are secreted extracellular matrix (ECM) metalloproteinases that exhibit adhesion and protease activity (MIM 605174). They interact with the ECM through thrombospondin domains, which show homology to the known potent anti-angiogenic motifs of thrombospondin-1 (*THBS1*) (MIM 188060). *THBS1* inhibits angiogenesis of blood vessels and promotes central nervous synaptogenesis (Christopherson, 2005; Volpert, 2002). An anti-angiogenic effect has been demonstrated for several members of the

ADAMTS family, including the *ADAMTS8* gene identified in the GENOA whites. In fact, functional studies have shown that *ADAMTS8* has a stronger anti-angiogenic effect than *THBS1*, and that its effects on cell proliferation are specific to endothelial cells and do not have an effect on smooth muscle or fibroblast cells (Vazquez, 1999). Very little information is known about the specific function of the *ADAMTS19* gene (MIM 607513) identified in the GENOA African Americans.

The structure and function of the *ADAMTS* genes are similar to the genes in the *ADAM* family, which also affect vascular function and have been implicated in the development of dementia (MIM 602192). In vitro studies in human cell lines have found that at least one member of the *ADAM* family (*ADAM10*) mediates the effect of cholesterol on the amyloid precursor protein, the main constituent of the amyloid plaques present in AD (Kojro, 2001). In mice that carried an expressed version of the human amyloid beta a4 precursor protein (*APP*), overexpression of *ADAM10* reduced the formation of amyloid beta peptides and reduced plaque formation (Postina, 2004).

Several other top findings are consistent with the involvement of genes containing thrombospondin motifs, including the top finding in whites, *THSD7B* (thrombospondin-1, domain containing 7B), whose precise function is currently unknown. However, this gene showed suggestive evidence of association with neurocognitive function in a genome-wide study of 750 subjects conducted by Need and colleagues (2009), as did a member of the *ADAM* family. There is also a relationship between *THBS1* itself and the top finding in the meta-analysis of GENOA and ARIC participants that excluded those with at least

one copy of the *APOE* ε4 allele, integrin beta-1 (*ITGB1*). *THBS1* has been identified as a ligand of *ITGB1* (MIM 135630), and the interaction of these two proteins has been shown to affect cell proliferation and the development of new vasculature in animal models (Staniszewska, 2007). *ITGB1* is also a subunit of receptors for other ligands that are known to affect vasculature remodeling, including *VCAM1 (Garmy-Susini, 2005)*, and is a mediator of inflammatory response (Suzuki, 2007; Conrad, 2007). In animal models, *ITGB1* has also been shown to be involved in neuronal adhesion and migration (Dulabon, 2000).

The second strongest finding in whites is a SNP in an intron of angiopoietin 4 (*ANGPT4*) (MIM 603705). Angiopoietin growth factors are ligands that regulate the proliferation and maturation of vascular endothelial cells (Thomas, 2009). *ANGPT4* activates *TIE2*, an endothelial cell-specific tyrosine kinase receptor, which promotes vessel growth. Using family-based association testing and fine-mapping linkage methods in 30 families with mixed AD/vascular disease, *ANGPT4* was identified as the strongest finding associated with disease status (Sillen, 2010). The relationship between *ANGPT4* and AD has also been demonstrated by transcriptomic analysis of the brain, with a 65% difference in expression of *ANGPT4* in the brains of AD cases and controls (Chapuis, 2009).

Another interesting finding is that one of the top results in whites was in an intron of the catenin (cadherin-associated protein), alpha 3 (*CTNNA3*) gene, which plays a role in cadherin-mediated cell adhesion (MIM 607667). Cadherins have been implicated in playing a key role in the formation and maintenance of neurons, stabilization of cell-cell

contacts at synapses, assembly of synaptic molecules, and synaptic placticity (Suzuki, 2008). Interestingly, no SNPs from the *CTNNA3* gene were identified as top results in GENOA African Americans, however this gene did appear as a top result in the meta-analysis of African Americans excluding individuals with either one or two copies of the *APOE* ε4 allele. *CTNNA3* binds to catenin, beta 1 (*CTNNB1*) (MIM 116806), which interacts with *PSEN1*, a well-known susceptibility gene for early-onset familial AD (Ertekin-Taner, 2003). The association between SNPs in *CTNNA3* and late onset AD in females regardless of *APOE* ε4 allele status was demonstrated in a large case-control study (Miyashita, 2007). In addition, *CTNNB1*, the binding partner of *CTNNA3*, was identified as a top hit (p-value<10^{-7}) in the Framingham study of white matter hyperintensity conducted by Seshadri and colleagues (2007) as well as the genome-wide study with cognitive function phenotypes conducted by Need et al. (2009). A SNP in another member of cadherin family, protocadherin (*PCDH7*), was also a top finding in this study for GENOA whites.

One SNP in the glutamate receptor, ionotropic n-methyl-d-aspartate 3a (*GRIN3A*) gene was a top finding in GENOA African Americans. In vitro studies have shown that *GRIN3A* is a subunit of an NMDA glutamate receptor that appears to play a key role in mediating myelin-damaging calcium accumulation in response to ischemia (Micu, 2006). Several other top findings also relate to calcium metabolism in the brain. Neurexin 1 (*NRXN1*), for example, is required for proper functioning of synaptic calcium channels (MIM 600565). This gene was also the top finding in the Need genome-wide study on cognitive phenotypes (Need, 2009). *KCNMA1* (MIM 600150), a top finding in GENOA

African Americans, is a calcium-activated potassium channel that is responsive to oxygen deprivation (Williams, 2004). Another finding in the meta-analysis of GENOA and ARIC African Americans, endonuclease VIII-like 3, (*NEIL3*) (MIM 608934), plays a role in repairing DNA damage caused by reactive oxygen species (Bandaru, 2002).

Power and statistical considerations in GWAS

GWAS have been successful in identifying and validating novel common genetic variants associated with a variety of human diseases and quantitative traits including prostate and breast cancer, type 2 diabetes, coronary heart disease, asthma, lipids, and height (McCarthy, 2008; Manolio, 2010). However, most of the variants identified for these traits have very modest effect sizes, explaining less than 1% of the variance of quantitative traits (de Bakker, 2008). Even when multiple validated loci have been identified for a particular trait, the predictive value of these SNPs is often marginal at best (Lango, 2008). Though the success of GWAS to date has been less than optimal in identifying SNPs with strong predictive value, much has been learned about the genetic basis of these diseases due to the identification of previously unknown biological pathways, leading to new avenues for prevention, diagnosis, and treatment (Wellcome Trust Case Control Consortium, 2007).

In the past several years, much has been learned about the best practices and standards for conducting GWAS. Due to the relatively small effect sizes that most SNP variants confer on the outcomes of interest, quite large sample sizes are needed to obtain adequate power to detect these variants and make reliable inferences (Roberts, 2010). Since the vast

majority of studies are underpowered to detect these variants, collaboration across studies typically in the form of meta-analysis is routinely used to increase the effective sample size and power (de Bakker, 2008; McCarthy, 2008) While this is a relatively straightforward endeavor for traits commonly measured, it becomes difficult with traits that are specialized and complicated to assess, such as leukoaraiosis. Several factors prevent the rapid collaborative relationships necessary to obtain large sample sizes for meta-analysis of leukoaraiosis. There have been very few studies that have assessed this phenotype and also have genome-wide genotype data, particularly in non-European samples. Studies that have measured leukoaraiosis consist almost entirely of small sample sizes due to the expense and time required to measure this phenotype. In addition, different MRI technologies and scoring algorithms for leukoaraiosis quantity often hamper the comparison or compilation of results across studies. For example, the distribution of leukoaraiosis volume in the Framingham Heart Study was so different from that in GENOA due to differences in measurement technique, it was not suitable for use as a replication sample or in a meta-analysis for this study. Study-specific decisions regarding what constitutes leukoaraiosis (for example, the inclusion or exclusion of silent infarcts or areas of diffuse white matter hyperintensity), whether leukoaraiosis is quantified as a continuous volumetric measurement or as a categorically graded variable with various levels (typically between five and nine), and whether or not specific brain regions are included in the measurement makes phenotype alignment across studies difficult. However, collaborative relationships such as that established between GENOA and ARIC make meta-analysis possible even in non-European samples, though the total sample size remains less than optimal.

Though sample size is a limitation in this analysis, standards for conducting GWAS have been clearly specified in the literature and have been applied to this analysis in an effort to obtain as much power as possible to detect associations. Adjusting the trait for age, gender, and other covariates known to be associated with leukoaraiosis is standard protocol, as is adjustment with genotype principal components in populations with a significant amount of admixture. Quality control standards for genotype data have also been established and are imperative to protecting the integrity of the results from bias and error (de Bakker, 2008; McCarthy, 2008; Wellcome Trust Case Control Consortium, 2007). GENOA collaborators at the Mayo Clinic that were responsible for genotyping performed initial quality control measures such as examining the data for plate effects, discarding samples that exhibited sex mismatch, examining the data for duplicate samples, and discarding SNPs that showed excess heterozygosity or had a substantial substantial degree of missingness. For this analysis, samples and SNPs that had a call rate less than 95% have been excluded, as well as SNPs with a MAF less than 1%.

Epidemiological considerations in GWAS

It is widely acknowledged that allelic effects often differ according to their genetic and environmental context (Flint, 2009; Cheverud, 2000; Thornton-Wells, 2004). In animal studies, it is relatively straightforward to study single SNP effects by comparing phenotypes between strains that have an identical genetic background except for genetic loci of interest. In humans, the quantification of a single SNP effect is much more difficult, as both the genetic and environmental backgrounds of individuals vary greatly

(Flint, 2009). Efforts to reduce the genetic heterogeneity in GWAS studies include studying isolated population groups that exhibit a much higher degree of genetic homogeneity than non-isolated populations due to lack of outbreeding for many generations, selecting populations that are known to have similar ancestral backgrounds, and analyzing populations with known differences in genetic background separately. The GENOA and ARIC African Americans were drawn from the Mississippi delta area, and the majority of African Americans in this geographic area share ancestors from a limited area of West Africa and Northern Europe (Jackson, 2006). The whites from Rochester, MN, have ancestral background that is largely from Northern and Western Europe. Since the African Americans and whites in this study have known differences in ancestral populations and consequently genetic backgrounds, we chose to analyze them separately.

A further implication of the differing ancestral background of the African Americans and whites is the difference in the genomic structure of these two population groups. European populations are considerably less diverse genetically than African populations (Rosenberg, 2010) and have a much higher degree of LD and thus larger haplotype blocks in their genome structure (Gabriel, 2002). LD, the correlation structure between SNPs, is a key factor that underlies the ability of GWAS to represent relationships between underlying variation in putatively causal polymorphisms in genomic regions and traits of interest (Kruglyak, 1999). Due to the differences in haplotype block and correlation structure in the genome of African Americans and Europeans, meta-analysis of GWAS results at a SNP level is a less than optimal strategy for identifying genomic regions of interest, because different SNPs are likely to be in strong LD with the different

true causal variants in each population group. To illustrate this, LD plots generated from 60 singletons from the HapMap European (CEU) and African Yoruban (YRI) panels are shown in Appendix 3.5 for selected top findings in GENOA African Americans and whites. As a result, we chose not to perform meta-analysis for the white and African American GENOA samples.

Statistical limitations

Perhaps the most serious limitation of this GWAS is that it is underpowered to detect associated variants with small effect sizes, particularly for SNPs with low MAF. Power is limited due to the small sample size, and it is also slightly decreased compared to a sample of unrelated individuals of the same size because we have sibling data. In a sample of unrelated individuals of the same size as the GENOA whites, we estimate that we would have 80% power to detect effect sizes of 0.75 standard deviations (sd) for SNPs with MAF=0.01, 0.25sd for MAF=0.1, 0.2sd for MAF=0.25, and 0.15sd for MAF close to 0.5. For the same range of MAF in a sample of unrelated individuals the same size as GENOA African Americans, we would have 80% power to detect SNPs with effect sizes of 0.85sd, 0.28sd, 0.22sd, and 0.17sd, respectively.

While we may have lost a small amount of power due to the family structure of our data, Visscher and colleagues (2008) have shown that for GWAS of quantitative traits, there is only a slight reduction in power due to the inclusion of family data. The exact reduction in power depends on several factors including the correlation of the trait among relatives, the size of the families included, and the effect sizes of the SNPs. Visscher et al. estimate

that the range of power lost for sib-pairs within a reasonable range of parameters is between 1% and 7%, but may increase to as much as 20% for large sibships. Our data consisted primarily of sibships of three individuals or less, including a large number of singletons, so most likely does not suffer from more loss of power than would be expected for sib-pairs. Though inclusion of sibships likely did not result in a serious reduction of power, newly developed methods could have been used to incorporate family information more efficiently. For example, a family-based test of association for quantitative traits described by Chen and Abecasis that efficiently uses ancestry information to estimate variance components in a maximum likelihood framework may have been a better choice for our statistical test of association (Chen, 2007).

Limitations in interpreting and generalizing GWAS results

Leukoaraiosis is a downstream consequence of hypertension. Ideally, a GWAS study conducted on target damage due to hypertension would control for hypertension in the analysis in order to avoid detecting SNPs that simply increase risk for the development of hypertension itself. Although blood pressure at the time of the GENOA examination could have been used as an adjustment variable, this measurement is not an accurate proxy for hypertension because a large portion of participants with hypertension were taking anti-hypertensive medication that reduces blood pressure to varying degrees. There are several major drug classes for treating hypertension including diuretics, beta-blockers, and calcium channel blockers. Study participants with history of hypertension may be on a single drug treatment, combination therapy, or no treatment at all. Furthermore, response to hypertensive medications exhibits significant inter-individual

variation (Garcia, 2003), so including a dichotomous variable indicating whether or not a participant is treated would not adequately control for confounding introduced by treatment. In addition, a measurement of current hypertension would not have captured the effect of hypertension on the brain over the life span. Since participants vary widely in the amount of time they had hypertension before seeking treatment as well as the severity of hypertension prior to treatment, the physiological effects of hypertension cannot be captured in a simple blood pressure measurement or even a detailed assessment of patient medical history, which would be imprecise at best. Consequently, the findings of this study will likely contain SNPs that have an effect on hypertension susceptibility as well as susceptibility to the phenotype of interest, leukoaraiosis. SNPs identified in this study that are in genes that play a role in biological pathways related to the development of hypertension such as general vascular function and angiogenesis may only affect leukoaraiosis through their effects on hypertension.

In addition, some of the SNPs identified in this GWAS may affect atherosclerosis in general, not only specifically arteriolosclerosis in the brain. Pathways that are known to affect the atherosclerotic process include those related to endothelial function, response to reactive oxygen species, cholesterol metabolism, and inflammation. Though it has been shown that most of the common atherosclerotic risk factors have weak to no association with leukoaraiosis, the development of atherosclerosis in the large arteries of the heart and brain does reduce the elasticity of the vasculature, which requires the heart to increase blood pressure by pumping more forcefully. Therefore, it is possible that some SNPs that affect atherosclerosis as a whole may be included in these findings either

because of their effect on blood pressure or because of their involvement in aspects of endothelial dysfunction that play a role in atherosclerosis of both the large and small blood vessels. Adjusting the outcome variable for a measure of overall atherosclerosis, such as the quantity of coronary calcium detected by electrocardiogram, may help to limit findings to SNP effects that are specific to small vessel atherosclerosis. Genes identified in this GWAS that have the strongest evidence for affecting leukoaraiosis specifically based on known biological relationships include those that are expressed primarily in the brain, are involved in pathways such as reaction to hypoxia and ischemia, or have been identified as risk factors for measures affected by leukoaraiosis quantity such as cognitive decline or general cognitive function.

Since hypertension is known to have a variety of etiological causes, we chose to enroll families predisposed to developing early-onset essential hypertension in order to reduce the heterogeneity in the etiological causes of hypertension in the GENOA sample. While this recruitment strategy allowed us to maximize our ability to detect genetic variants impacting hypertension and its downstream consequences, it limits the generalization of findings from GENOA to people in families with increased risk of hypertension. Findings from the meta-analysis with ARIC, a population-based sample, may be more generalizable to the broader hypertensive and normotensive population.

Conclusions and future directions

One strength of this study is that this GWAS was conducted on a trait that is an intermediate phenotype between hypertension and the clinical endpoints of ischemic

stroke and vascular dementia. The molecular pathways that begin with genotypic variations which lead to functional changes in proteins and end in clinical endpoints are complex and, in many cases, not well characterized. Investigation of asymptomatic subclinical phenotypes such as leukoaraiosis that precede clinical detectable outcomes presents an opportunity to begin to unravel the etiological complexity of the molecular pathways that affect these outcomes. Most of the GWAS studies in neuroepidemiology of late-onset diseases have been conducted on clinical endpoints that have multiple etiologies (such as ischemic stroke), limiting the amount of information that can be gleaned regarding the underlying physiological pathways. Studying the genetic determinants of subclinical traits such as leukoaraiosis that are predictive of clinical outcomes independent of other risk factors (such as smoking and diabetes) is an important avenue of research for identifying etiological factors that may be missed in other types of studies and providing novel insights into the development of clinical endpoints.

A further strength of this GWAS is that it was conducted in two distinct ethnic groups. The vast majority of GWAS to date have been conducted only in European samples, due in part to the complication of confounding by population substructure and admixture in other ethnic groups. This practice limits the applicability of study findings to European populations only and doesn't allow for comparison of findings in genetically distinct groups. Since allele frequencies and environmental factors have been shown to vary considerably across ethnic groups, future studies of the genetic contribution to complex

traits must include an analysis of how these factors contribute differently across ethnic groups (Altshuler, 2008).

A further strength of this study is that we performed meta-analysis on the GWAS results from the GENOA and ARIC African Americans after excluding participants with either two copies or at least one copy of the *APOE* ε4 allele in order to reduce the genetic heterogeneity of the study sample with respect to susceptibility to AD. Though we did not formally statistically evaluate the effects of genetic context on the GWAS results, we did observe that the most significant p-values in the meta-analysis were from the analysis that excluded participants with at least one ε4 allele. Differences in both the genes identified and the greater statistical significance in the ε4-excluded analysis suggest that more formalized evaluation of the effects of genetic context may be warranted.

Future directions include collaborating with other groups that have measured leukoaraiosis in whites in order to assess replication of these findings across studies, looking for replication of SNPs and gene regions in GENOA that have been associated with leukoaraiosis-related phenotypes such as silent infarcts and ischemic stroke in published GWAS, and using pathway analysis techniques to integrate SNP data with transcriptomic and epigenetic data to explore the underlying biological processes that affect leukoaraiosis through pathway analysis.

Table 3.1. Sibship structure in GENOA whites

Number of Siblings in Sibship	Number of Sibships	Total Number of Individuals	Percentage of Total Number of Individuals
8	2	16	2.1%
7	1	7	0.9%
6	3	18	2.4%
5	7	35	4.6%
4	15	60	7.9%
3	47	141	18.6%
2	183	366	48.2%
1	116	116	15.3%
Totals	**374 Sibships**	**759 individuals**	**100.0%**

Table 3.2. Sibship structure in GENOA African Americans

Number of Siblings in Sibship	Number of Sibships	Total Number of Individuals	Percentage of Total Number of Individuals
6	1	6	1.1%
5	5	25	4.5%
4	5	20	3.6%
3	41	123	22.2%
2	92	184	33.3%
1	195	195	35.3%
Totals	**339 Sibships**	**553 individuals**	**100.0%**

Table 3.3. Comparison of the characteristics of GENOA whites and African Americans

	GENOA whites		GENOA African Americans		T-test p-value
	N	Mean (SD)	N	Mean (SD)	
Leukoaraiosis Volume, cm^3	759	7.89 (6.66)	553	10.55 (11.77)	2.14E-06
Total Intracranial Volume, cm^3	759	1466 (147)	553	1374 (135)	<2.2E-16
Total Brain Volume, cm^3	759	1158 (123)	553	1067 (114)	<2.2E-16
Brain Atrophy (TIV – Brain), cm^3	759	307.4 (74.6)	553	307.5 (75.2)	0.9910
Ventricular Volume, cm^3	759	25.5 (17.0)	553	22.3 (13.6)	0.0002
Time from Exam to MRI, years	759	1.20 (0.77)	553	1.16 (0.80)	0.4177
Age at MRI, years	759	60.5 (9.89)	553	63.5 (8.90)	2.46E-08
Age at Phase II Exam, years	759	59.3 (10.1)	553	62.3 (8.74)	1.68E-08
Body Mass Index, kg/m^2	759	30.4 (5.8)	552	31.2 (5.8)	0.0152
Height, cm	759	168 (9.2)	552	169 (9.0)	0.0390
Weight, kg	552	85.9 (18.2)	552	88.9 (16.6)	0.0023
Waist-to-hip ratio	759	0.91 (0.11)	552	0.89 (0.06)	0.0034
Systolic Blood Pressure, mm Hg	758	131.3 (16.5)	553	136.9 (20.3)	1.25E-07
Diastolic Blood Pressure, mm Hg	758	74.1 (9.1)	553	79.7 (10.5)	<2.2E-16
Pulse pressure, mm Hg	758	57.3 (15.3)	553	57.2 (16.4)	0.9545
Total cholesterol, mg/dL	759	197.8 (33.7)	547	201.9 (42.5)	0.0583
HDL cholesterol, mg/dL	759	52.1 (14.7)	547	57.7 (19.2)	1.20E-08
LDL cholesterol, mg/dL	759	120.6 (31.8)	547	123.1 (39.9)	<2.2E-16
Physical activity	759	-9.93 (7.52)	553	-13.3 (4.74)	<2.2E-16
Female, n (%)	759	446 (58.8%)	553	381 (68.9%)	0.0002
Ever smoker, n (%)	759	366 (48.2%)	553	334 (60.4%)	0.0023
Coronary heart disease, n (%)	759	57 (7.5%)	553	24 (4.3%)	0.0251
Hypertension, n (%)	759	557 (73.4%)	553	428 (77.4%)	0.1111

T-tests compare GENOA whites and African Americans.
All variables except for brain measures were measured at the Phase II GENOA examination.

Table 3.4. Multivariable linear mixed model regression with adjustment covariates as fixed effects, family as the random effect, and ln(leukoaraiosis+1) as the outcome

	Whites			African Americans		
Covariate	**N**	**β**	**p-value**	**N**	**β**	**p-value**
(Intercept)	759	-0.6674	0.0019	553	-1.0601	0.0011
Age at MRI	759	0.0289	1.01E-51	553	0.0285	2.85E-22
Gender	759	-0.0700	0.0653	553	0.0962	0.1026
TIV	759	0.0007	4.02E-07	553	0.0010	8.98E-07

Figure 3.1. Quantile-quantile plot of observed and expected p-values from single SNP associations with residual values of ln(leukoaraiosis+1) in GENOA whites

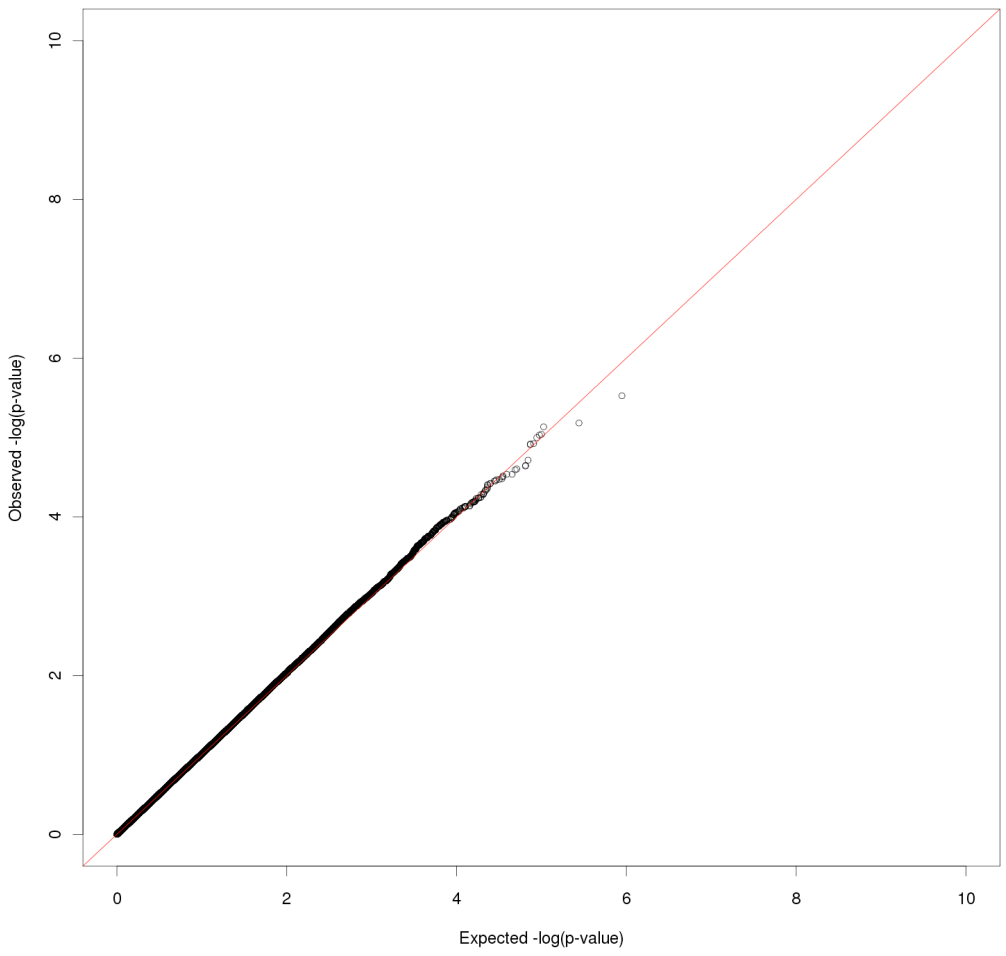


Figure 3.2. Quantile-quantile plot of observed and expected p-values from single SNP associations with residual values of ln(leukoaraiosis+1) in GENOA African Americans

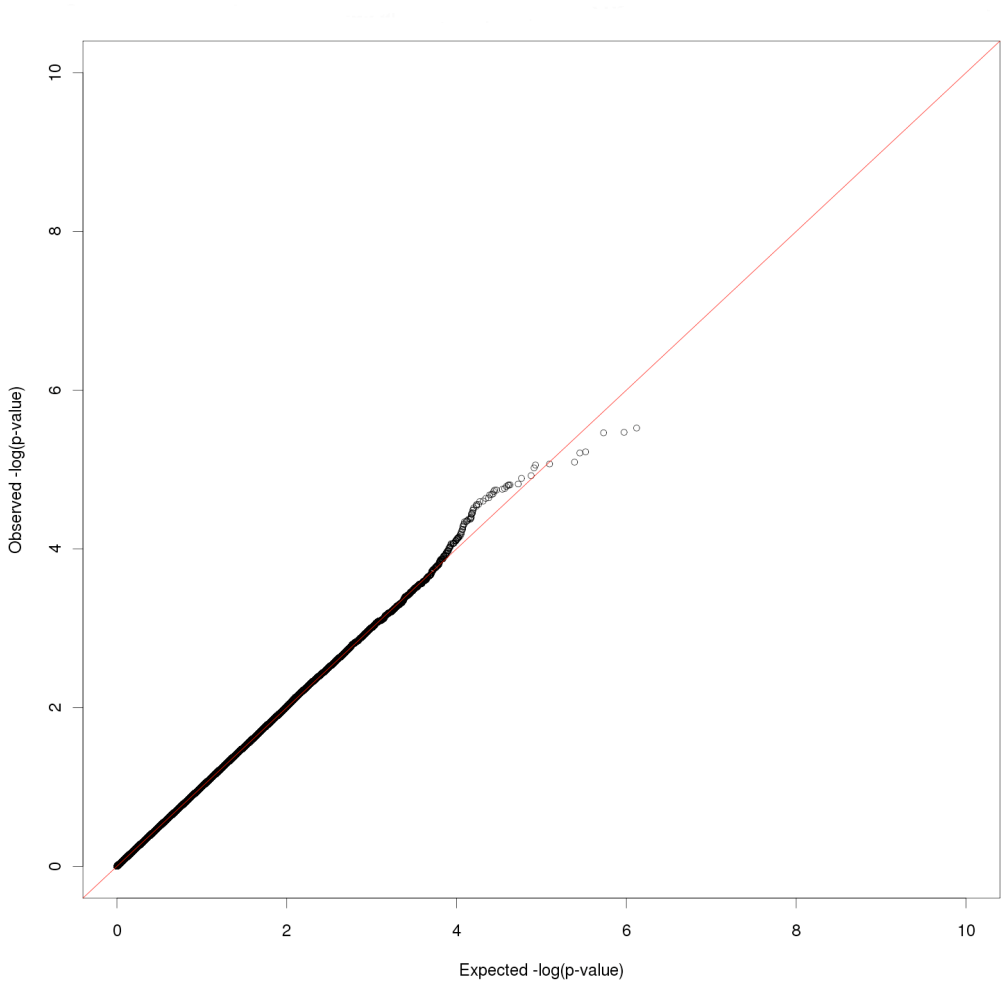


Figure 3.3. Distribution of –log p-values from single SNP associations with residual values of ln(leukoaraiosis+1) in GENOA whites

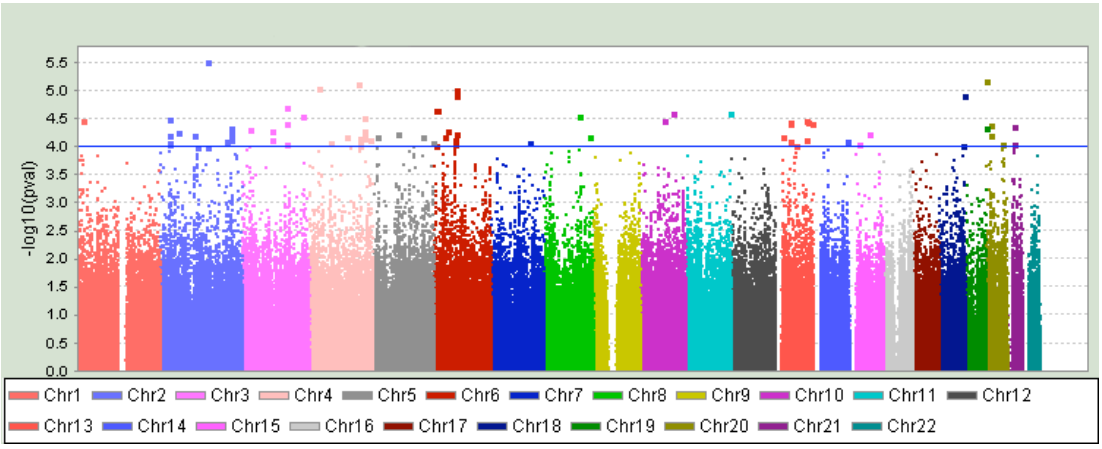


Figure 3.4. Distribution of –log p-values from single SNP associations with residual values of ln(leukoaraiosis+1) in GENOA African Americans

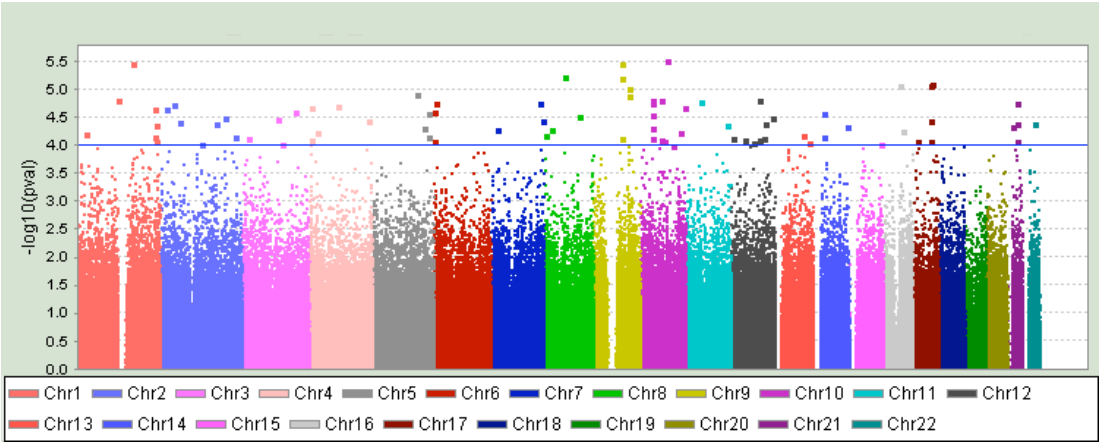


Table 3.5. Top 20 SNP association results in GENOA whites

SNP	Chr	Position	Beta	MAF	Nearest Gene (bp)	Type	LR p-value
rs11686818	2	138152809	0.172	0.094	THSD7B (1052)	intergenic	2.98E-06
rs3787569	20	838278	0.124	0.204	ANGPT4 (0)	intronic	6.57E-06
rs6537339	4	146462918	0.096	0.491	OTUD4 (142636)	intergenic	7.33E-06
rs4312821	4	28917503	-0.128	0.197	PCDH7 (1413631)	intergenic	9.11E-06
rs1459651	6	63941094	0.103	0.327	LGSN (102721)	intergenic	9.33E-06
rs9343759	6	63923583	-0.104	0.322	LGSN (120232)	intergenic	1.00E-05
rs17266916	18	72194302	-0.162	0.101	ZNF516 (6304)	intergenic	1.19E-05
rs1459652	6	63941114	0.100	0.332	LGSN (102701)	intergenic	1.21E-05
rs12195233	6	63929207	-0.104	0.314	LGSN (114608)	intergenic	1.23E-05
rs6801556	3	129576124	-0.295	0.024	EEFSEC (0)	intronic	1.93E-05
rs470778	6	10568858	0.150	0.114	C6orf218 (26098)	unknown	2.25E-05
rs9503670	6	3611366	-0.208	0.051	C6orf145 (56468)	unknown	2.29E-05
rs12572897	10	96104825	-0.131	0.141	NOC3L (0)	missense	2.49E-05
rs11222084	11	129778440	0.095	0.379	ADAMTS8 (1587)	intergenic	2.55E-05
rs16898906	8	102098172	0.198	0.056	YWHAZ (63373)	intergenic	2.91E-05
rs4857784	3	177801480	0.102	0.284	TBL1XR1 (419755)	intergenic	2.91E-05
rs720329	4	164191184	0.195	0.054	NAF1 (76125)	intergenic	3.05E-05
rs10190878	2	25796878	-0.098	0.325	ASXL2 (18878)	intergenic	3.15E-05
rs1335951	13	89668716	0.108	0.240	GPC5 (1180219)	intergenic	3.31E-05
rs10762151	10	68643621	-0.133	0.135	CTNNA3 (0)	intronic	3.37E-05

Table 3.6. Top 20 SNP association results in GENOA African Americans

SNP	Chr	Position	Beta	MAF	Nearest Gene (bp)	Type	LR p-value
rs4979887	10	78921964	0.232	0.123	KCNMA1 (0)	intronic	3.00E-06
rs12554999	9	81232311	-0.179	0.200	CHCHD9 (35541)	intergenic	3.40E-06
rs4575062	1	163257670	-0.373	0.038	PBX1 (170001)	intergenic	3.45E-06
rs16921267	8	57977541	-0.294	0.063	IMPAD1 (55500)	intergenic	5.99E-06
rs11138235	9	81259492	-0.174	0.200	CHCHD9 (62722)	intergenic	6.19E-06
rs2687065	17	54610592	0.349	0.047	PRR11 (0)	intronic	8.08E-06
rs9911667	17	52134263	0.212	0.132	C17orf67 (90009)	unknown	8.53E-06
rs16948448	16	48980696	-0.364	0.038	BRD7 (20366)	intergenic	8.78E-06
rs10760850	9	103992278	-0.242	0.093	GRIN3A (451595)	intergenic	9.54E-06
rs13158524	5	128915938	0.140	0.365	ADAMTS19 (0)	intronic	1.20E-05
rs10121972	9	103991720	0.207	0.128	GRIN3A (451037)	intergenic	1.29E-05
rs1325510	1	117455026	-0.449	0.023	TRIM45 (184)	near-gene-3	1.52E-05
rs4351766	10	62505612	0.132	0.454	RHOBTB1 (74408)	intergenic	1.56E-05
rs1335989	10	33104406	-0.154	0.259	C10orf68 (0)	intronic	1.57E-05
rs7977839	12	83607592	0.272	0.069	SLC6A15 (169807)	intergenic	1.58E-05
rs7114911	11	42546233	0.172	0.194	API5 (743847)	intergenic	1.64E-05
rs7089366	10	32998579	-0.153	0.255	C10orf68 (0)	intronic	1.74E-05
rs10952598	7	144330199	0.235	0.092	TPK1 (166120)	intergenic	1.77E-05
rs4959299	6	4437078	-0.161	0.238	CDYL (214313)	intergenic	1.81E-05
rs2254160	21	29174433	0.175	0.175	N6AMT1 (0)	intronic	1.82E-05

Table 3.7. Characteristics of ARIC and GENOA African American participants included in meta-analysis

	ARIC (N=428) Mean (SD)	GENOA (N=501) Mean (SD)
Age at MRI, years	71.8 (4.5)	62.7 (8.9)
Ln (Leukoaraiosis+1)	2.41 (0.71)	2.17 (0.59)
Total Intracranial Volume, cm^3	1405 (139)	1372 (134)
Systolic Blood Pressure, mm Hg	130.2 (20.5)	136.9 (20.4)
Body Mass Index, kg/m^2	29.0 (4.8)	31.4 (5.9)
Total cholesterol, mg/dL	208.4 (40.4)	201.8 (42.3)
Female, %	269 (62.9)	352 (70.3)
Coronary Heart Disease, %	1.2	4.4
Ever Smoker, %	48.2	39.3
Hypertension, %	57.5	77.0

Table 3.8. SNP association results with p-value <$1x10^{-5}$ in the meta-analysis of the full GENOA and ARIC African American samples

SNP	Chr	Position	Beta	Beta SE	MAF	Beta (ARIC)	Beta (GENOA)	Beta SE (ARIC)	Beta SE (GENOA)	Nearest Gene (bp)	Type	P-value
rs1945938	11	130484613	0.15	0.03	0.28	0.19	0.13	0.05	0.04	SNX19 (193021)	intergenic	5.16E-07
rs10016567	4	83622092	-0.28	0.06	0.05	-0.31	-0.27	0.11	0.07	FLJ12993 (2535)	intergenic	1.93E-06
rs11019290	11	90488809	-0.27	0.06	0.06	-0.31	-0.25	0.09	0.08	CHORDC1 (892629)	intergenic	3.49E-06
rs12554999	9	81232311	-0.15	0.03	0.19	-0.14	-0.16	0.06	0.04	CHCHD9 (35541)	intergenic	3.52E-06
rs10863562	1	218995778	-0.13	0.03	0.62	-0.14	-0.12	0.05	0.04	MOSC2 (0)	intronic	3.67E-06

Figure 3.5. Distribution of –log p-values from single SNP associations with residual values of ln(leukoaraiosis+1) in the meta-analysis of the full GENOA and ARIC African American samples (N=929)

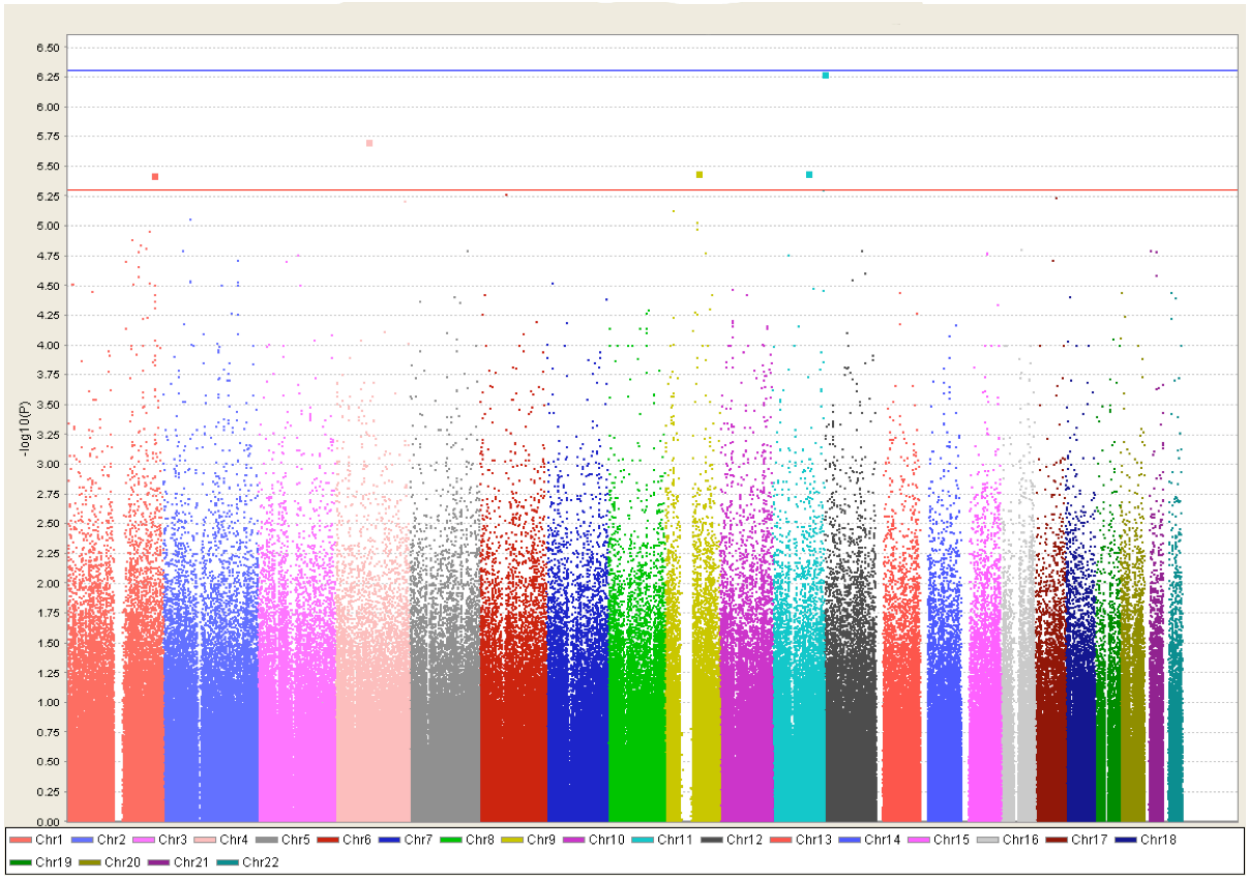


Table 3.9. SNP association results with p-value <1x10^{-5} in the meta-analysis of the GENOA and ARIC African American samples excluding participants with two *APOE* ε4 alleles (N=883)

SNP	Chr	Position	Beta	Beta SE	MAF	Beta (ARIC)	Beta (GENOA)	Beta SE (ARIC)	Beta SE (GENOA)	Nearest Gene (bp)	Type	P-value
rs10016567	4	83622092	-0.29	0.06	0.05	-0.29	-0.29	0.11	0.07	FLJ12993 (2535)	intergenic	5.39E-07
rs13061385	3	104496273	0.15	0.03	0.21	0.12	0.16	0.06	0.04	ZPLD1 (814898)	intergenic	3.47E-06
rs4299533	4	178321087	-0.15	0.03	0.23	-0.12	-0.16	0.05	0.04	NEIL3 (146897)	intergenic	3.65E-06
rs1456390	4	154939028	-0.25	0.05	0.07	-0.36	-0.20	0.10	0.06	SFRP2 (9350)	intergenic	3.66E-06
rs9299702	10	33331688	-0.13	0.03	0.33	-0.12	-0.14	0.05	0.04	ITGB1 (44389)	intergenic	4.74E-06
rs10863562	1	218995778	-0.12	0.03	0.39	-0.12	-0.13	0.05	0.03	MOSC2 (0)	intronic	4.90E-06
rs2687065	17	54610592	-0.29	0.06	0.05	-0.15	-0.36	0.12	0.08	PRR11 (0)	intronic	6.25E-06
rs1945938	11	130484613	0.13	0.03	0.28	0.16	0.12	0.05	0.04	SNX19 (193021)	intergenic	8.18E-06
rs7898823	10	33356504	0.16	0.04	0.18	0.13	0.17	0.06	0.04	ITGB1 (69205)	intergenic	9.20E-06
rs11019290	11	90488809	-0.26	0.06	0.06	-0.26	-0.26	0.09	0.08	CHORDC1 (892629)	intergenic	9.37E-06

Figure 3.6. Distribution of –log p-values from single SNP associations with residual values of ln(leukoaraiosis+1) in the meta-analysis of the GENOA and ARIC African American samples excluding participants with two *APOE* ε4 alleles (N=883)

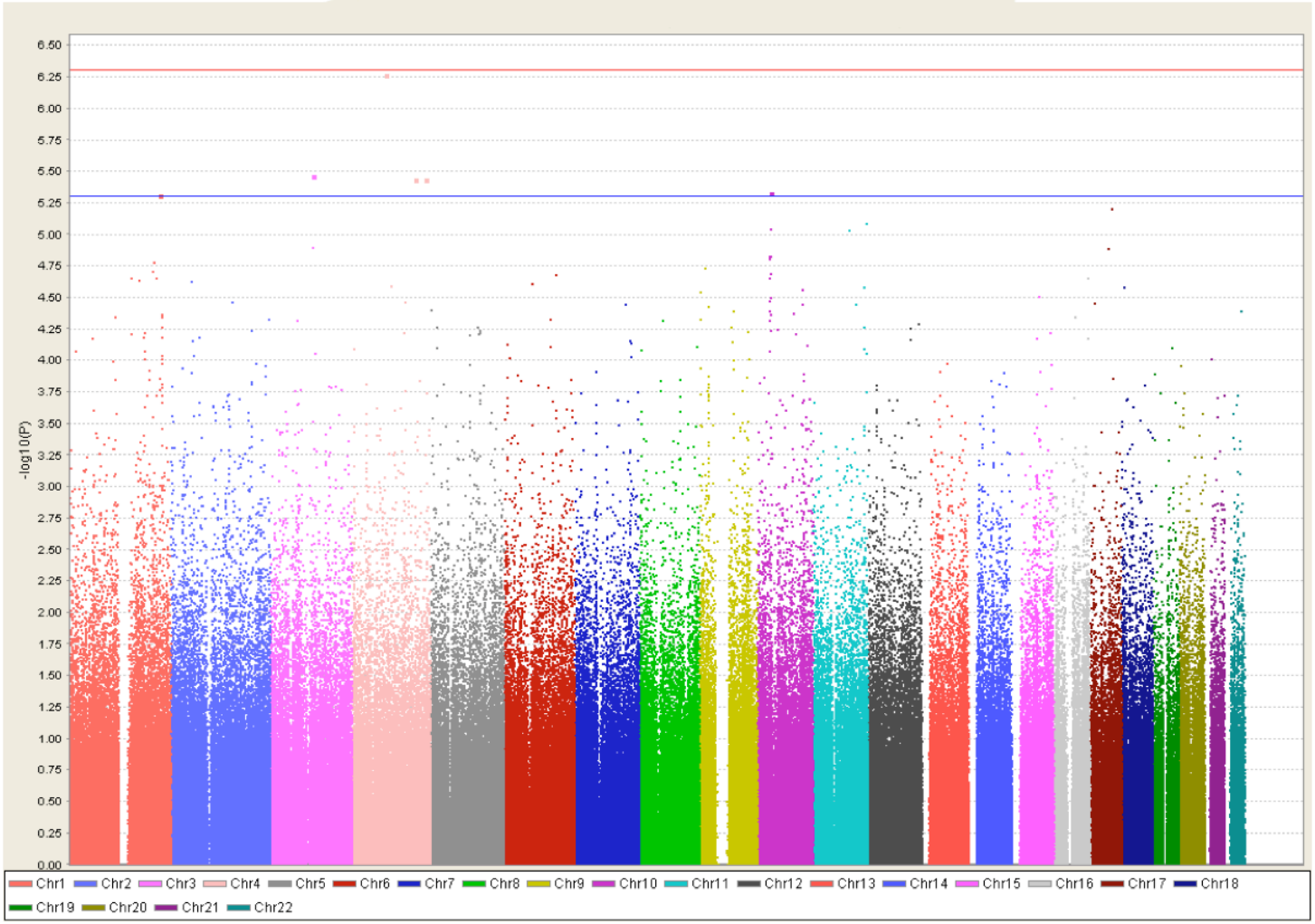


Table 3.10. SNP association results with p-value <1x10^{-5} in the meta-analysis of the GENOA and ARIC African American samples excluding participants with at least one *APOE* ε4 allele (N=602)

SNP	Chr	Position	Beta	Beta SE	MAF	Beta (ARIC)	Beta (GENOA)	Beta SE (ARIC)	Beta SE (GENOA)	Nearest Gene (bp)	Type	P-value
rs9299702	10	33331688	-0.21	0.04	0.33	-0.22	-0.20	0.07	0.05	ITGB1 (44389)	intergenic	3.73E-08
rs7898823	10	33356504	0.23	0.05	0.18	0.21	0.24	0.08	0.05	ITGB1 (69205)	intergenic	3.20E-07
rs10891151	11	110009254	-0.22	0.04	0.20	-0.17	-0.25	0.07	0.06	ARHGAP20 (0)	intronic	3.42E-07
rs12264562	10	68361028	0.23	0.05	0.15	0.17	0.27	0.08	0.06	CTNNA3 (0)	intronic	5.92E-07
rs17627013	10	36679140	-0.42	0.09	0.03	-0.56	-0.38	0.17	0.10	FZD8 (708772)	intergenic	7.38E-07
rs17615892	10	36675415	0.43	0.09	0.03	0.56	0.37	0.17	0.11	FZD8 (705047)	intergenic	2.69E-06
rs1456390	4	154939028	-0.30	0.07	0.07	-0.44	-0.25	0.12	0.08	SFRP2 (9350)	intergenic	3.05E-06
rs6741482	2	195955990	-0.28	0.06	0.09	-0.21	-0.31	0.11	0.07	SLC39A10 (273786)	intergenic	3.51E-06
rs7958457	12	52166920	-0.33	0.07	0.05	-0.29	-0.36	0.11	0.09	MAP3K12 (0)	intronic	4.09E-06
rs9365351	6	162260612	0.15	0.03	0.46	0.22	0.12	0.06	0.04	PARK2 (0)	intronic	5.50E-06
rs12062763	1	150679890	0.18	0.04	0.25	0.16	0.19	0.07	0.05	CRNN (26516)	intergenic	6.36E-06
rs719534	9	118515703	-0.17	0.04	0.25	-0.15	-0.18	0.07	0.05	ASTN2 (0)	intronic	6.62E-06
rs13339693	17	12059882	-0.18	0.04	0.21	-0.15	-0.19	0.07	0.05	MAP2K4 (72106)	intergenic	7.64E-06
rs1368269	3	71754567	-0.30	0.07	0.06	-0.20	-0.33	0.12	0.08	FOXP1 (38737)	intergenic	7.76E-06
rs9365352	6	162260940	0.15	0.03	0.41	0.23	0.10	0.05	0.04	PARK2 (0)	intronic	8.03E-06
rs11893056	2	50731206	0.31	0.07	0.07	0.48	0.25	0.13	0.08	NRXN1 (0)	intronic	8.24E-06

Figure 3.7. Distribution of –log p-values from single SNP associations with residual values of ln(leukoaraiosis+1) in the meta-analysis of the GENOA and ARIC African American samples excluding participants with at least one *APOE* ε4 allele (N=602)

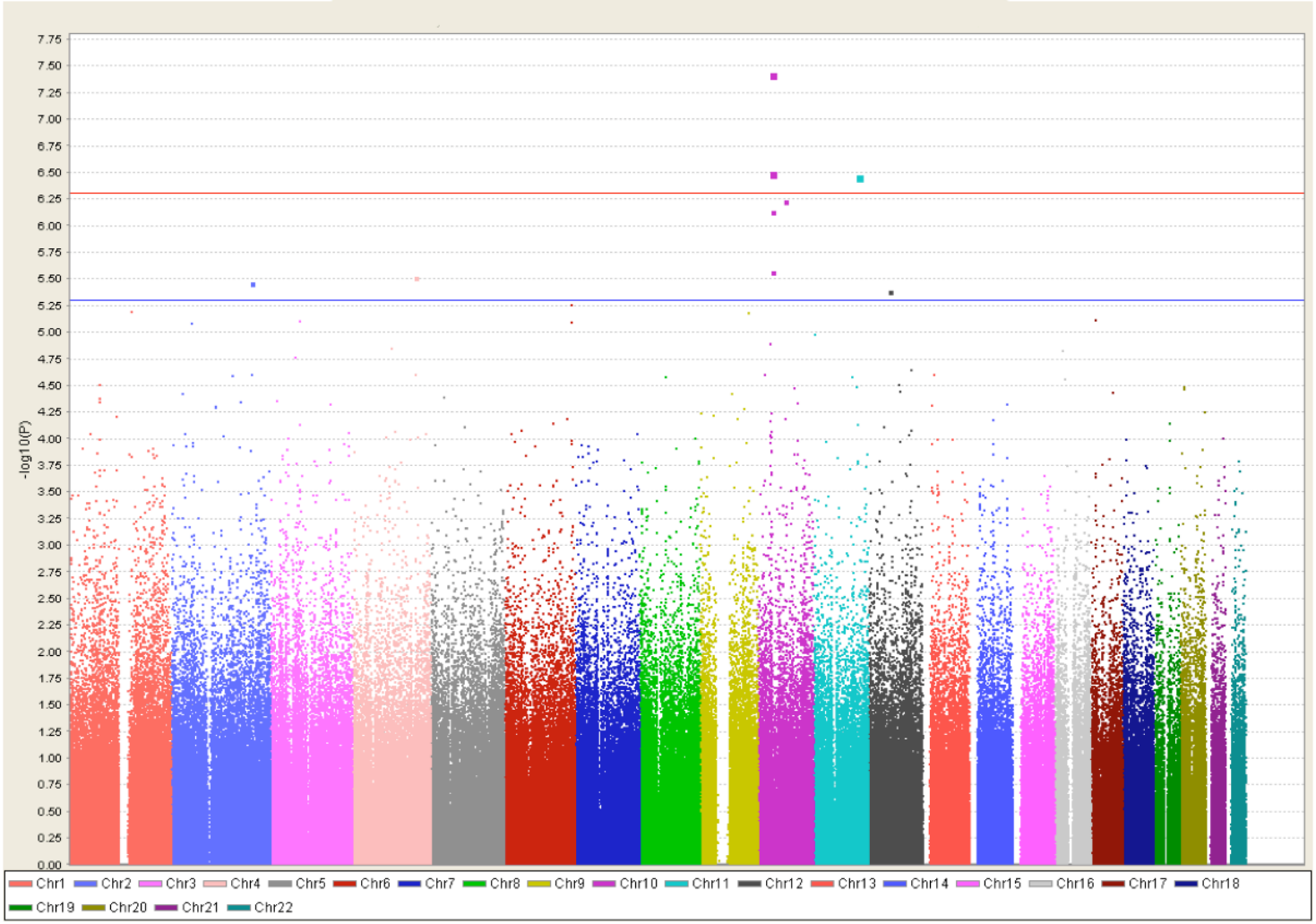
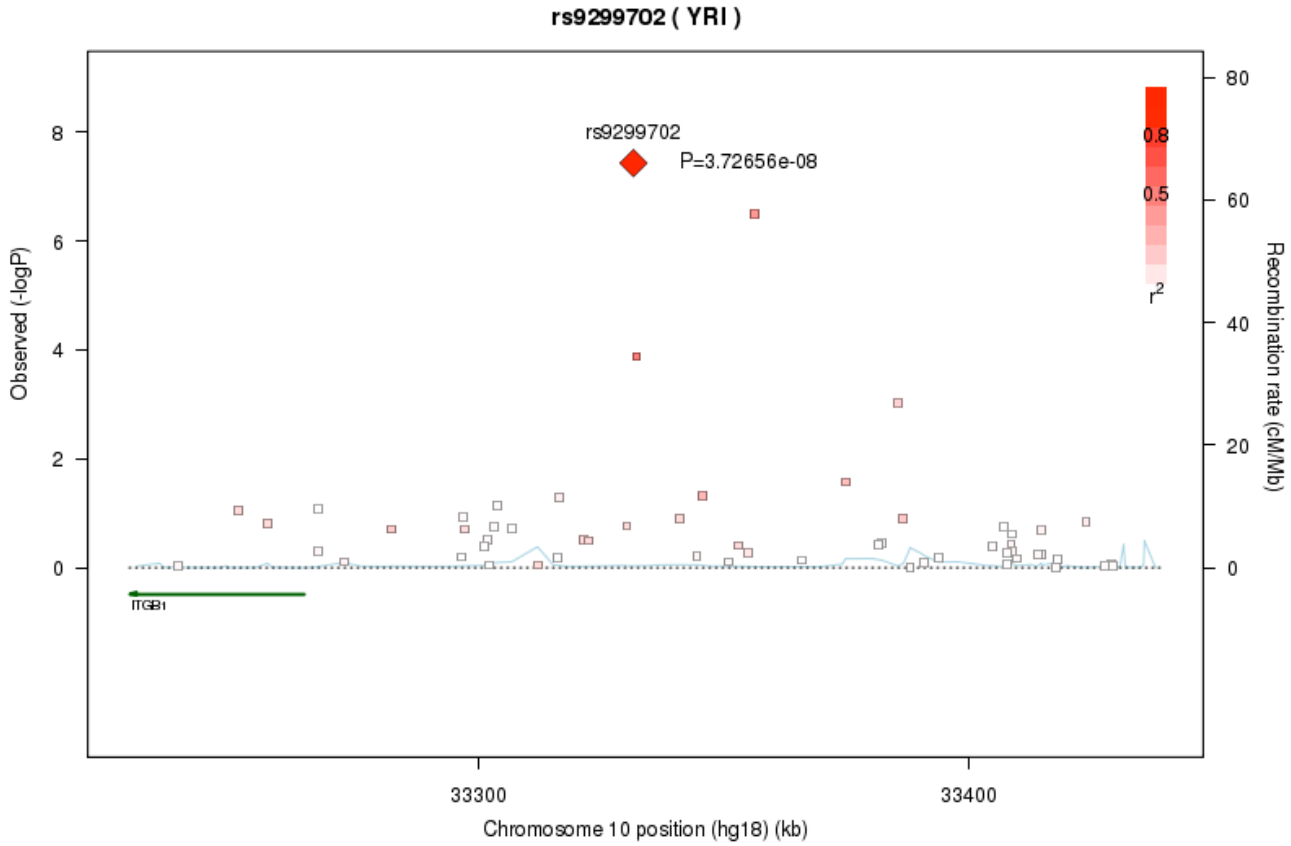
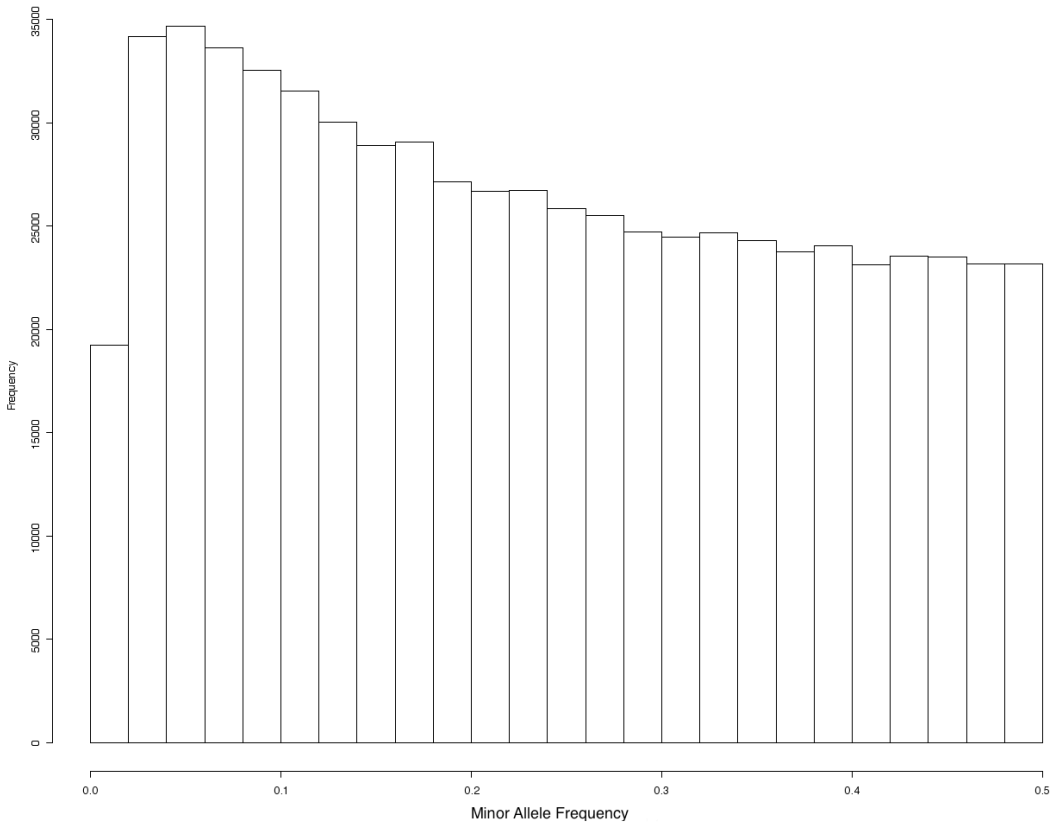


Figure 3.8. Top single SNP association with residual values of ln(leukoaraiosis+1) in the meta-analysis of the GENOA and ARIC African American samples excluding participants with at least one *APOE* ε4 allele (N=602)

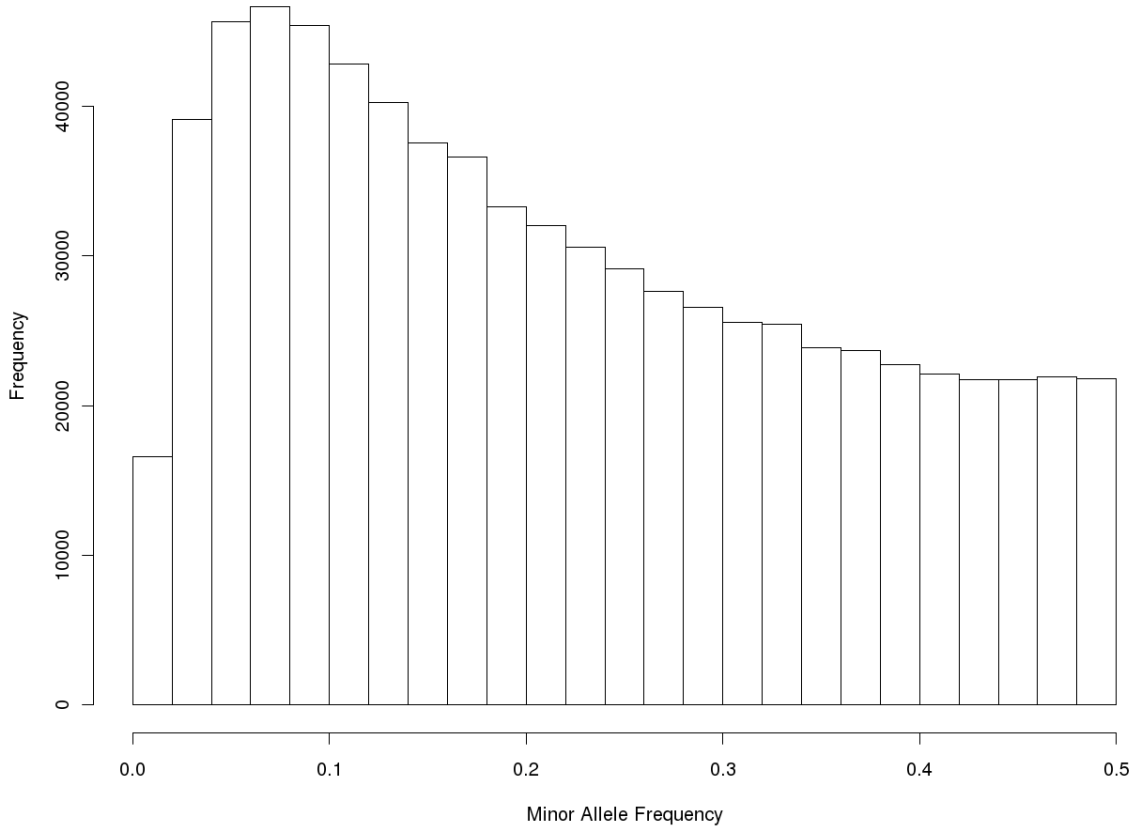


Appendix 3.1. Distribution of minor allele frequencies for Affymetrix 6.0 SNPs

Distribution of minor allele frequencies for the 668,293 Affymetrix 6.0 SNPs in GENOA whites



Distribution of minor allele frequencies for the 762,766 Affymetrix 6.0 SNPs in GENOA African Americans



Appendix 3.2. Characteristics of GENOA participants with and without Affymetrix 6.0 genotype data

GENOA whites

	Participants with Genotype Data		Participants without Genotype Data		T-test p-value
	N	Mean (SD)	N	Mean (SD)	
Leukoaraiosis Volume, cm^3	759	7.89 (6.66)	122	6.39 (3.68)	0.0003
Total Intracranial Volume, cm^3	759	1466 (147)	122	1472 (147)	0.6393
Total Brain Volume, cm^3	759	1158 (123)	122	1166 (116)	0.5106
Brain Atrophy (TIV – Brain), cm^3	759	307.4 (74.6)	122	306.6 (69.1)	0.9053
Ventricular Volume, cm^3	759	25.5 (17.0)	122	24.7 (12.8)	0.5569
Time from Exam to MRI, years	759	1.20 (0.77)	122	1.32 (0.75)	0.0839
Age at MRI, years	759	60.5 (9.89)	122	58.5 (9.31)	0.0255
Age at Phase II Exam, years	759	59.3 (10.1)	122	57.2 (9.63)	0.0216
Body Mass Index, kg/m^2	759	30.4 (5.8)	122	30.9 (6.3)	0.3977
Height, cm	759	168 (9.2)	122	168 (8.6)	0.9030
Weight, kg	552	85.9 (18.2)	122	87.6 (20.2)	0.3877
Waist-to-hip ratio	759	0.91 (0.11)	122	0.89 (0.10)	0.1185
Systolic Blood Pressure, mm Hg	758	131.3 (16.5)	122	130.3 (17.2)	0.5230
Diastolic Blood P, mm Hg	758	74.1 (9.1)	122	74.6 (8.8)	0.5262
Pulse pressure, mm Hg	758	57.3 (15.3)	122	55.7 (13.6)	0.2347
Total cholesterol, mg/dL	759	197.8 (33.7)	122	199.6 (37.2)	0.6149
HDL cholesterol, mg/dL	759	52.1 (14.7)	122	52.9 (14.9)	0.5602
LDL cholesterol, mg/dL	759	120.6 (31.8)	122	121.3 (35.0)	0.8277
Physical activity	759	-9.93 (7.52)	122	-10.0 (8.04)	0.9229
Female, n (%)	759	446 (58.8%)	122	80 (65.6%)	0.1854
Ever smoker, n (%)	759	366 (48.2%)	122	58 (47.5%)	0.9665
Coronary heart disease, n (%)	759	57 (7.5%)	122	3 (2.5%)	0.0626
Hypertension, n (%)	759	557 (73.4%)	122	87 (71.3%)	0.7117

T-tests compare participants with and without genotype data.
All variables except for brain measures were measured at the Phase II GENOA examination.

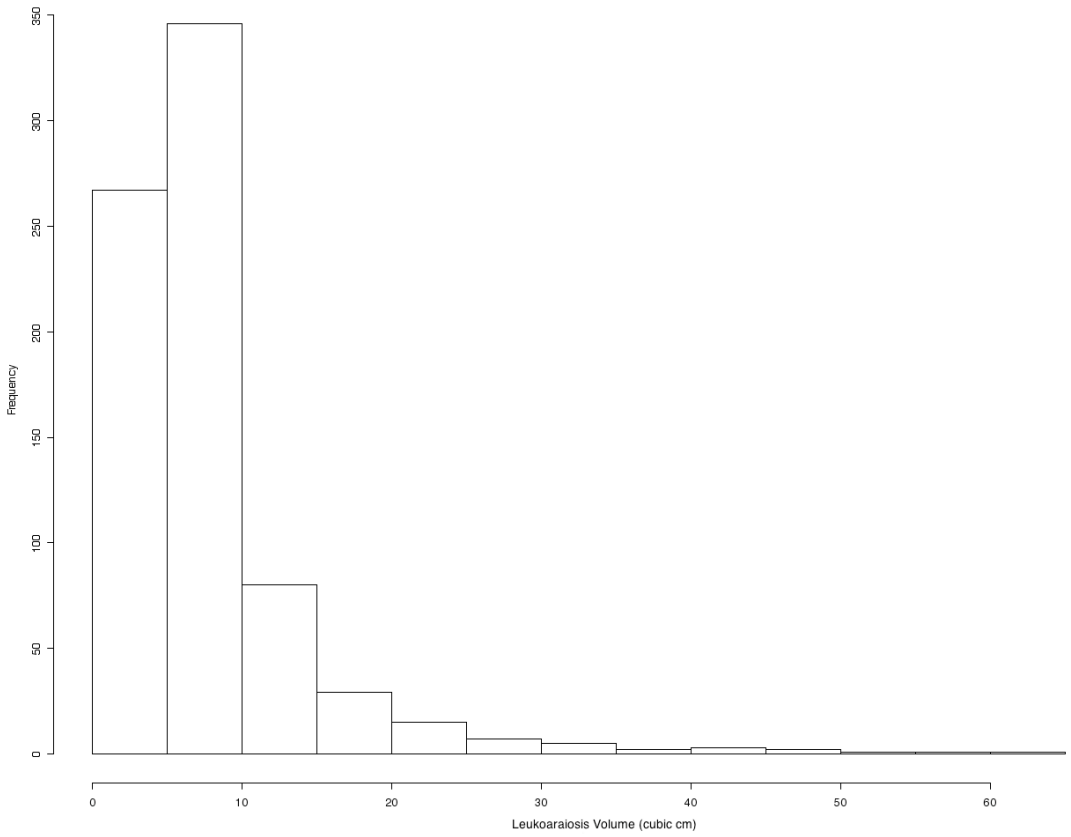
GENOA African Americans

	Participants with Genotype Data		Participants without Genotype Data		T-test p-value
	N	Mean (SD)	N	Mean (SD)	
Leukoaraiosis Volume, cm^3	553	10.55 (11.77)	242	10.43 (10.84)	0.8888
Total Intracranial Volume, cm^3	553	1374 (135)	242	1375 (132)	0.9775
Total Brain Volume, cm^3	553	1067 (114)	242	1063 (111)	0.6920
Brain Atrophy (TIV – Brain), cm^3	553	307.5 (75.2)	242	311.2 (71.8)	0.5081
Ventricular Volume, cm^3	553	22.3 (13.6)	242	22.8 (12.9)	0.6381
Time from Exam to MRI, years	553	1.16 (0.80)	242	1.22 (0.84)	0.3520
Age at MRI, years	553	63.5 (8.90)	242	64.7 (8.92)	0.0655
Age at Phase II Exam, years	553	62.3 (8.74)	242	63.5 (8.72)	0.0729
Body Mass Index, kg/m^2	552	31.2 (5.8)	242	30.5 (5.8)	0.1106
Height, cm	552	169 (9.0)	242	169 (8.3)	0.8676
Weight, kg	552	88.9 (16.6)	242	87.0 (17.8)	0.1601
Waist-to-hip ratio	552	0.89 (0.06)	241	0.88 (0.07)	0.0105
Systolic Blood Pressure, mm Hg	553	136.9 (20.3)	242	137.9 (19.3)	0.5343
Diastolic Blood P, mm Hg	553	79.7 (10.5)	242	79.2 (10.9)	0.5172
Pulse pressure, mm Hg	553	57.2 (16.4)	242	58.7 (17.1)	0.2567
Total cholesterol, mg/dL	547	201.9 (42.5)	241	203.3 (47.1)	0.6929
HDL cholesterol, mg/dL	547	57.7 (19.2)	241	60.2 (19.8)	0.0985
LDL cholesterol, mg/dL	547	123.1 (39.9)	241	122.2 (44.3)	0.7858
Physical activity	553	-13.3 (4.74)	242	-13.5 (4.37)	0.4541
Female, n (%)	553	381 (68.9%)	242	167 (69.0%)	0.9585
Ever smoker, n (%)	553	334 (60.4%)	242	152 (62.8%)	0.5735
Coronary heart disease, n (%)	553	24 (4.3%)	242	13 (5.4%)	0.6508
Hypertension, n (%)	553	428 (77.4%)	242	179 (74.0%)	0.3389

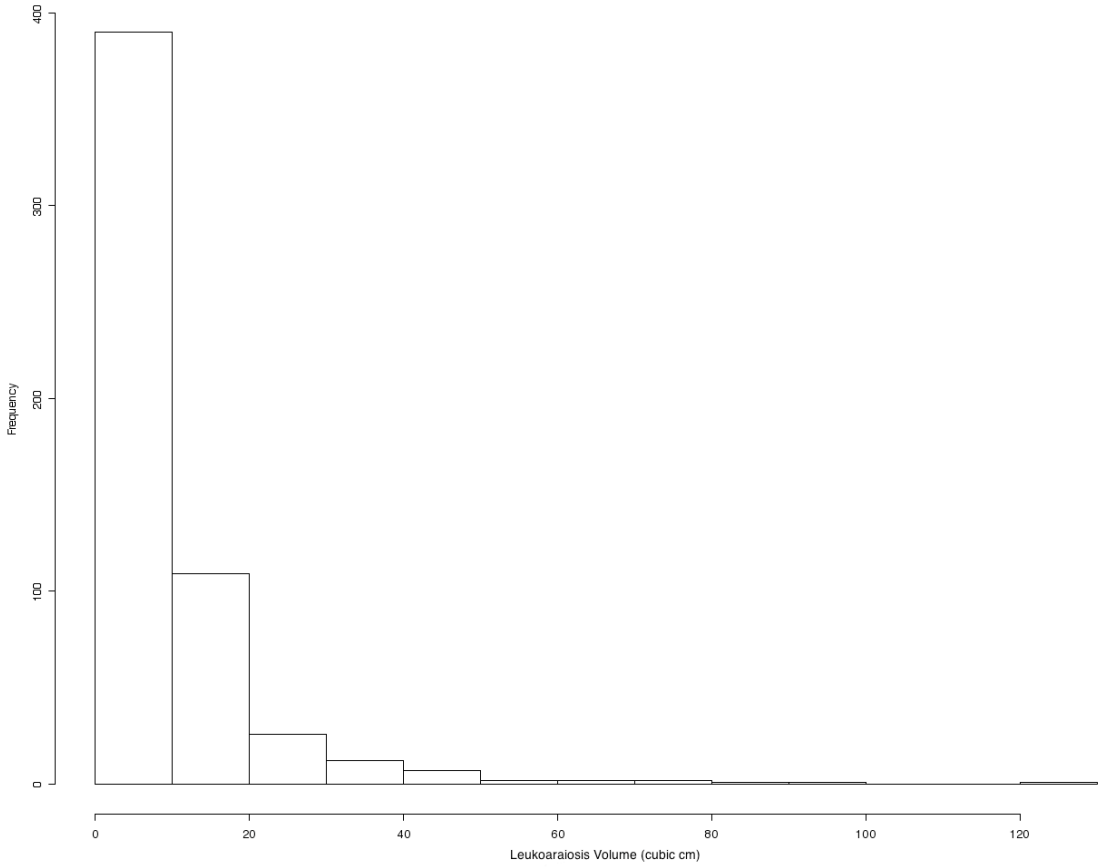
T-tests compare participants with and without genotype data.
All variables except for brain measures were measured at the Phase II GENOA examination.

Appendix 3.3. Distribution of leukoaraiosis volume before and after natural log transformation and adjustment for covariates

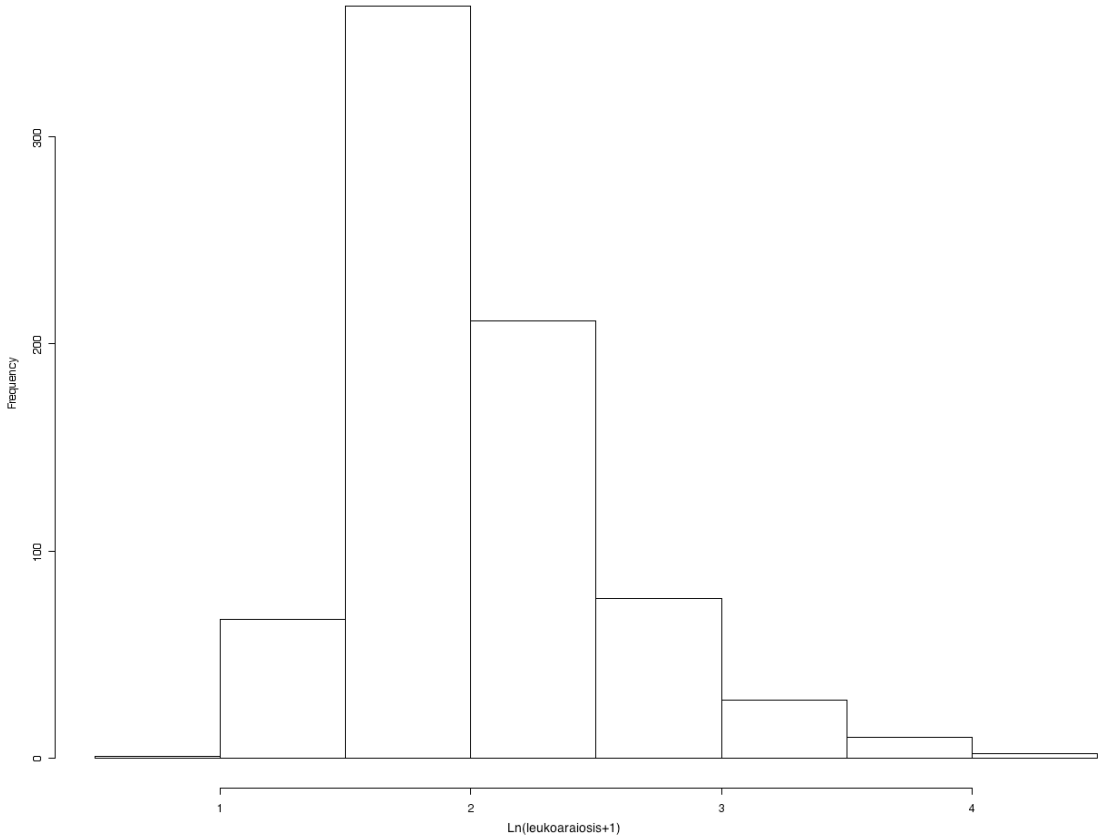
Distribution of leukoaraiosis volume in GENOA whites



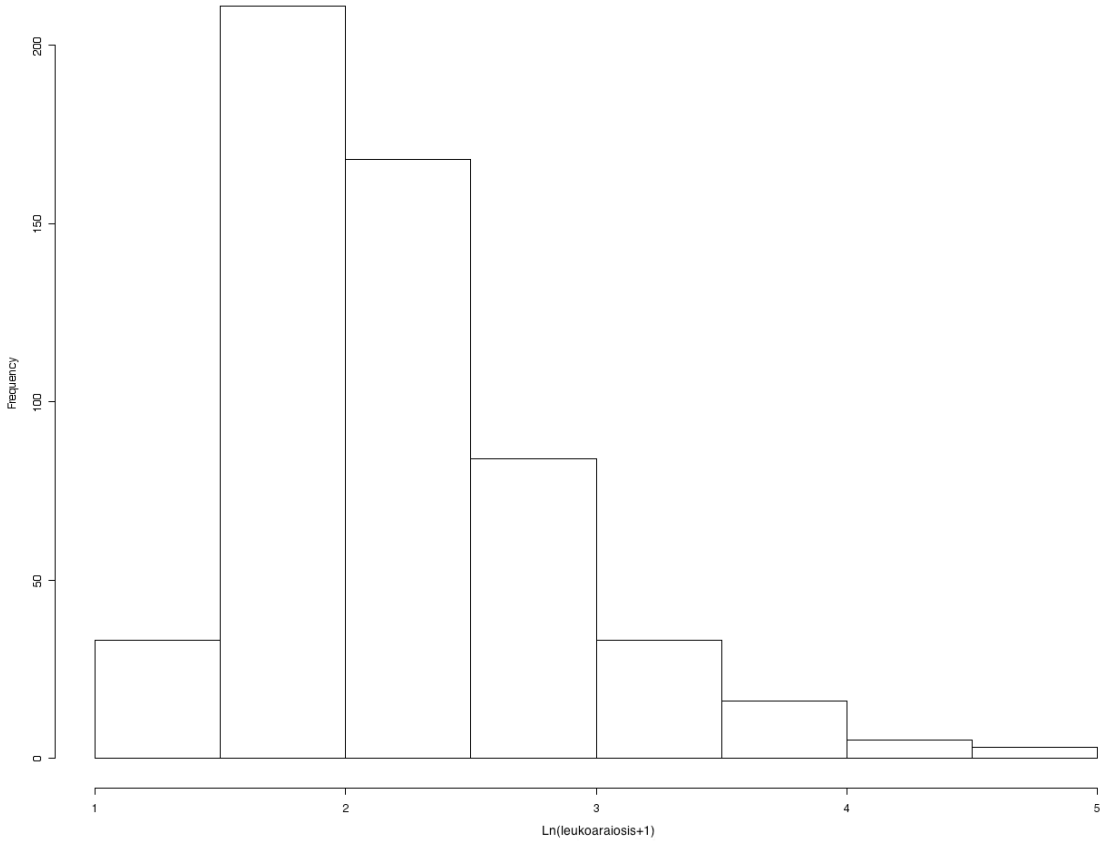
Distribution of leukoaraiosis volume in GENOA African Americans



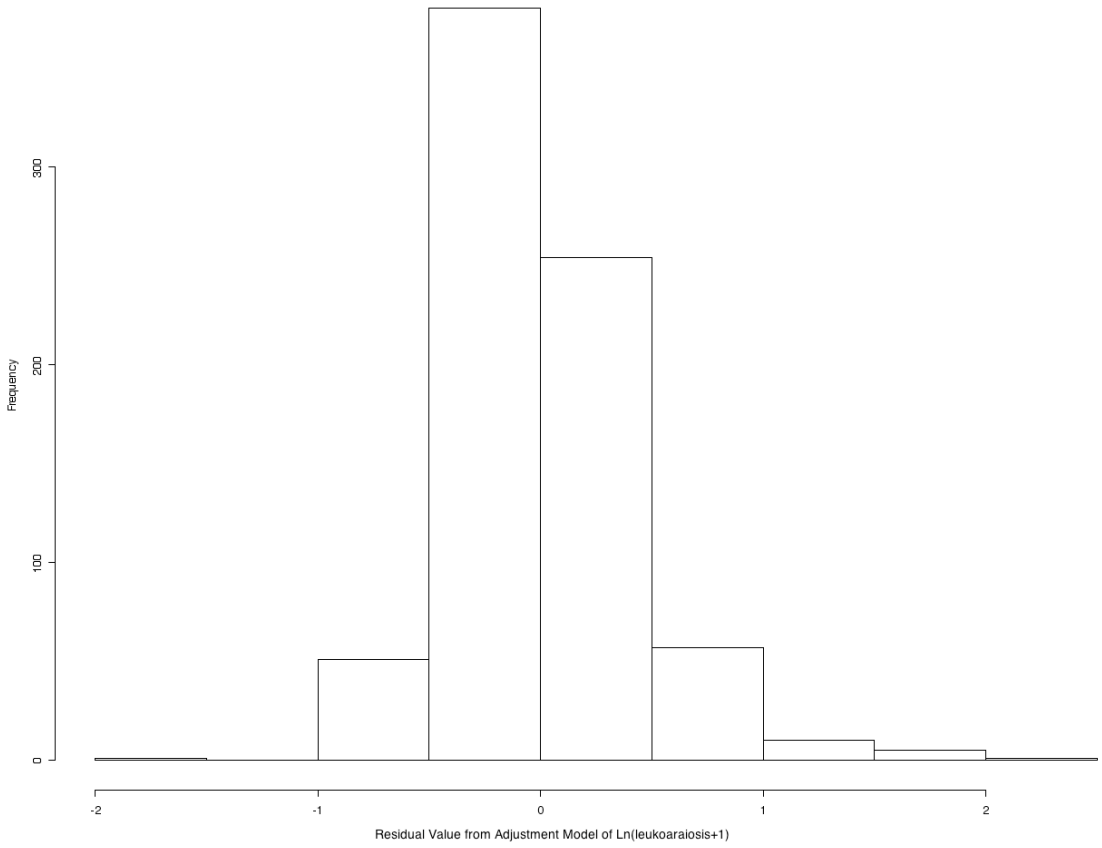
Distribution of ln(leukoaraiosis+1) in GENOA whites



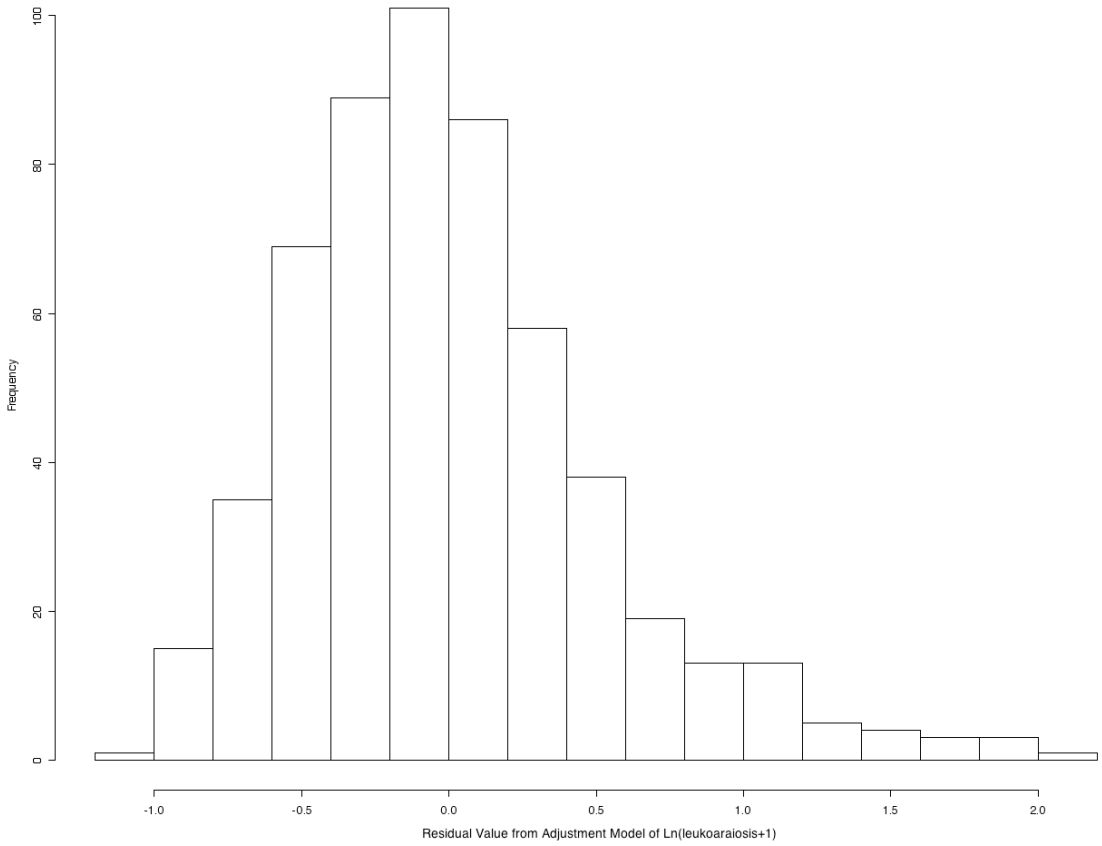
Distribution of ln(leukoaraiosis+1) in GENOA African Americans



Distribution of the residual values of ln(leukoaraiosis+1) after adjustment for age at MRI, gender, and total intracranial volume in GENOA whites



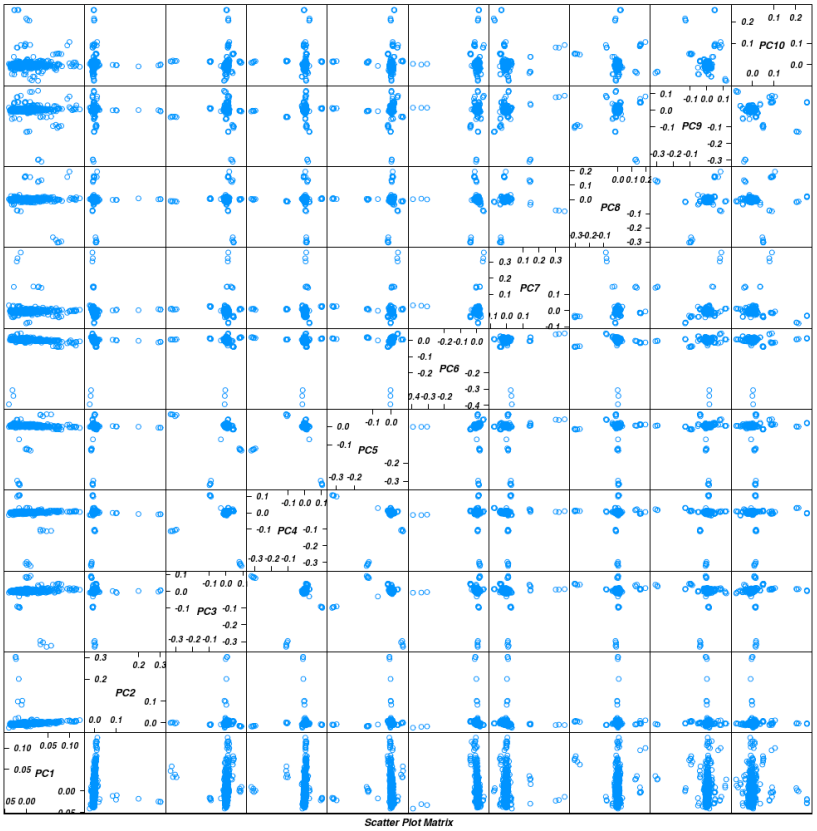
Distribution of the residual values of ln(leukoaraiosis+1) after adjustment for age at MRI, gender, total intracranial volume, and the first 10 principal components of Affymetrix genotypes in GENOA African Americans



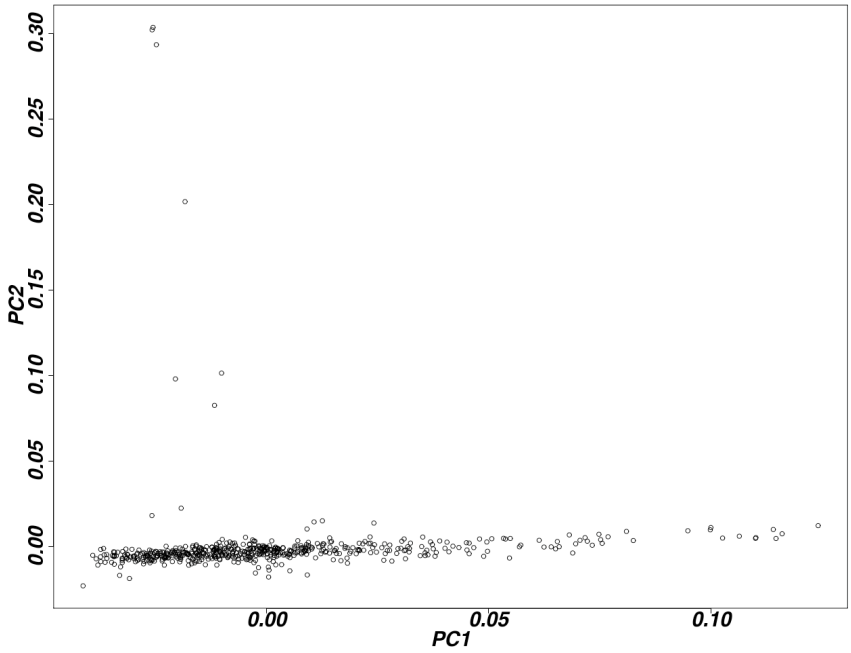
Appendix 3.4. Principal components from Affymetrix 6.0 genotype data in GENOA African Americans

In order to adjust for population structure in the African American sample, the first ten principal components of the Affymetrix 6.0 genotypes were calculated, and are presented in a scatterplot matrix below. Because it appeared that the some of the principal components may be capturing single sibships, a careful examination of the principal components of the full sample of African Americans (N=1,263 GENOA participants genotyped at the Mayo Clinic and an additional 183 GENOA participants genotyped at the University of Texas as part of the ARIC study) was conducted. Plots of the first and second principal components of the GENOA study sample (N=553) and the full GENOA sample (N=1,446) are presented below. Upon close examination, it is clear that the outliers in the second principal component were not due to a single family, since there are more points in the furthest outlying group than are in a single sibship. This pattern was observed for the majority of the other principal components as well. No outliers were removed, and all ten principal components were used to adjust for population structure, although only principal component nine was associated with ln(leukoaraiosis+1) adjusted for age, gender, and TIV.

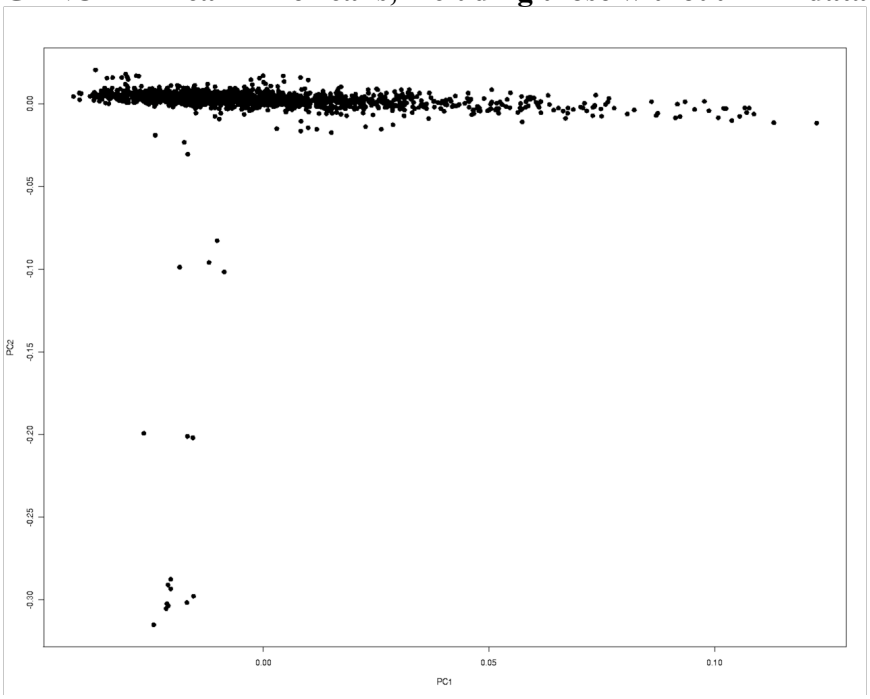
Scatterplot of the top ten principal components from Affymetrix 6.0 genotype data in GENOA African Americans with MRI data (N=553)



Scatterplot of PC1 and PC2 from Affymetrix 6.0 genotype data in GENOA African Americans with MRI data (N=553)



Scatterplot of PC1 and PC2 from Affymetrix 6.0 genotype data in the full sample of GENOA African Americans, including those without MRI data (N=1446)



Association of residual values of ln(leukoaraiosis) after adjustment for age, gender, and TIV with the top ten principal components from Affymetrix 6.0 genotype data in GENOA African Americans with MRI data (N=553)

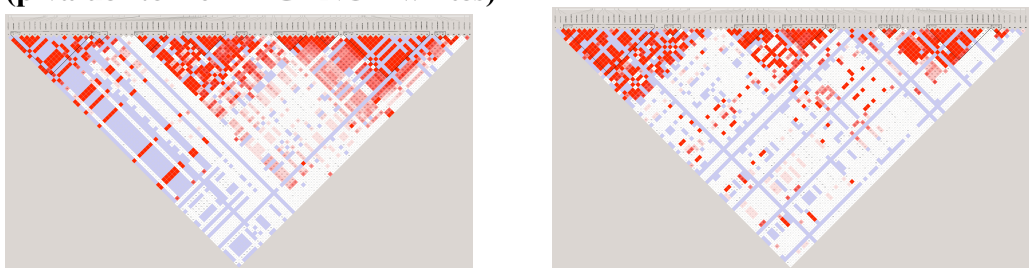
Principal Component	N	β	β SE	p-value
PC1	553	-0.9781	0.8475	0.2497
PC2	553	-0.7120	1.0326	0.4913
PC3	553	-0.0546	0.9188	0.9527
PC4	553	0.2527	0.9231	0.7845
PC5	553	0.4664	0.9210	0.6131
PC6	553	0.5488	0.9958	0.5821
PC7	553	-0.5822	0.8239	0.4806
PC8	553	0.3304	0.7946	0.6780
PC9	553	1.7521	0.8680	0.0448*
PC10	553	0.7066	0.8689	0.4170

* Significance indicated at p<=0.05 for t-tests comparing GENOA whites and African Americans

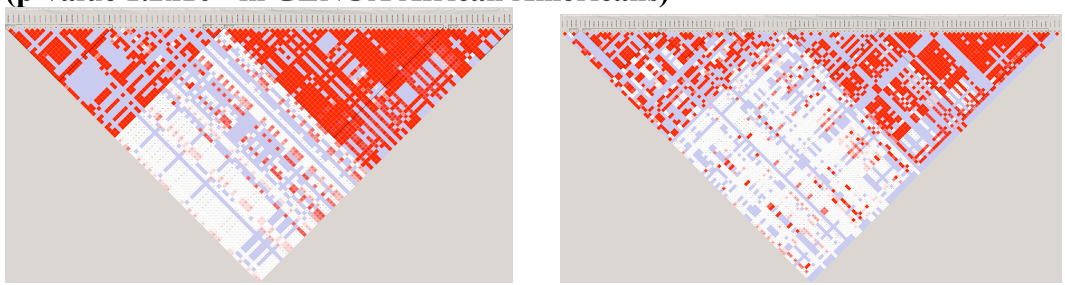
Appendix 3.5. Plots of linkage disequilibrium in regions near selected top GENOA findings in 60 singletons from the HapMap CEPH European (CEU) and African Yoruban (YRI) panels

Below are LD plots for two regions harboring a SNP that showed association with leukoaraiosis in either the white or African American GENOA sample. LD plots are color-coded according to the following scheme: white=(|D′|<1, LOD<2) shades of pink/red=(|D′|<1, LOD≥2), blue=(|D′|=1, LOD<2), bright red=(|D′|=1, LOD≥2). The numbers in the squares are the D′ value. Presented first are LD plots within 75kb of the associated SNP. These plots illustrate that while there are similarities in the overall pattern of regional LD in European and African populations, the extent of LD is much greater in Europeans. Thus, the SNP of interest identified in GENOA may be highly correlated with a larger number of neighboring SNPs in whites than in African Americans. Next, LD plots of the same region but within only 20kb of the SNP of interest are presented, with the SNP of interest indicated by a red arrow. These plots show that the SNPs of interest are correlated with sets of neighboring SNPs to a different degree in the populations.

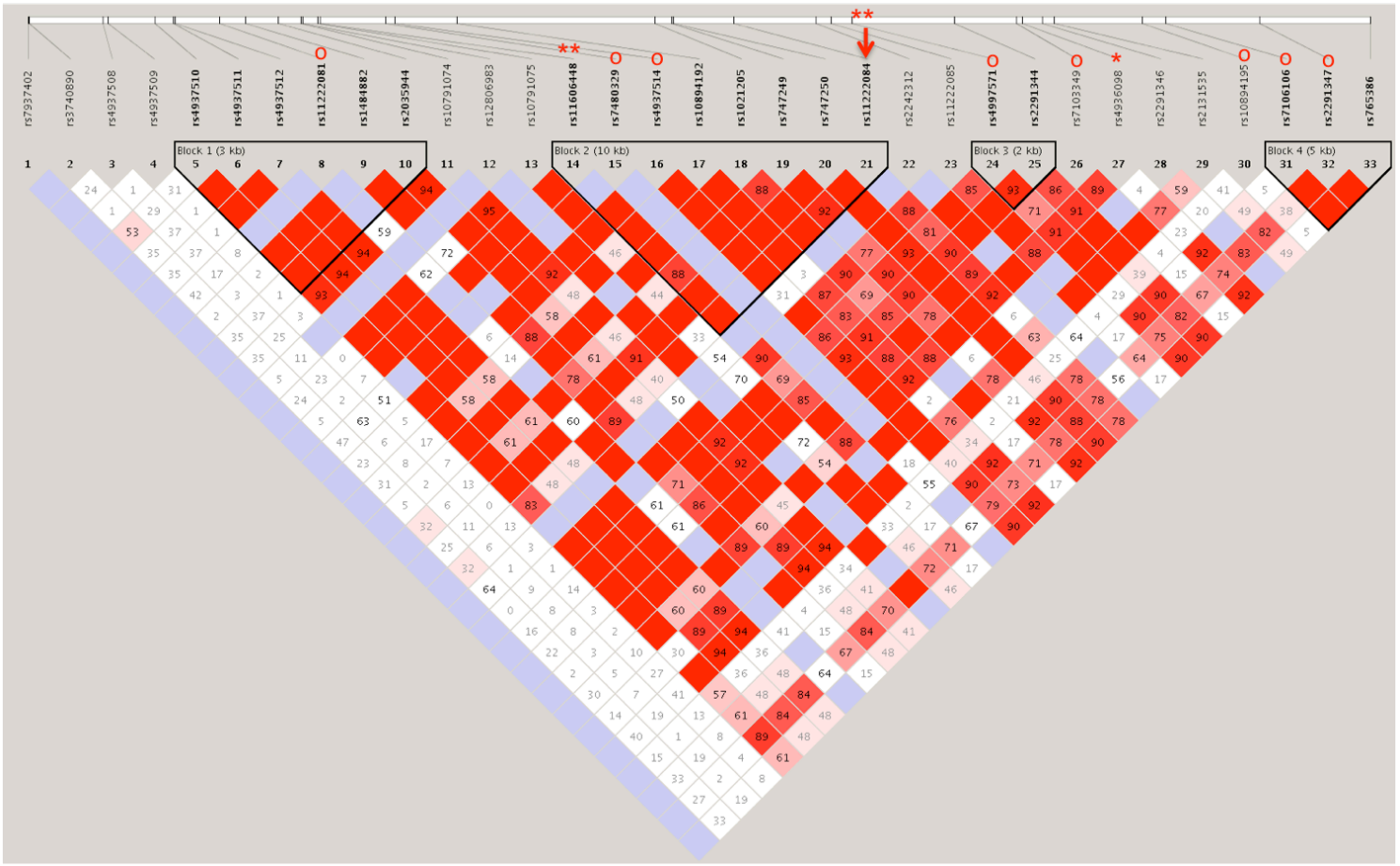
LD in the HapMap CEU panel (left) and YRI panel (right) within 75kb of rs11222084 near *ADAMTS8* (p-value 2.6x10^{-5} in GENOA whites)



LD in the HapMap CEU panel (left) and YRI panel (right) within 75kb of rs13158524 in an intron of *ADAMTS19* (p-value 1.2x10^{-5} in GENOA African Americans)

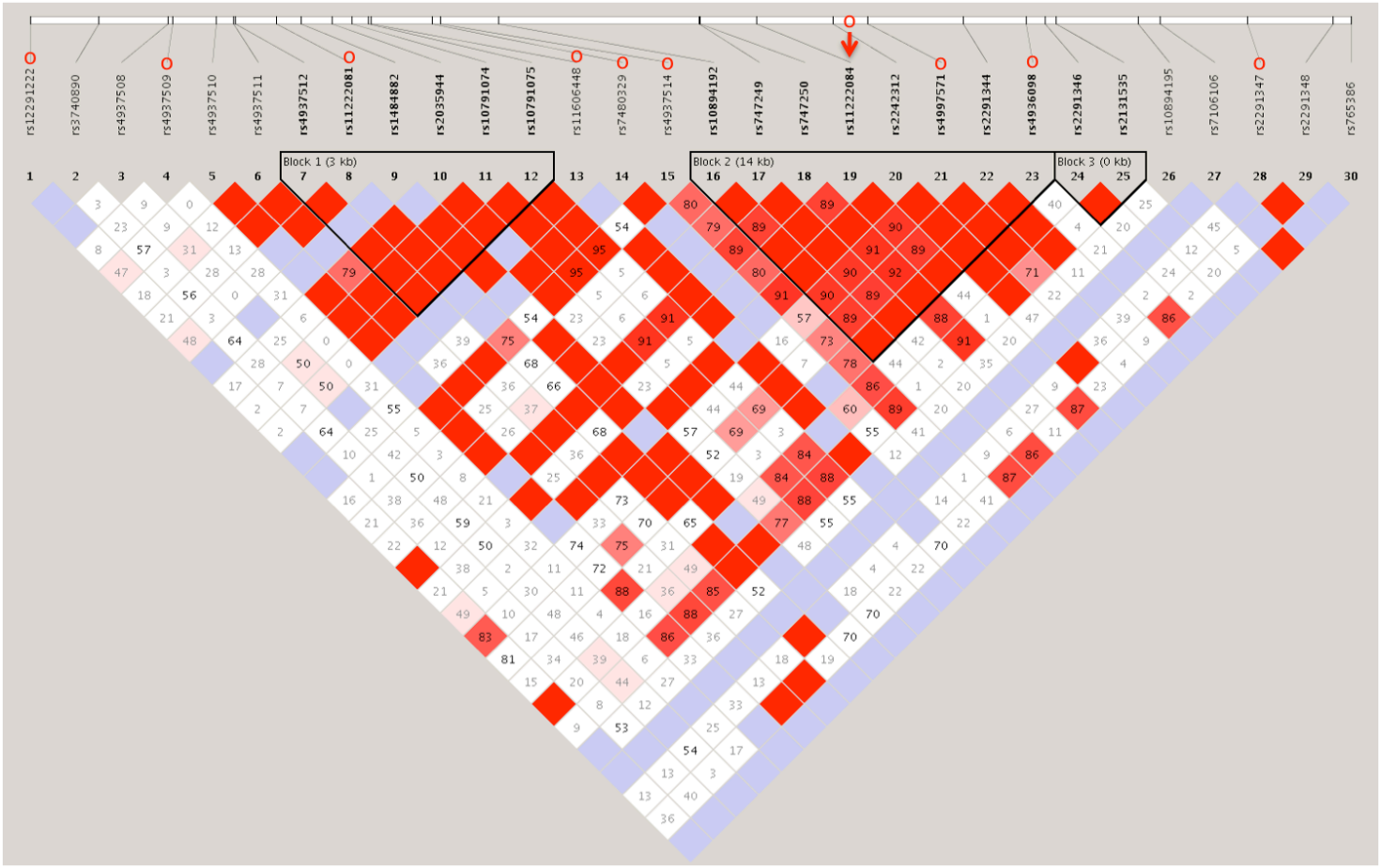


LD in the HapMap CEU panel within 20kb of rs11222084 near *ADAMTS8* (p-value 2.6x10^{-5} in GENOA whites)



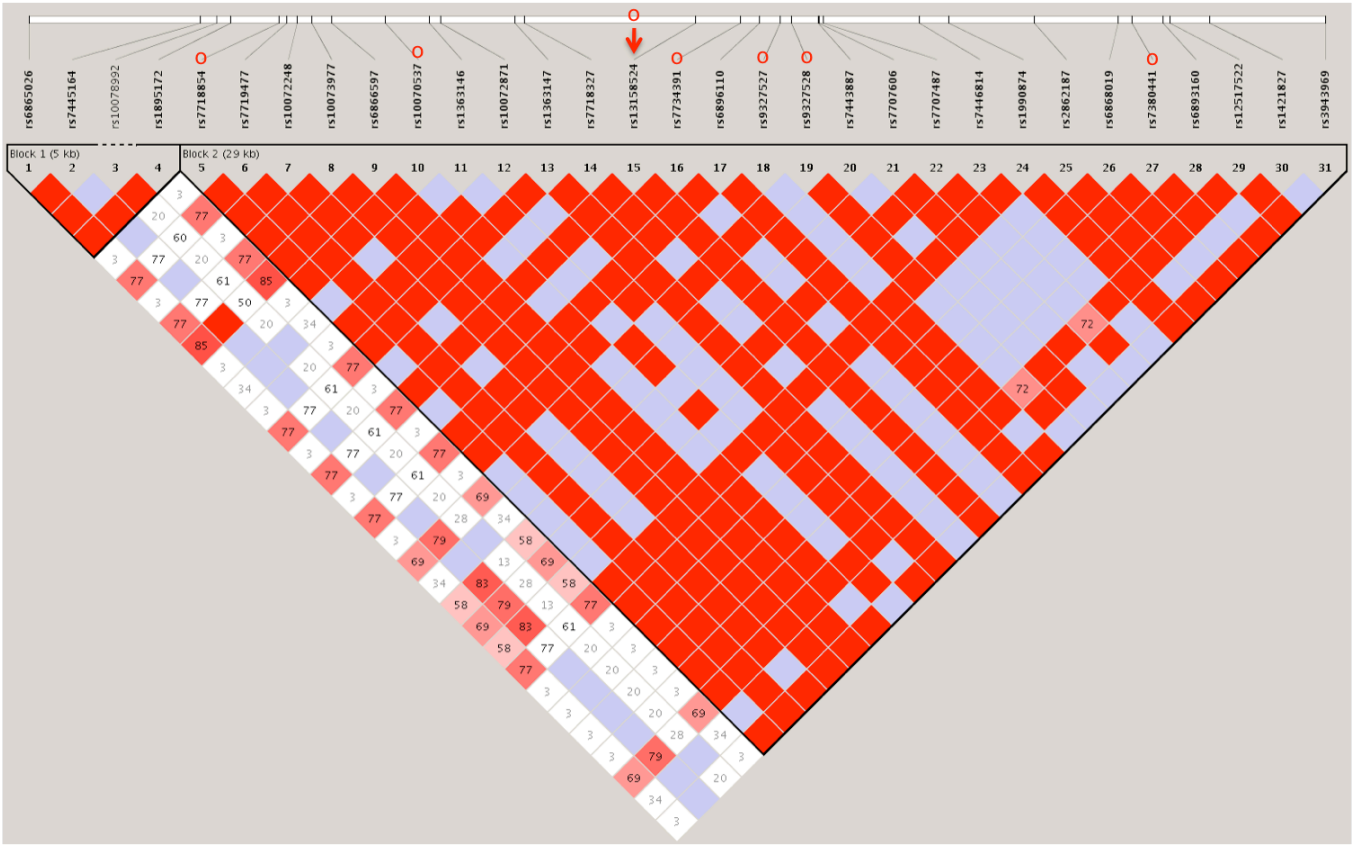
The region shown is between positions 129758000 and 129798000 on chromosome 11.
rs11222084 is SNP number 21, located in haplotype block 2 (10kb).
** = p-value < 1x10^{-3} in GENOA whites
* = 1x10^{-3} < p-value < 0.01 in GENOA whites
° = p-value < 0.01 in GENOA whites
No marking = not genotyped on the Affymetrix 6.0 platform

LD in the HapMap YRI panel within 20kb of rs11222084 near *ADAMTS8* (p-value 2.6x10^{-5} in GENOA whites)



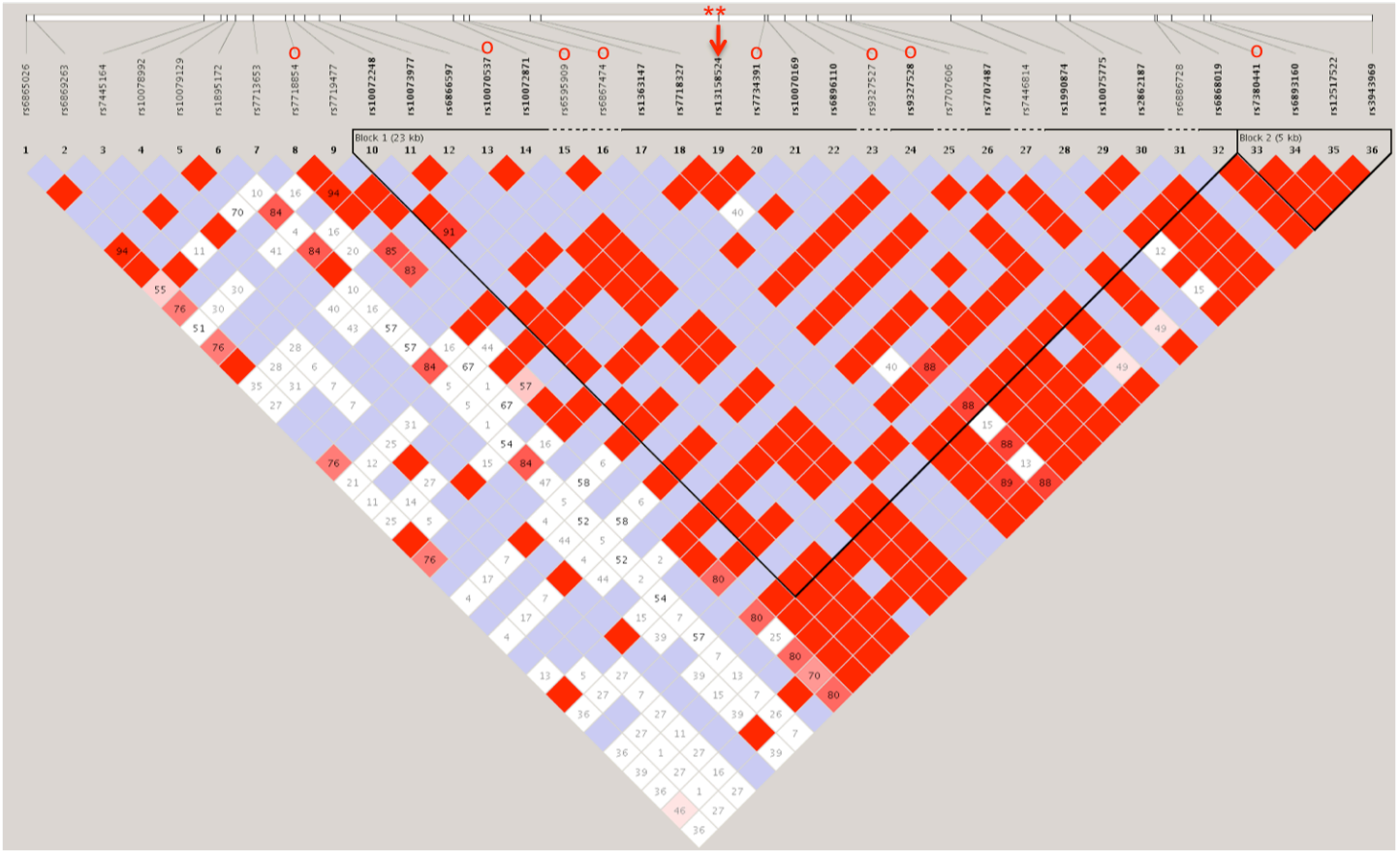
The region shown is between positions 129758000 and 129798000 on chromosome 11.
rs11222084 is SNP number 19, located in haplotype block 2 (14kb)
** = p-value < 1x10^{-3} in GENOA African Americans
* = 1x10^{-3} < p-value < 0.01 in GENOA African Americans
° = p-value < 0.01 in GENOA African Americans
No marking = not genotyped on the Affymetrix 6.0 platform

LD in the HapMap CEU panel within 20kb of rs13158524 in an intron of *ADAMTS19* (p-value 1.2x10^{-5} in GENOA African Americans)



The region shown is between positions 128895000 and 128935000 on chromosome 5.
rs13158524 is SNP number 15, located in haplotype block 2 (29kb)
** = p-value < 1x10^{-3} in GENOA whites
* = 1x10^{-3} < p-value < 0.01 in GENOA whites
° = p-value < 0.01 in GENOA whites
No marking = not genotyped on the Affymetrix 6.0 platform

LD in the HapMap YRI panel within 75kb of rs13158524 in an intron of *ADAMTS19* (p-value 1.2x10^{-5} in GENOA African Americans)



The region shown is between positions 128895000 and 128935000 on chromosome 5.
rs13158524 is SNP number 19, located in haplotype block 1 (23kb)
** = p-value < 1x10^{-3} in GENOA African Americans
* = 1x10^{-3} < p-value < 0.01 in GENOA African Americans
° = p-value < 0.01 in GENOA African Americans
No marking = not genotyped on the Affymetrix platform

References

Affymetrix. (2007a). Affymetrix® Genome-Wide Human SNP Nsp/Sty 6.0 User Guide.

Affymetrix. (2007b). Data Sheet: Affymetrix® Genome-Wide Human SNP Array 6.0.

Altshuler, D., Daly, M.J., and Lander, E.S. (2008). Genetic Mapping in Human Disease. *Science* **322(5903):** 881-888.

ARIC Investigators. (1989). The Atherosclerosis Risk in Communities (ARIC) Study: Design and Objectives. *Am J Epidemiol* **129(4):** 687-702.

Atwood, L.D., Wolf, P.A., Heard-Costa, N.L., Massaro, J.M., Beiser, A., D'Agostino, R.B., and DeCarli, C. (2004). Genetic Variation in White Matter Hyperintensity Volume in the Framingham Study. *Stroke* **35(7):** 1609-1613.

Aulchenko, Y.S., Ripke, S., Isaacs, A., and van Duijn, C.M. (2007). GenABEL: An R Library for Genome-Wide Association Analysis. *Bioinformatics* **23(10):** 1294-1296.

Bandaru, V., Sunkara, S., Wallace, S.S., and Bond, J.P. (2002). A Novel Human DNA Glycosylase that Removes Oxidative DNA Damage and is Homologous to Escherichia Coli Endonuclease VIII. *DNA Repair (Amst)* **1(7):** 517-529.

Bornstein, N., Silvestrelli, G., Caso, V., and Parnetti, L. (2006). Arterial Hypertension and Stroke Prevention: An Update. *Clin Exp Hypertens* **28(3-4):** 317-326.

Breteler, M.M. (2000). Vascular Risk Factors for Alzheimer's Disease: An Epidemiologic Perspective. *Neurobiol Aging* **21(2):** 153-160.

Bryan, R.N., Manolio, T.A., Schertz, L.D., Jungreis, C., Poirier, V.C., Elster, A.D., and Kronmal, R.A. (1994). A Method for using MR to Evaluate the Effects of Cardiovascular Disease on the Brain: The Cardiovascular Health Study. *Am J Neuroradiol* **15(9):** 1625-1633.

Carmelli, D., DeCarli, C., Swan, G.E., Jack, L.M., Reed, T., Wolf, P.A., and Miller, B.L. (1998). Evidence for Genetic Variance in White Matter Hyperintensity Volume in Normal Elderly Male Twins. *Stroke* **29(6):** 1177-1181.

Chapuis, J., Hot, D., Hansmannel, F., Kerdraon, O., Ferreira, S., Hubans, C., Maurage, C.A., Huot, L., Bensemain, F., Laumet, G. et al. (2009). Transcriptomic and Genetic Studies Identify IL-33 as a Candidate Gene for Alzheimer's Disease. *Mol Psychiatry* **14(11):** 1004-1016.

Chen, W.M., and Abecasis, G.R. (2007). Family-Based Association Tests for Genomewide Association Scans. *Am J Hum Genet* **81(5):** 913-926.

Cheverud, J.M. (2000). *Chapter 4.* In *Epistasis and the Evolutionary Process.* Oxford University Press: New York, NY, pp. 58-81.

Christopherson, K.S., Ullian, E.M., Stokes, C.C., Mullowney, C.E., Hell, J.W., Agah, A., Lawler, J., Mosher, D.F., Bornstein, P., and Barres, B.A. (2005). Thrombospondins are Astrocyte-Secreted Proteins that Promote CNS Synaptogenesis. *Cell* **120(3):** 421-433.

Conrad, C., Boyman, O., Tonel, G., Tun-Kyi, A., Laggner, U., de Fougerolles, A., Kotelianski, V., Gardner, H., and Nestle, F.O. (2007). Alpha1beta1 Integrin is Crucial for Accumulation of Epidermal T Cells and the Development of Psoriasis. *Nat Med* **13(7):** 836-842.

Cutler, D.J., Zwick, M.E., Carrasquillo, M.M., Yohn, C.T., Tobin, K.P., Kashuk, C., Mathews, D.J., Shah, N.A., Eichler, E.E., Warrington, J.A. et al. (2001). High-Throughput Variation Detection and Genotyping using Microarrays. *Genome Res* **11(11):** 1913-1925.

Daniels, P.R., Kardia, S.L., Hanis, C.L., Brown, C.A., Hutchinson, R., Boerwinkle, E., Turner, S.T., and Genetic Epidemiology Network of Arteriopathy study. (2004). Familial Aggregation of Hypertension Treatment and Control in the Genetic Epidemiology Network of Arteriopathy (GENOA) Study. *Am J Med* **116(10):** 676-681.

de Bakker, P.I., Ferreira, M.A., Jia, X., Neale, B.M., Raychaudhuri, S., and Voight, B.F. (2008). Practical Aspects of Imputation-Driven Meta-Analysis of Genome-Wide Association Studies. *Hum Mol Genet* **17(R2):** R122-8.

Devlin, B., and Roeder, K. (1999). Genomic Control for Association Studies. *Biometrics* **55(4):** 997-1004.

Dickstein, D.L., Walsh, J., Brautigam, H., Stockton, S.D.,Jr, Gandy, S., and Hof, P.R. (2010). Role of Vascular Risk Factors and Vascular Dysfunction in Alzheimer's Disease. *Mt Sinai J Med* **77(1):** 82-102.

Dulabon, L., Olson, E.C., Taglienti, M.G., Eisenhuth, S., McGrath, B., Walsh, C.A., Kreidberg, J.A., and Anton, E.S. (2000). Reelin Binds alpha3beta1 Integrin and Inhibits Neuronal Migration. *Neuron* **27(1):** 33-44.

Elias, M.F., Wolf, P.A., D'Agostino, R.B., Cobb, J., and White, L.R. (1993). Untreated Blood Pressure Level is Inversely Related to Cognitive Functioning: The Framingham Study. *Am J Epidemiol* **138(6):** 353-364.

Ertekin-Taner, N., Ronald, J., Asahara, H., Younkin, L., Hella, M., Jain, S., Gnida, E., Younkin, S., Fadale, D., Ohyagi, Y. et al. (2003). Fine Mapping of the Alpha-T Catenin Gene to a Quantitative Trait Locus on Chromosome 10 in Late-Onset Alzheimer's Disease Pedigrees. *Hum Mol Genet* **12(23):** 3133-3143.

FBPP Investigators. (2002). Multi-Center Genetic Study of Hypertension: The Family Blood Pressure Program (FBPP). *Hypertension* **39(1):** 3-9.

Flint, J., and Mackay, T.F. (2009). Genetic Architecture of Quantitative Traits in Mice, Flies, and Humans. *Genome Res* **19(5):** 723-733.

Freedman, M.L., Reich, D., Penney, K.L., McDonald, G.J., Mignault, A.A., Patterson, N., Gabriel, S.B., Topol, E.J., Smoller, J.W., Pato, C.N. et al. (2004). Assessing the Impact of Population Stratification on Genetic Association Studies. *Nat Genet* **36(4):** 388-393.

Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M. et al. (2002). The Structure of Haplotype Blocks in the Human Genome. *Science* **296(5576):** 2225-2229.

Garcia, E.A., Newhouse, S., Caulfield, M.J., and Munroe, P.B. (2003). Genes and Hypertension. *Curr Pharm Des* **9(21):** 1679-1689.

Garmy-Susini, B., Jin, H., Zhu, Y., Sung, R.J., Hwang, R., and Varner, J. (2005). Integrin alpha4beta1-VCAM-1-Mediated Adhesion between Endothelial and Mural Cells is Required for Blood Vessel Maturation. *J Clin Invest* **115(6):** 1542-1551.

Hirschhorn, J.N., and Daly, M.J. (2005). Genome-Wide Association Studies for Common Diseases and Complex Traits. *Nat Rev Genet* **6(2):** 95-108.

Hoggart, C.J., Parra, E.J., Shriver, M.D., Bonilla, C., Kittles, R.A., Clayton, D.G., and McKeigue, P.M. (2003). Control of Confounding of Genetic Associations in Stratified Populations. *Am J Hum Genet* **72(6):** 1492-1504.

Jack, C.R.,Jr, O'Brien, P.C., Rettman, D.W., Shiung, M.M., Xu, Y., Muthupillai, R., Manduca, A., Avula, R., and Erickson, B.J. (2001). FLAIR Histogram Segmentation for Measurement of Leukoaraiosis Volume. *J Magn Reson Imaging* **14(6):** 668-676.

Jack, C.R.,Jr, Twomey, C.K., Zinsmeister, A.R., Sharbrough, F.W., Petersen, R.C., and Cascino, G.D. (1989). Anterior Temporal Lobes and Hippocampal Formations: Normative Volumetric Measurements from MR Images in Young Adults. *Radiology* **172(2):** 549-554.

Jackson, F.L. (2006). Illuminating Cancer Health Disparities using Ethnogenetic Layering (EL) and Phenotype Segregation Network Analysis (PSNA). *J Cancer Educ* **21(1 Suppl):** S69-79.

Knopman, D., Boland, L.L., Mosley, T., Howard, G., Liao, D., Szklo, M., McGovern, P., Folsom, A.R., and Atherosclerosis Risk in Communities (ARIC) Study Investigators. (2001). Cardiovascular Risk Factors and Cognitive Decline in Middle-Aged Adults. *Neurology* **56(1):** 42-48.

Kojro, E., Gimpl, G., Lammich, S., Marz, W., and Fahrenholz, F. (2001). Low Cholesterol Stimulates the Nonamyloidogenic Pathway by its Effect on the Alpha - Secretase ADAM 10. *Proc Natl Acad Sci U S A* **98(10):** 5815-5820.

Kruglyak, L. (1999). Prospects for Whole-Genome Linkage Disequilibrium Mapping of Common Disease Genes. *Nat Genet* **22(2):** 139-144.

Kukull, W.A., and Bowen, J.D. (2002). Dementia Epidemiology. *Med Clin North Am* **86(3):** 573-590.

Kuller, L.H., Lopez, O.L., Jagust, W.J., Becker, J.T., DeKosky, S.T., Lyketsos, C., Kawas, C., Breitner, J.C., Fitzpatrick, A., and Dulberg, C. (2005). Determinants of Vascular Dementia in the Cardiovascular Health Cognition Study. *Neurology* **64(9):** 1548-1552.

Kuller, L.H., Longstreth, W.T.,Jr, Arnold, A.M., Bernick, C., Bryan, R.N., Beauchamp, N.J.,Jr, and Cardiovascular Health Study Collaborative Research Group. (2004). White Matter Hyperintensity on Cranial Magnetic Resonance Imaging: A Predictor of Stroke. *Stroke* **35(8):** 1821-1825.

Lango, H., UK Type 2 Diabetes Genetics Consortium, Palmer, C.N., Morris, A.D., Zeggini, E., Hattersley, A.T., McCarthy, M.I., Frayling, T.M., and Weedon, M.N. (2008). Assessing the Combined Impact of 18 Common Genetic Variants of Modest Effect Sizes on Type 2 Diabetes Risk. *Diabetes* **57(11):** 3129-3135.

Launer, L.J., Ross, G.W., Petrovitch, H., Masaki, K., Foley, D., White, L.R., and Havlik, R.J. (2000). Midlife Blood Pressure and Dementia: The Honolulu-Asia Aging Study. *Neurobiol Aging* **21(1):** 49-55.

Manolio, T.A. (2010). Genomewide Association Studies and Assessment of the Risk of Disease. *N Engl J Med* **363(2):** 166-176.

Markus, H.S. (2008). Genes, Endothelial Function and Cerebral Small Vessel Disease in Man. *Exp Physiol* **93(1):** 121-127.

Markus, H.S., Hunt, B., Palmer, K., Enzinger, C., Schmidt, H., and Schmidt, R. (2005). Markers of Endothelial and Hemostatic Activation and Progression of Cerebral White Matter Hyperintensities: Longitudinal Results of the Austrian Stroke Prevention Study. *Stroke* **36(7):** 1410-1414.

McCarthy, M.I., Abecasis, G.R., Cardon, L.R., Goldstein, D.B., Little, J., Ioannidis, J.P., and Hirschhorn, J.N. (2008). Genome-Wide Association Studies for Complex Traits: Consensus, Uncertainty and Challenges. *Nat Rev Genet* **9(5):** 356-369.

Micu, I., Jiang, Q., Coderre, E., Ridsdale, A., Zhang, L., Woulfe, J., Yin, X., Trapp, B.D., McRory, J.E., Rehak, R. et al. (2006). NMDA Receptors Mediate Calcium Accumulation in Myelin during Chemical Ischaemia. *Nature* **439(7079):** 988-992.

Miyashita, A., Arai, H., Asada, T., Imagawa, M., Matsubara, E., Shoji, M., Higuchi, S., Urakami, K., Kakita, A., Takahashi, H. et al. (2007). Genetic Association of CTNNA3 with Late-Onset Alzheimer's Disease in Females. *Hum Mol Genet* **16(23):** 2854-2869.

Mosley, T.H.,Jr, Knopman, D.S., Catellier, D.J., Bryan, N., Hutchinson, R.G., Grothues, C.A., Folsom, A.R., Cooper, L.S., Burke, G.L., Liao, D. et al. (2005). Cerebral MRI Findings and Cognitive Functioning: The Atherosclerosis Risk in Communities Study. *Neurology* **64(12):** 2056-2062.

Nakaoka, H., and Inoue, I. (2009). Meta-Analysis of Genetic Association Studies: Methodologies, between-Study Heterogeneity and Winner's Curse. *J Hum Genet* **54(11):** 615-623.

Need, A.C., Attix, D.K., McEvoy, J.M., Cirulli, E.T., Linney, K.L., Hunt, P., Ge, D., Heinzen, E.L., Maia, J.M., Shianna, K.V. et al. (2009). A Genome-Wide Study of Common SNPs and CNVs in Cognitive Performance in the CANTAB. *Hum Mol Genet* **18(23):** 4650-4661.

O'Sullivan, M. (2008). Leukoaraiosis. *Pract Neurol* **8(1):** 26-38.

Pantoni, L., Poggesi, A., and Inzitari, D. (2007). The Relation between White-Matter Lesions and Cognition. *Curr Opin Neurol* **20(4):** 390-397.

Paternoster, L., Chen, W., and Sudlow, C.L. (2009). Genetic Determinants of White Matter Hyperintensities on Brain Scans: A Systematic Assessment of 19 Candidate Gene Polymorphisms in 46 Studies in 19,000 Subjects. *Stroke* **40(6):** 2020-2026.

Petretto, E., Liu, E.T., and Aitman, T.J. (2007). A Gene Harvest Revealing the Archeology and Complexity of Human Disease. *Nat Genet* **39(11):** 1299-1301.

Postina, R., Schroeder, A., Dewachter, I., Bohl, J., Schmitt, U., Kojro, E., Prinzen, C., Endres, K., Hiemke, C., Blessing, M. et al. (2004). A Disintegrin-Metalloproteinase Prevents Amyloid Plaque Formation and Hippocampal Defects in an Alzheimer Disease Mouse Model. *J Clin Invest* **113(10):** 1456-1464.

Prins, N.D., van Dijk, E.J., den Heijer, T., Vermeer, S.E., Koudstaal, P.J., Oudkerk, M., Hofman, A., and Breteler, M.M. (2004). Cerebral White Matter Lesions and the Risk of Dementia. *Arch Neurol* **61(10):** 1531-1534.

R Core Development Team. (2008). R: A Language and Environment for Statistical Computing.

Rabbee, N., and Speed, T.P. (2006). A Genotype Calling Algorithm for Affymetrix SNP Arrays. *Bioinformatics* **22(1):** 7-12.

Raudenbush, S.W., and Bryk, A.S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods, 2nd Ed.* Sage Publications, Inc: Thousand Oaks, CA.

Roberts, R., Wells, G.A., Stewart, A.F., Dandona, S., and Chen, L. (2010). The Genome-Wide Association Study--a New Era for Common Polygenic Disorders. *J Cardiovasc Transl Res* **3(3):** 173-182.

Roger, V.L., Go, A.S., Lloyd-Jones, D.M., Adams, R.J., Berry, J.D., Brown, T.M., Carnethon, M.R., Dai, S., de Simone, G., Ford, E.S., et al. (2011). Heart Disease and Stroke Statistics 2011 Update: A Report From the American Heart Association. *Circulation* **123**: e000-e000 (epub ahead of print).

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(5):** 356-366.

Schmidt, R., Petrovic, K., Ropele, S., Enzinger, C., and Fazekas, F. (2007). Progression of Leukoaraiosis and Cognition. *Stroke* **38(9):** 2619-2625.

Schmidt, R., Scheltens, P., Erkinjuntti, T., Pantoni, L., Markus, H.S., Wallin, A., Barkhof, F., and Fazekas, F. (2004). White Matter Lesion Progression: A Surrogate Endpoint for Trials in Cerebral Small-Vessel Disease. *Neurology* **63(1):** 139-144.

Seshadri, S., DeStefano, A.L., Au, R., Massaro, J.M., Beiser, A.S., Kelly-Hayes, M., Kase, C.S., D'Agostino RB, S., Decarli, C., Atwood, L.D. et al. (2007). Genetic Correlates of Brain Aging on MRI and Cognitive Test Measures: A Genome-Wide Association and Linkage Analysis in the Framingham Study. *BMC Med Genet* **8 (Suppl 1):** S15.

Sillen, A., Brohede, J., Lilius, L., Forsell, C., Andrade, J., Odeberg, J., Ebise, H., Winblad, B., and Graff, C. (2010). Linkage to 20p13 Including the ANGPT4 Gene in Families with Mixed Alzheimer's Disease and Vascular Dementia. *J Hum Genet* **55(10):** 649-655.

Staniszewska, I., Zaveri, S., Del Valle, L., Oliva, I., Rothman, V.L., Croul, S.E., Roberts, D.D., Mosher, D.F., Tuszynski, G.P., and Marcinkiewicz, C. (2007). Interaction of alpha9beta1 Integrin with Thrombospondin-1 Promotes Angiogenesis. *Circ Res* **100(9):** 1308-1316.

Suzuki, K., Okuno, T., Yamamoto, M., Pasterkamp, R.J., Takegahara, N., Takamatsu, H., Kitao, T., Takagi, J., Rennert, P.D., Kolodkin, A.L. et al. (2007). Semaphorin 7A Initiates T-Cell-Mediated Inflammatory Responses through alpha1beta1 Integrin. *Nature* **446(7136):** 680-684.

Suzuki, S.C., and Takeichi, M. (2008). Cadherins in Neuronal Morphogenesis and Function. *Dev Growth Differ* **50 (Suppl 1):** S119-30.

Swan, G.E., DeCarli, C., Miller, B.L., Reed, T., Wolf, P.A., Jack, L.M., and Carmelli, D. (1998). Association of Midlife Blood Pressure to Late-Life Cognitive Decline and Brain Morphology. *Neurology* **51(4):** 986-993.

Szolnoki, Z., Somogyvari, F., Kondacs, A., Szabo, M., Fodor, L., Bene, J., and Melegh, B. (2004). Specific APO E Genotypes in Combination with the ACE D/D Or MTHFR 677TT Mutation Yield an Independent Genetic Risk of Leukoaraiosis. *Acta Neurol Scand* **109(3):** 222-227.

Thomas, M., and Augustin, H.G. (2009). The Role of the Angiopoietins in Vascular Morphogenesis. *Angiogenesis* **12(2):** 125-137.

Thornton-Wells, T.A., Moore, J.H., and Haines, J.L. (2004). Genetics, Statistics and Human Disease: Analytical Retooling for Complexity. *Trends Genet* **20(12):** 640-647.

Turner, S.T., Fornage, M., Jack, C.R.,Jr, Mosley, T.H., Knopman, D.S., Kardia, S.L., Boerwinkle, E., and de Andrade, M. (2009). Genomic Susceptibility Loci for Brain Atrophy, Ventricular Volume, and Leukoaraiosis in Hypertensive Sibships. *Arch Neurol* **66(7):** 847-857.

Turner, S.T., Jack, C.R., Fornage, M., Mosley, T.H., Boerwinkle, E., and de Andrade, M. (2004). Heritability of Leukoaraiosis in Hypertensive Sibships. *Hypertension* **43(2):** 483-487.

Vazquez, F., Hastings, G., Ortega, M.A., Lane, T.F., Oikemus, S., Lombardo, M., and Iruela-Arispe, M.L. (1999). METH-1, a Human Ortholog of ADAMTS-1, and METH-2 are Members of a New Family of Proteins with Angio-Inhibitory Activity. *J Biol Chem* **274(33):** 23349-23357.

Visscher, P.M., Andrew, T., and Nyholt, D.R. (2008). Genome-Wide Association Studies of Quantitative Traits with Related Individuals: Little (Power) Lost but Much to be Gained. *Eur J Hum Genet* **16(3):** 387-390.

Volpert, O.V., Zaichuk, T., Zhou, W., Reiher, F., Ferguson, T.A., Stuart, P.M., Amin, M., and Bouck, N.P. (2002). Inducer-Stimulated Fas Targets Activated Endothelium for Destruction by Anti-Angiogenic Thrombospondin-1 and Pigment Epithelium-Derived Factor. *Nat Med* **8(4):** 349-357.

Weir, B.S. (1996). *Genetic Data Analysis II: Methods for Discrete Population Genetic Data.* Sinauer Associates: Sunderland, MA.

Wellcome Trust Case Control Consortium. (2007). Genome-Wide Association Study of 14,000 Cases of Seven Common Diseases and 3,000 Shared Controls. *Nature* **447(7145):** 661-678.

Williams, S.E., Wootton, P., Mason, H.S., Bould, J., Iles, D.E., Riccardi, D., Peers, C., and Kemp, P.J. (2004). Hemoxygenase-2 is an Oxygen Sensor for a Calcium-Sensitive Potassium Channel. *Science* **306(5704):** 2093-2097.

Zeggini, E., and Ioannidis, J.P. (2009). Meta-Analysis in Genome-Wide Association Studies. *Pharmacogenomics* **10(2):** 191-201.

Chapter 4

Pleiotropy among Leukoaraiosis and Measures of Cognitive Function

Abstract

Ischemic brain injury due to inadequately controlled hypertension is associated with stroke and dementia, clinical endpoints that aggregate in families and are likely to be the consequence of complex interactions between many genetic and environmental factors. Leukoaraiosis, a subclinical measure of hypertension-related brain injury, is a powerful predictor of stroke and dementia as well as cognitive decline prior to clinical endpoints. We investigated the extent of pleiotropy, the condition in which variation in a single gene affects multiple traits, between leukoaraiosis and seven measures of cognitive function using both unmeasured genetic (biometrical) and measured genetic (GWAS) approaches in whites and African Americans of the Genetic Epidemiology Network of Arteriopathy (GENOA) study. Heritability analysis showed strong evidence of genetic contribution to both leukoaraiosis and the investigated cognitive traits, ranging from 0.3 to 0.78 in whites and from 0.14 to 0.81 in African Americans. Similar patterns of genetic correlations (evidence of pleiotropy) were observed for pairs of cognitive traits in whites (significant correlations ranged from 0.26 to 0.92) and African Americans (0.36 to 0.92), though genetic correlations reached the level of significance more frequently in whites (67% of cognitive pairs in whites and 19% in African Americans). Only one cognitive trait, a

measure of total learning, showed evidence of pleiotropy with leukoaraiosis. The percentage of nominally significant SNP associations from GWAS that were shared between pairs of traits showed a stronger relationship with genetic correlation than environmental correlation, indicating that percentage of shared SNP associations may be a useful metric for assessing pleiotropy from GWAS results. A greater understanding of the genetic architecture and the underlying pleiotropic mechanisms contributing to leukoaraiosis and cognitive function has the potential to allow for earlier identification of individuals at increased risk for disease, the development of more efficacious treatments, and the tailoring of particular treatments to people most likely to respond positively.

Introduction

Public health importance of studying leukoaraiosis and measures of cognitive function

Hypertension affects approximately 1 in 3 American adults (76.4 million people), and accounts for $43.5 billion in yearly direct and indirect costs in the United States (Roger, 2011). Hypertension is a leading risk factor for ischemic stroke (Roger, 2011; Kannel, 1995) and for cognitive decline leading to vascular dementia (Elias, 1993; Launer, 2000). Inadequately controlled hypertension gives rise to ischemic damage of the brain that is thought to be the manifestation of underlying cerebrovascular disease (Turner, 2000). Areas of ischemic damage, known as leukoaraiosis, appear as hyperintense spots on MRI of the white matter of the brain, and the extent of subclinical damage can be quanitifed as the volume of leukoaraiosis. Development of leukoaraiosis is thought to be one of the major mechanistic pathways between hypertension and clinical endpoints such as

ischemic stroke and vascular dementia, and is a known risk factor for both of these endpoints (O'Sullivan, 2008; Markus, 2005; Kuller, 2005; Fu, 2005).

Ischemic stroke accounts for 87% of all strokes, a leading cause of morbidity, mortality, and economic burden in the US (Roger, 2011). Over 7 million Americans currently living with the cognitive and physical consequences of stroke (Roger, 2011), and it has been estimated that stroke account for approximately 4% of all direct health care costs in the US (Donnan, 2008). Dementia is also a leading cause of morbidity and economic burden in the US (Kukull, 2002; Haan, 2004), affecting 3%-11% of people older than 65 and 25%-47% of people older than 85 (Boustani, 2003). Dementia is a heterogeneous group of disorders with variable etiology that involves impairment in cognitive domains such as memory, executive function, and language as well as specific physical impairments such as gait abnormalities that cause significant impairment in social or occupational function and represent a decline from a previous level of functioning (American Psychiatric Association, 2000). The differential diagnosis of vascular dementia (VaD), incorporates the underlying vascular cause as well as the cognitive and physical symptomology (Pohjasvaara, 2000), specifically "focal neurological signs and symptoms or laboratory evidence indicative of cerebrovascular disease (multiple infarctions involving cortex and underlying white matter) that are judged to be etiologically related to the disturbance) (American Psychiatric Association, 2000). Therefore, leukoaraiosis is not only a risk factor for VaD, but also is part of the diagnostic criteria.

Role of genetics in leukoaraiosis and cognitive function

Though the main risk factors for leukoaraiosis are elevated blood pressure and lack of blood pressure control (van Dijk, 2004; Liao, 1996), there is a significant amount of inter-individual variation in leukoaraiosis volume among subjects with similar duration and severity of hypertension (Szolnoki, 2006; Schmidt, 2004). In addition inter-individual variability in cognitive functioning and brain structure is highly variable and it is likely that genetic variability accounts for a significant portion of the variation (Deary, 2004). Heritability studies, candidate gene studies, and genome-wide association studies are beginning to shed light on the biological processes involved in the progression from hypertension to the development of leukoaraiosis and cognitive decline that are asymptomatic indicators of increased risk of stroke and dementia.

A comprehensive review of the linkage, candidate gene, and genome-wide association studies for leukoaraiosis and cognitive traits are presented in Chapter 1 of this dissertation. Briefly, candidate gene studies for leukoaraiosis have primarily concentrated on genes in pathways known to be involved in hypertension, vasculature, and endothelial damage, and although initial findings have been encouraging, no specific polymorphism has been unequivocally shown to be associated with this trait (Paternoster, 2009). For cognitive decline, the most promising candidate genes include those that are associated with hypertension, leukoaraiosis, Alzheimer's Disease (AD), normal cognitive functioning, cardiovascular function, oxidative stress, and inflammation (Deary, 2004), though there haven't been enough candidate gene studies conducted on these traits in humans to isolate any specific polymorphisms that affect cognitive decline at this time.

Genome-wide association studies (GWAS) for these traits have also been limited. Seshadri et al. (2007) identified several regions associated with leukoaraiosis in a sample of 705 related white participants, and a large meta-analysis of 9,401 white participants identified a an intronic SNP in the *MACRO domain containing 2* (*MACROD2*) gene in the downstream region of the *fibronectin leucine-rich transmembrane protein 3* (*FLTR3*) gene (Debette, 2010). Though the SNP did not replicate in an independent white sample of 1,822 participants, four SNPs within 200kb from the original SNP did show association with the trait in a sample of 644 black participants.

Pleiotropy

Pleiotropy is most simply defined as the condition in which variation in a single gene affects multiple traits (Hodgkin, 1998). Defined in this manner, pleiotropic genes range from those that encode proteins involved in a single biological pathway that influences multiple disease processes and/or organ systems to those that play entirely different roles in multiple biological pathways. In some instances, pleiotropy is "the phenomenon in which a single gene controls several distinct, seemingly unrelated, phenotypic effects" (Zou, 2008). In humans, an example is recessive mutation in the structural locus of the tyrosine kinase (*TYR*) gene, which obstructs the biosynthesis of melanin. This leads to hypopigmentation of the skin (albinism) as well as ocular disturbances such as severely reduced visual acuity. Though the traits affected are seemingly unrelated, both are the result of disruption of the underlying biological pathway involving conversion of tyrosine to melanin (Carden, 1998).

Another well-known example of pleiotropy in humans is the ε4 allele of the apolioprotein E (*APOE*) gene (Dickstein, 2010). *APOE* is a plasma cholesterol transport molecule that resides primarily on very low density lipoproteins, and the ε4 allele has been shown to be a risk factor for coronary heart disease and stroke through mechanisms related directly to lipid transport. However, the ε4 alelle is also a risk factor for AD and cognitive decline, with those carrying the allele having a younger age of onset as well as an accelerated pace of cognitive decline. Though the precise mechanism by which ε4 leads to cognitive decline is not known, the main hypotheses are through pathways not directly related or only tangentially related to lipid transport. *APOE* appears to affect brain traits through its role as a chaperone for the amyloid beta protein and/or mediation of the phosphorylation of the tau protein.

GWAS studies have also revealed some surprising pleiotropic findings in humans. Several genic regions have been implicated in cancers in addition to at least one other seemingly unrelated human disease, as have individual single nucleotide polymorphisms (SNPs). For example, at least three SNPs associated with type 2 diabetes have also been found to be associated with both prostate and colon cancers (Winckler, 2007; Gudmundsson, 2007; Zeggini, 2008; Thomas, 2008; Slattery, 2008).

Pleiotropic genetic variation has been studied extensively in model organisms using quantitiative genetic methods, primarily linkage, in order to inform breeding programs, examine evolutionary mechanisms, and elucidate molecular pathways. Studies of genetic variation of quantitative traits in model organisms have revealed unexpectedly complex

genetic architectures with a large degree of pleiotropy (Mackay, 2010; Zhou, 2009). Mackay and colleagues (2005) studied the genetics of Drosophila mating behavior and found that nearly 20% of the genome was associated with this trait, including genes that have also been implicated in a variety of seemingly unrelated or tangentially-related biological pathways such as neurogenesis, learning and memory, olfaction, metabolism, and development as well as other behavioral traits such as circadian rhythm and geotaxis. Zou and colleagues (2008) found that more than half of the genes involved in early embryogenesis of C. elegans show a marginal degree of pleiotropy and that approximately 3% of genes were highly pleiotropic, noting that signaling proteins exhibit a very high degree of pleiotropy. A further finding in studies of pleiotropy in model organisms is that pleiotropic effects are often sex- and environment-specific (Mackay, 2010; Zhou, 2009; Mackay, 2009).

The study of pleiotropy in model organisms and humans serves several functions. In model organisms, it serves to further the understanding and elucidation of the complex biological pathways that regulate the development of traits, providing information about normal cellular function, normal development and function at the organismal level, connections between previously unrecognized biological processes, and increased predictive ability in breeding programs (Hodgkin, 1998). It also it provides insight into the mechanisms of evolution, as effects on multiple traits due to a single genetic variant may pose severe evolutionary constraints (Cheverud, 2004). The findings from pleiotropy studies in model organisms, particularly the high degree of connectivity among transcriptional modules, have strong implications for understanding the pleiotropic

genetic mechanisms that are also operating in humans. A greater understanding of the underlying pleiotropic mechanisms contributing to human health and disease has the potential to allow for earlier identification of individuals at increased risk for disease, the development of more efficacious treatments, and the tailoring of particular treatments to people most likely to respond positively.

Bivariate variance component analysis to assess pleiotropy

Preliminary studies of pleiotropy in humans have generally consisted of bivariate genetic analysis using variance decomposition techniques and linkage analysis in biologically related groups of traits. Variance decomposition techniques are used to parse the total phenotypic correlation in a pair of traits into the correlation due to genetic influences (genetic correlation) and the correlation due to environmental influences (environmental correlation) using family relationships.

Bivariate variance decomposition techniques have been used to study pleiotropy in humans for a variety of purposes. Comuzzie and colleagues (1994) estimated the genetic and environmental correlations among eight measures of skinfolds in order to inform epidemiologic studies that study these measures as risk factors for heart disease and diabetes. The authors argue that studying pleiotropy in risk factors is important because shared genetic or environmental effects may confound analyses using these traits if these effects are unrecognized. Using variance decomposition and linkage analysis, Voruganti et al. (2009) investigated the genetic and environmental correlations between uric acid and several cardiovascular risk factors with the purpose similar to Comuzzie et al. of

informing epidemiologic studies. The studies were both conducted in Mexican Americans, a population that is at increased risk for obesity and diabetes, both significant risk factors for cardiovascular disease (Comuzzie, 1994; Voruganti, 2009).

Cassidy-Bushrow et al. (2007) examined pleiotropy using bivariate genetic analysis for a different purpose. They examined the genetic contribution to coronary artery calcification (CAC) by quantifying the genetic effects that are shared between a measure of CAC at a single point in time (baseline) and CAC progression over time to determine the extent to which a common set of genes were involved in the different phases of the atherosclerotic process. A further use of bivariate genetic analysis is to identify measureable endophenotypes that can be used to study the genetic underpinnings of complex diseases with multiple etiologies. For example, Charlesworth et al. (2010) examined the genetic correlations between several quantitative characteristics of the eye and primary open-angle glaucoma in order to determine the most appropriate endophenotypes to focus on in genetic association studies.

GWAS studies as a tool for identifying pleiotropic genes

While there has been a large amount of work that examines the genetic contribution to traits and genetic correlations among traits using variance components analysis as well as great advances and interest in genetic association studies to identify loci that contribute to trait variation, there has historically been little integration or agreement between these two approaches. As it becomes more computationally and economically feasible to measure and impute millions of genetic markers on groups of individuals, there is

considerable interest in developing techniques for studying pleiotropy with measured genetic variations such as SNPs. Comparing the similarities and differences in the properties of these two techniques is important in order to determine the utility that GWAS results may have in identifying pleiotropic genes.

Taking a SNP-based approach to exploring pleiotropy is important for several reasons. First, traditional bivariate variance component approaches can only be used with family structured data, while the GWAS approach is typically performed on unrelated individuals. Given the drastically increasing utilization of the GWAS approach by the statistical genetics community as well as the growing abundance of samples with GWAS data and multiple traits, learning how to use GWAS data to examine pleiotropy will be greatly beneficial to examining the genetic architecture of complex traits. Second, variance component approaches do not give insight into which genes, specifically, have pleiotropic effects on the outcome of interest. These approaches can only give an estimate of the extent of pleiotropy, while SNP-based approaches can be used to identify the underlying genetic variations that are pleiotropic because they make use of directly genotyped information instead of only family relationships. Developing methods to identify genes with pleiotropic effects will greatly enhance our ability to map the genetic architecture of complex traits by shedding light on the inter-relationships of the underlying biological pathways that contribute to the development of the traits. It will also help us to identify the specific genes that have a broad impact on a particular system.

To date, there has been little research into the use of GWAS data to explore pleiotropy in humans. In a recently published paper, Karasik et al. (2010) began to explore the relationship between variance components analysis and GWAS results on pairs of bone-related traits. In this study, the authors compared the genetic, environmental, and overall phenotypic correlation among pairs of traits with the percentage of marginally significant (p-value < 0.01) GWAS results that overlap between the traits ("shared associated SNPs"). They also used a simulation method to determine whether the percentage of shared associated SNPs in pairs of traits were more significant than what is expected by chance alone. Using this approach, the authors concluded that the percentage of shared associated SNPs is strongly associated with the genetic correlation between traits and not environmental correlation, and thus may be an acceptable metric for measuring pleiotropy.

In this chapter, we focus on estimating the heritabilities and genetic correlations between leukoaraiosis and seven measures of neurocognitive function in a sample of 759 whites and 720 African Americans to examine patterns of pleiotropy using a bivariate variance components analysis approach and a GWAS approach. Evidence of pleiotropy detected by the two approaches will be compared in order to examine the utility of using GWAS results as a metric for assessing pleiotropy in biologically related traits. Findings of this work will help inform method development for studying pleiotropy in humans as well as further the understanding of the genetic relationships among leukoaraiosis and measures of cognitive function. A deeper understanding the genetics of subclinical leukoaraiosis development and its impact on cognitive decline in individuals free of overt

neurocognitive disorders may help to inform pharmacogenomic drug development and preventive strategies for identifying individuals at increased risk of stroke and dementia. Research into the genetic architecture of leukoaraiosis and cognitive decline in samples that are presymptomatic is particularly important because preventive interventions for dementia would need to start early, preferably before any brain damage occurs (DeKosky, 2003).

Methods

Sample

The National Heart, Lung and Blood Institute established the Family Blood Pressure Program (FBPP) in 1996 from four existing research networks that were investigating the genetics of hypertension and its sequelae (FBPP Investigators, 2002), including The Genetic Epidemiology Network of Arteriopathy (GENOA). GENOA recruited hypertensive sibships from Rochester, Minnesota and Jackson, Mississippi for linkage and association studies to investigate the genetic underpinnings of hypertension and target organ damage related to hypertension (Daniels, 2004).

In the initial phase of the GENOA study (Phase I: 1996-2001), all members of sibships containing ≥ 2 individuals with essential hypertension clinically diagnosed before age 60 were invited to participate, including both hypertensive and normotensive siblings (1,583 non-Hispanic whites and 1,841 African Americans). The diagnosis of essential hypertension was established based on blood pressure levels measured at the study visit (>140 mmHg average systolic BP or >90 mmHg average diastolic BP) or a prior

diagnosis of hypertension and current treatment with antihypertensive medications. Exclusion criteria were secondary hypertension, alcoholism or drug abuse, pregnancy, insulin-dependent diabetes mellitus, or active malignancy. In the second phase of the GENOA study (Phase II: 2000-2004), 1,241 white and 1,482 African American participants were successfully re-recruited to measure potential target organ damage due to hypertension. Phase I and II GENOA data consist of demographic information, medical history, clinical characteristics, lifestyle factors, and blood samples for genotyping and biomarker assays. Written informed consent was obtained from all subjects and approval was granted by participating institutional review boards. All reported phenotype and covariate data used for this dissertation was collected during the Phase II exam.

The Genetics of Microangiopathic Brain Injury (GMBI) study (2001-2006) is an ancillary study of GENOA undertaken to investigate susceptibility genes for ischemic brain injury. Phase II GENOA participants that had a sibling willing and eligible to participate in the GMBI study underwent a neurocognitive testing battery to assess several domains of cognitive function including learning, memory, attention, concentration, and language (967 whites and 1,010 African Americans). Ischemic brain damage to the subcortical and periventricular white matter (leukoaraiosis) was quantified by magnetic resonance imaging (MRI) in subjects who had no history of stroke or neurological disease and no implanted metal devices (916 whites and 830 African Americans).

Participants were excluded from this analysis if they were less than 45 years of age (56 whites, 23 African Americans), had missing data for age (2 African Americans), or had a history of stroke (22 whites, 51 African Americans). Since the GWAS analyses in this study were conducted in part to serve as a replication sample for the Cohorts in Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium, participants that were also part of a CHARGE cohort were excluded from the analysis. African American participants who also participated in the cognitive assessment performed as part of the Atherosclerosis Risk in Communities (ARIC) study, one of the cohorts in the CHARGE Consortium, were excluded from the GENOA GWAS (N=118). Excluding participants that did not have genome-wide genotype data left a final sample of 762 whites in 378 sibships and 720 African Americans in 413 sibships. Of the white participants, 58.1% were female, and the average age was 61.27 years. Of the African American participants, 72.6% were female, and the average age was 63.29 years.

Outcome measures

Leukoaraiosis

Leukoaraiosis volume (cm^3) was obtained via MRI in a separate clinical visit. All MRI scans were performed on identically equipped Signa 1.5 T MRI scanners (GE Medical Systems, Waukesha, WI, USA) and images were centrally processed at the Mayo Clinic. Symmetric head positioning with respect to orthogonal axes was verified by a series of short scout scans. Total intracranial volume (head size) was measured from T1-weighted spin echo sagittal images, each set consisting of 32 contiguous 5 mm thick slices with no interslice gap, field of view = 24 cm, matrix = 256 x 192, obtained with the following

sequence: scan time = 2.5 min, echo time = 14 ms, repetitions = 2, replication time = 500 ms (Jack, 1989). Total brain and leukoaraiosis volumes were determined from axial fluid-attenuated inversion recovery (FLAIR) images, each set consisting of 48 contiguous 3-mm interleaved slices with no interslice gap, field of view = 22 cm, matrix = 256 x 160, obtained with the following sequence: scan time = 9 min, echo time = 144.8 ms, inversion time = 2,600 ms, repetition time = 26,002 ms, bandwidth = +/- 15.6 kHz, one signal average. A FLAIR image is a T2-weighted image with the signal of the cerebrospinal fluid nulled, such that brain pathology appears as the brightest intracranial tissue. Interactive imaging processing steps were performed by a research associate who had no knowledge of the subjects' personal or medical histories or biological relationships. A fully automated algorithm was used to segment each slice of the edited multi-slice FLAIR sequence into voxels assigned to one of three categories: brain, cerebrospinal fluid, or leukoaraiosis. The mean absolute error of this method is 1.4% for brain volume and 6.6% for leukoaraiosis volume, and the mean test-retest coefficient of variation is 0.3% for brain volume and 1.4% for leukoaraiosis volume (Jack, 2001). White matter hyperintensities in the corona-radiata and periventricular zone, as well as central gray infarcts (ie, lacunes) were included in the global leukoaraiosis measurements. Brain scans with cortical infarctions were excluded from the analyses because of the distortion of the leukoaraiosis volume estimates that would be introduced in the automated segmentation algorithm.

Neuropsychological testing battery

Neuropsychological tests were conducted in a private room that was free of noise and other distractions by trained interviewers (Alves de Moraes, 2002). In order to assure accuracy and comparability in test administrator performance, a portion (approximately 5%) of all interviews were tape recorded and evaluated for accuracy to provide feedback to test administrators. The neuropsychological outcome measures used for this analysis are presented in Table 4.1 along with the cognitive functions assessed.

Rey's Auditory Verbal Learning Test (RAVLT)

Rey's Auditory Verbal Learning Test (RAVLT) is a brief test that assesses learning and memory through multiple learning trials and a 30-minute delayed recall (Rey, 1964; Spreen, 1998). Specifically, the measure assesses immediate memory span, new learning, vulnerability to interference in learning, and recognition memory. RAVLT testing norms for individuals aged 55 and older were developed through Mayo's Older Americans Normative Studies (MOANS), and the testing procedure followed in GMBI was identical to that used in MOANS (Ivnik, 1992).

The examiner begins by reading a list of 15 common words aloud, and participants are asked to recall as many of the words as possible in any order. The same procedure is repeated four more times using the same list of 15 words. The total number of words that the participant remembers correctly over the five trials is recorded and forms the basis of the RAVLT total learning outcome measure for this analysis, which assesses immediate memory. Delayed recall is assessed by asking the participant to again name as many

words as he/she remembers after a 30-minute delay, forming the basis of the RAVLT delayed learning outcome measure for this analysis. During the 30-minute interim, an interference task is performed in which the interviewer reads another set of words aloud, and the participant is asked to recall them. Thus, the delayed learning outcome assesses both delayed memory as well as vulnerability to interference.

Wechsler Adult Intelligence Scale Revised (WAIS-R) Digit Symbol Substitution Task (DSST)

The Digit Symbol Substitution task (DSS) from the Wechsler Adult Intelligence Scale Revised (WAIS-R) (Wechsler, 1981b; Wechsler, 1981a) is a timed translation test designed to measure complex visual attention, sustained and focused concentration, response speed, and visuomotor coordination (Lezak, 1995). In this test, participants are given a key in which each number corresponds to a special symbol. The task consists of filling in empty boxes below a series of random numbers with the symbol corresponding to the appropriate number (translating the numbers to symbols). After a practice session to ensure that the participant understands the task, participants were given a 90 second time limit to complete as many items as possible. The DSST outcome measure for this analysis was the number of correct symbols completed in 90 seconds.

Controlled Oral Word Association Test (COWA) of the Multilingual Aphasia Examination

The Multilingual Aphasia Examination was developed to diagnose the presence of aphasic disorders (any type of acquired language impairment), and the Controlled Oral

Word Association Test (COWA) is a subset of this examination designed to measure verbal fluency (Lezak, 1995; Benton, 1994). Two measures of verbal fluency were used as outcomes for the present study, one of letter fluency (Word Fluency Test (WFT)) and one of category fluency (Animal Naming).

The Word Fluency Test of the COWA assesses letter fluency (phonetic association) by asking subjects to generate words orally that begin with a specific letter of the alphabet ("F", "A", and "S") for a period of 60 seconds. These letters were chosen because they have been demonstrated to allow more vocabulary choices overall than other letters. Scoring of this test consisted of adding the total number of admissible words generated for each of the three letters. Inadmissable words include proper nouns as well as variations, plurals, and repetitions of previously stated words.

The Animal Naming portion of the COWA assesses category fluency (semantic association) by asking subjects to name as many animals as possible in a period of 60 seconds (Lezak, 1995). Scoring of this test is the sum of all admissible animals. Inadmissable animals include extinct, imaginary, or magical animals, proper names, and variations of previously stated animals.

Stroop Color Word (CW) Test

The Stroop Color Word (CW) Test is primarily a measure of concentration effectiveness, specifically the ability to shift perceptual sets to correspond with changing demands and the ability to inhibit a customary response to stimulus in favor of a more novel one

(Spreen, 1998; Lezak, 1995; Stroop, 1935). Administration and scoring of the test to GMBI participants followed procedures outlined in the standardized version of the CW test developed by Golden, 1978 (Golden, 1978).

This test consists of three pages: the word page, the color page, and the color-word page. The word page consists of the words "RED", "GREEN", and "BLUE" arranged randomly and printed in black ink. The color page consists of sets of "XXXX" printed in red, green, or blue ink. The color-word page consists of the words from the word page printed in the colors on the color page, but no word matches the color in which it is printed (for example, the word "RED" is printed in either green or blue ink). For this study, the participant was first asked to read the word page as fast as he/she could for 45 seconds, and the total number of correct words was recorded. If the participant stated an incorrect word, the interviewer said, "No," and the participant was instructed to read the same word again to correct their error. The same procedure was followed for naming the colors on the color page. The participant was then asked to state the colors of the words on the color-word page as fast as he/she could in 45 seconds, and the total number of correctly stated colors were recorded.

Two measures from this test are used as outcomes in the present study. The color-word (CW) score is the total number of correctly stated colors out of 100 from the color-word page. The Stroop difference score is the difference in scores between the color page and the color-word page.

Genotyping and SNP imputation

A total of 1387 white participants and 1263 African American participants from GENOA were genotyped on the Affymetrix® Genome-Wide Human SNP Array 6.0 array using the protocol outlined by Affymetrix (Affymetrix, 2007) at the Mayo Clinic in Rochester, Minnesota. Some of the stored blood samples for African Americans contained DNA of poor quality, and we were unable to genotype these samples using the Affymetrix 6.0 platform. However, we were able to obtain high quality genotyping using the Illumina® Human1M-Duo BeadChip (Illumina, 2010) for an additional 269 African Americans. Since the African American sibships for the GENOA study were identified using hypertensive subjects from the ARIC study as probands, we also obtained genotypes for 92 additional GENOA/ARIC participants that could not be genotyped on either platform using the GENOA blood sample. Genotyping for the ARIC study was done at the Broad Institute on the Affymetrix 6.0 platform.

For all genotyping platforms used, samples and SNPs with a call rate <95% were removed. Samples demonstrating sex mismatch, duplicate samples, and samples with low identity-by-state with all other samples were also removed. Pedigree information was used as a quality check to identify mislabeled samples. The sample used in this analysis consisted of 568 African Americans genotyped on the Affymetrix platform as part of the GENOA study, 118 genotyped on the Illumina platform, and 34 genotyped on the Affymetrix platform as part of the ARIC study.

SNP imputation increases the density of available genotypes, fills in missing genotypes, provides the ability to perform analysis on the full sample of African Americans regardless of genotyping platform, and allows collaboration with other groups who have genotyped their samples on different platforms. Imputation was performed separately for whites and the three sub-samples of African Americans using the single-step approach implemented in Markov Chain Haplotyper (MaCH) 1.0.16 (Li, 2006). The reference panel for imputation in whites was composed of the HapMap phased haplotypes (release 22) from 60 unrelated CEU samples (Utah residents with Northern and Western European ancestry) (The International HapMap Consortium, 2003). The reference panel for African Americans was composed of the HapMap phased haplotypes (release 22) from 60 unrelated CEU and 60 unrelated YRI samples (Yoruba from Ibadan, Nigeria). We have examined how three variables affect imputation results for GENOA African Americans: 1) using siblings versus only unrelated individuals, 2) the one-step versus the two-step procedure, and 3) using the CEU+YRI panel versus using the complete HapMap panel, and have concluded that these variables make very little difference in the imputation of genotypes in this sample (manuscript in preparation).

Prediction accuracy of the imputed depends upon LD between measured and unmeasured SNPs, so there is variability in the accuracy of the imputed genotypes for a single SNP in a single individual (de Bakker, 2008). For each SNP in each individual, posterior probabilities of each genotype are calculated and the effective allelic dosage (the expected number of copies of a specific allele, ranging from 0 to 2) is reported. We used allelic dosages for imputed SNPs in our analysis rather than the 'best guess' genotype (0,

1, or 2) in order to incorporate the uncertainty of genotype designation. The final SNP dataset used for association analysis uses directly genotyped Affymetrix genotypes for whites when available; otherwise dosages are used for imputed SNPs. Since only a small number of directly genotyped SNPs overlap on the Affymetrix and Illumina platforms used to genotype African Americans, imputed dosages were used for all association analyses in the African American sample.

A total of 2,543,889 SNPs in whites and 2,203,609 SNPs in African Americans were available for analysis in whites and African Americans, respectively. In order to prevent false positive associations due to a small number of people in a single genotype category, SNPs with a minor allele frequency less than 0.01 were removed. As an additional quality measure, imputed SNPs with an average posterior probability of less than 0.8 were also removed. Following these quality control measures, the total number of SNPs available for analysis was 2,401,820 in whites and 2,150,041 in African Americans.

Statistical analysis

Descriptive statistics

Data management and statistical analyses were conducted primarily in R version 2.8.0 (R Core Development Team, 2008). Allele frequencies and genotype frequencies were calculated for all SNPs using HelixTree (http://goldenhelix.com). Distributional plots indicated that the measures of leukoaraiosis volume are severely right-skewed, so this variable was transformed by taking the natural log of (leukoaraiosis + 1). The cognitive traits appear to have relatively normal distributions; thus, no variable transformations

were applied to these variables. T-tests were conducted for the outcome measures to test whether there were significant differences in the white and African American study participants.

Creation of residual phenotypes for GWAS

In the African American sample, Helix Tree and R were used to conduct principal components analysis in a random sample of unrelated individuals. Principal components were constructed using the genotypes of the SNPs that overlapped in the Affymetrix and Illumina platforms and were also in HapMap (207,565 SNPs). An additive model was assumed for the SNPs, which were standardized with a mean of 0 and variance of 1. A complete description of the procedure used to obtain principal components and results from association testing between the top ten principal components and outcome measures adjusted for covariates are presented in Appendix 4.1. Since different principal components were associated with different outcome measures, all of the first ten principal components were included as adjustment variables in the creation of all residual phenotypes for GWAS in the African American sample.

Age and gender were included as adjustment covariates because both have been historically used as adjustment variables for this trait. Age is a very strong independent predictor for leukoaraiosis, and gender has been shown to have a marginal association with leukoaraiosis in some samples, including GENOA whites. To account for differences in brain size, intracranial volume was also included as an adjustment variable.

Performance on cognitive testing is known to vary by age and gender, and relative performance on these tests is determining using age- and gender-specific population-based norms. Education also affects performance on some cognitive tests, as people with higher educational attainment tend to perform better (Staff, 2004; Valenzuela, 2006). To explore the relationships between adjustment variables and each outcome of interest in GENOA whites and African Americans, we conducted multivariable linear mixed models with adjustment covariates as predictor variables for each unadjusted outcome measure.

For each cognitive trait, the outcome variable for this analysis was the residual value of the trait adjusted for age at cognitive testing, gender, and education. In African-Americans, the first ten principal components were also added to the model to reduce the effects of population substructure. The adjustment model for the natural logarithm of leukoaraiosis plus one included all of the same covariates as well as total intracranial volume (TIV). Education was categorized (less than high school, completed high school (GED), some college, and completed college (4+ years) and coded as a continuous variable 0 to 3 scale, respectively.

Biometrical genetic modeling

The expected covariance of a trait between a pair of individuals is modeled as a function of the variance parameters and the expected correlation between the individuals for genetic effects (Sing, 1987). The additive genetic effects are unobservable and can only be modeled using the variance-covariance matrix of the trait, expressed as a function of identity-by-descent relationships as expressed below. Here, shared residual (non-genetic)

effects are assumed to be zero because the siblings are all adults and have reported living in separate households. Specifically, $\text{cov}(P_x, P_y) = 2\Phi_{x,y}V(G) + I_{x,y}V(R)$ where: $\Phi_{x,y}$ = kinship coefficient for two individuals, *x* and *y*. The kinship coefficient is the probability that a randomly selected allele from each person at the same locus is identical-by-descent. The kinship coefficient is ½ for an individual (if *x*=*y*) and ¼ for full siblings; $I_{x,y} = 1$ if *x* and *y* are the same individual, 0 otherwise (identity matrix); V(G) = deviation attributable to additive genetic factors; V(R) = deviation attributable to random residual effects (individual factors) including measurement error. A more detailed description of biometrical genetics is presented in Appendix 4.2.

In this study, SOLAR (Sequential Oligogenic Linkage Analysis Routines) (Almasy, 1998) was used to implement a variance component regression based on maximum likelihood estimation to partition phenotypic variance according to the following model:

$$y_i = \mu + \sum_{j=1}^{m} \beta_j x_{ij} + g_i + \varepsilon_i$$

Where:

$i = 1, 2, \ldots n$ (individuals)

$j = 1, 2, \ldots m$ (covariates)

y_i = outcome measure for the i^{th} individual

μ = population mean of *y*

x_{ij} = i^{th} value for the j^{th} covariate

β_j = regression coefficient associated with j^{th} covariate

g_i = additive genetic effect where $g_i \sim N(0, \sigma^2_g)$

ε_i = random residual effect where $\varepsilon_i \sim N(0, \sigma^2_e)$

The model is constrained so that $\sigma^2_g + \sigma^2_e = 1$. In this way, the heritability estimate (σ^2_g) is the heritability of the residual variance of the trait that is not accounted for by adjustment variables (the heritability of the trait after adjustment for measured covariates). Thus, the contribution of genetic effects to the total phenotypic variance when accounting for measured covariates is given by [(1-proportion of variance explained by covariates)*h^2]*100.

The heritability of the trait can be tested for significance by comparing the log-likelihood of the full model above to the log-likelihood of a model with σ^2_g constrained to 0 (i.e. $\mathbf{H_o}$: $\sigma^2_g = 0$ versus $\mathbf{H_a}$: $\sigma^2_g \neq 0$). The null distribution of the likelihood ratio test statistic is a 50:50 mixture of a χ^2_1 and a point mass at zero. Narrow sense heritabilities were estimated for the outcome variables (all cognitive traits and ln(leukoaraiosis+1)) both with and without adjustment covariates included in the biometric models.

Phenotypic, genetic, and environmental correlations

To examine the relationships among pairs of the cognitive testing and leukoaraiosis measures, two methods were used to obtain estimates of the correlation coefficient, Pearson's correlation coefficient and the estimate of phenotypic correlation using the SOLAR software package. In addition to estimating total phenotypic correlation among pairs of traits, the SOLAR software package simultaneously estimates the genetic and environmental correlations of the traits.

The SOLAR software package uses maximum likelihood methods to simultaneously estimate heritabilities of trait pairs (as described above) and the phenotypic, genetic, and environmental correlations between pairs of traits according to the formula below. The advantage of using SOLAR to estimate phenotypic correlation is that it uses the family relationships among the participants to perform estimations; thus, it properly accounts for the sibship structure in the GENOA data. The disadvantage is that due to the method of estimation, only a point estimate of the phenotypic correlation is given. Since the standard error for this parameter is not estimable, the significance of the phenotypic correlation cannot be tested. The phenotypic, genetic, and environmental correlations among all pairs of traits in both ethnic groups were estimated in SOLAR, both with and without adjustment covariates in the biometric models.

$$\rho_p = \rho_g \sqrt{h_1^2 h_2^2} + \rho_e \sqrt{(1 - h_1^2)(1 - h_2^2)}$$

Where:

ρ_p = phenotypic correlation between the traits

ρ_g = genetic correlation between the traits

ρ_e = environmental correlation between the traits

h_1^2 = heritability of trait 1

h_2^2 = heritability of trait 2

The genetic and environmental correlations between the traits estimated in SOLAR, ρ_g and ρ_e, can be tested for significance by comparing the log-likelihood of the model in which the parameter of interest is estimated to that of the model in which the parameter is fixed to 0. The test for pleiotropy, or evidence of shared genetic influences is as follows:

$\mathbf{H_o}$: $\rho_g = 0$ vs. $\mathbf{H_a}$: $\rho_g \neq 0$. The null distribution of the likelihood ratio test statistic statistic is a χ^2_1. Rejection of the null hypothesis provides evidence of pleiotropy.

The genetic correlation can also be tested for evidence of complete pleiotropy (all genetic influences are shared between the pair of traits) according to the following test: $\mathbf{H_o}$: $\rho_g = 1$ vs. $\mathbf{H_a}$: $\rho_g \neq 1$. The null distribution of the likelihood ratio test statistic is a χ^2_1. This statistic is used to test the null hypothesis of $\rho_g = 1$ (all genetic influences are shared between the traits). Rejection of the null hypothesis provides evidence that there are differences in the genetic influences on the traits.

Finally, the presence of shared environmental influences can be determined by the following test: $\mathbf{H_o}$: $\rho_e = 0$ vs. $\mathbf{H_a}$: $\rho_e \neq 0$. The null distribution of the likelihood ratio test statistic is a χ^2_1. This statistic is used to test the null hypothesis of $\rho_e = 0$ (no environmental influences beyond the adjustment variables are shared between the traits). Rejection of the null hypothesis provides evidence that there are shared environmental influences on the traits.

Genome-wide association

Genome-wide association was performed on the residual values of the seven cognitive variables and the natural log of leukoaraiosis+1, described above, separately in each ethnic group. Linear mixed effects modeling was used to test all associations between a single SNP and the outcome, as described in Chapter 3 (Raudenbush, 2002). Measured genotypes were coded additively as SNP = 0, 1, 2 for people that are homozygous for one

allele of a particular SNP, heterozygous, and homozygous for the other allele, respectively (Weir, 1996). Effective allelic dosages (between 0 and 2) were used for imputed genotypes.

Examining pleiotropy with GWAS

Assessing enrichment of GWAS results at a nominal significance level

The method developed by Karasik et al. (2010) for examining pleiotropy using percentage of shared SNP associations as a metric begins with determining the number of nominally significant GWAS results at $\alpha = 0.01$. The number of results expected by chance alone at this significance level for a particular trait, however, is dependent on the genome-wide inflation factor for the GWAS of that trait. Genome-wide inflation factors were calculated for each trait by dividing the median of the Wald statistic [$\beta/SE(\beta)$] of all SNPs included in the GWAS by the median of the absolute value of the T statistic with the appropriate degrees of freedom under a null distribution. Using a Binomial distribution, we next determined the p-value for obtaining a number greater than or equal to the number of nominally significant SNPs using the genome-wide inflation factor as the probability of success. For example, if a particular GWAS with 2,400,000 SNPs had 25,000 nominally significant SNP associations and a genome-wide inflation factor of 1.011, we calculated a p-value for obtaining 25,000 or greater nominally associated SNPs under a Bin(2,400,000, 0.01011) distribution.

Relationship between genetic/environmental correlations and GWAS results

To examine evidence for pleiotropy using results from the GWAS analysis, the number

of shared marginally significant associations (p-value < 0.01) was used to calculate the percentage of shared SNP associations for each pair of traits according to the method used by Karasik et al. (2010): [(number of shared associated SNPs) / (total number of non-union associated SNPs)]*100. For example, if there are X SNPs associated with trait 1, Y SNPs associated with trait 2, and Z SNPs associated with both traits 1 and 2, the percentage of shared SNP associations would be [Z/(X+Y-Z)]*100. This percentage represents a quantitative measure of the genetic similarity between trait pairs. The percentage of shared SNP associations was plotted against the genetic and environmental correlations estimated in SOLAR to examine the relationship between these two methods of assessing genetic similarity.

Permutation tests

For a subset of traits in whites, permutation testing was used to empirically generate a null distribution for determining whether the number and percentage of shared associated SNPs for a given trait pair was greater than expected by chance alone using two approaches. In the first approach, each trait was permuted individually, disturbing the correlation structure among the traits ("unpaired" approach). The permutation was performed three times to obtain three permutations of each trait. In the second approach, all traits were permuted as a vector, preserving the correlation structure among traits ("paired" approach). This permutation procedure was also performed three times to obtain three permutations of each trait vector. For both approaches, GWAS was performed for each trait (three GWAS total for each permuted trait under each approach).

For each GWAS, SNPs were split into 48 groups of 50,000 SNPs each according to their chromosomal location (the first group consisted of the first 50,000 SNPs genotyped on chromosome 1, the next group consisted of the next 50,000 SNPs genotyped on chromosome 1, etc). For each of the 48 groups of SNPs, the number and percentage of shared associated SNPs at α=0.01 was calculated as described above. The expected number of shared associated SNPs that would have been present in a full GWAS using 2,401,820 SNPs was extrapolated by multiplying the number of shared associated SNPs in each group by 2,401,820/50,000. Next, for each pair of traits under each approach, the results from the three GWAS were combined to obtain a null distribution of the number and percentage of shared SNP associations expected by chance alone, each with 144 datapoints. Finally, the observed number and percentage of shared SNP associations were compared to the distributions from the permutation tests to determine an empirical one-sided p-value. The null hypothesis for these permutation tests is that the observed number or percentage of shared SNP associations is not greater than chance alone either given the correlation structure of the traits ("paired approach") or under an assumption of independence between the traits ("unpaired approach").

Results

Descriptive statistics

Descriptive statistics of the outcome measures and adjustment variables for the 762 white and 720 African American participants, as well as T-tests comparing the samples, are presented in Tables 4.2-4.4. GENOA whites are 58.1% female, have a mean age at the time of cognitive testing of 61.3 years, and have a mean leukoaraiosis volume of

8.11cm^3. GENOA African Americans have a much larger percentage of females (72.6%), have a higher mean age of cognitive testing (63.3 years), and have a higher mean volume of leukoaraiosis with greater variability (9.56cm^3). Approximately half of both whites and African Americans attended at least some college; however, only 5.2% of white participants did not graduate from high school or obtain a GED while this was true for 28.3% of African American participants. The mean values for all outcome measures were significantly different in whites and African Americans except for the Stroop difference. Leukoaraiosis was strongly right skewed in both populations, but had a relatively normal distribution after taking the natural log of leukoaraiosis+1. Distributional plots of the cognitive outcomes show that they also have relatively normal distributions (Appendix 4.3).

Associations between adjustment covariates and outcome measures

In order to explore the relationship between adjustment variables and each outcome of interest, we conducted multivariable linear mixed models with adjustment covariates as predictor variables for each unadjusted outcome measure (Tables 4.5-4.6) In both whites and African Americans, age, gender, and education were significant predictors for all cognitive measures except that education was not a significant predictor of Stroop difference in whites and gender was not a significant predictor of Stroop color word in African Americans after accounting for the other adjustment variables. As expected, increasing age was associated with lower cognitive scores, while increasing education was associated with higher cognitive scores. Female gender also showed a trend of being associated with higher cognitive scores. In both groups, increasing age and total

intracranial volume were associated with increasing leukoaraiosis volume, while gender and education were not associated with this measure.

The amount of variance explained by the adjustment covariates, as measured by R^2, showed a consistent pattern between the two groups. R^2 was lowest for Stroop difference (0.018 in whites and 0.074 in African Americans) and highest for DSST (0.411 in whites and 0.529 in African Americans). The amount of variance explained by adjustment variables in the remainder of the cognitive measures ranged from 0.122 to 0.309 in whites and from 0.207 to 0.294 in African Americans. Variance of ln(leukoaraiosis+1) explained by adjustment variables was higher in whites (0.309) than in African Americans (0.213).

Genetic variance (heritabilities)

In order to examine the contribution of genetic factors to the observed variation in the traits, we used a biometrical approach to estimate the proportion of variance in the traits explained by genetic factors (heritabilities) both with and without inclusion of adjustment variables in the models (Tables 4.7-4.8). Heritabilities of all traits were highly significant in both whites and African Americans (p-value<0.001) with the exception of Stroop difference in African Americans that showed only a marginally significant heritability, illustrating that all of the traits under study are influenced by genetic factors. Similar patterns of heritability were observed between the two groups, though African Americans tended to have lower heritabilities than whites for most traits.

Heritabilities in unadjusted traits were lowest for Stroop difference (0.302 in whites, 0.135 in African Americans) and highest for DSST (0.774 in whites, 0.81 in African Americans), with the majority of heritabilities in the range of 0.45 to 0.6. Leukoaraiosis had a higher heritability in whites (0.656) than in African Americans (0.485). After including adjustment covariates in the biometric models, heritabilities for the traits were generally lower but remained highly significant. Again, Stroop difference had the lowest heritability in both groups (0.275 in whites, 0.154 in African Americans) and DSST had the highest (0.843 in whites, 0.556 in African Americans), with the remaining traits ranging between 0.33 and 0.54. Leukoaraiosis still showed higher heritability in whites (0.529) than in African Americans (0.432).

The proportion of the observed trait variance accounted for by adjustment covariates estimated with biometric modeling mirrored the relationships we observed in multivariate linear mixed modeling, described above. The lowest proportion of variance explained was for Stroop difference (0.023 in whites, 0.075 in African Americans) and the highest was for DSST (0.403 in whites, 0.525 in African Americans). For the remainder of the traits, the proportion of variance explained by adjustment covariates ranged from 0.139 (COWA FAS) to 0.311 (leukoaraiosis) in whites and from 0.203 (RAVLT delayed recall) to 0.3 (COWA FAS) in African Americans.

In order to determine the proportion of variation in the traits explained by genetic factors, we multiplied the proportion of variation not explained by the covariates by the heritability. Expressed as a percentage of total variation, genetic factors explain the

lowest amount of variation in Stroop difference in both groups (26.87% in whites, 14.25% in African Americans). The largest amount of variation explained by genetic factors in whites was for DSST (50.33%) followed by RAVLT delayed recall (40.92%) and leukoaraiosis (36.45%). In African Americans, genetic factors explained the largest percentage of variation in COWA FAS (37.52%) followed by leukoaraiosis (33.83%) and Stroop color word (33.79%). For the majority of traits, the amount of variation explained by genetic factors was lower in African Americans than in whites, but most traits in both groups had at least 25% of variation explained by genetic factors, showing that genetics has an important influence on all of these traits.

Phenotypic correlations between pairs of traits

Pearson's correlation coefficient

To examine the relationships among pairs of the cognitive testing and leukoaraiosis measures, both Pearson's correlation coefficient and biometric modeling in SOLAR were used to obtain estimates of trait correlations. Tables 4.9-4.10 show correlations estimated using Pearson's correlation coefficient. In both samples, the majority of unadjusted cognitive trait pairs exhibited positive and highly significant (p-value<0.001) correlations. Correlation was lowest for Stroop difference and RAVLT delayed recall (0.053 in whites and 0.148 in African Americans) and highest for RAVLT delayed recall and RAVLT total learning (0.816 in whites and 0.789 in African Americans). The exception was that Stroop difference and Stroop color word showed a negative correlation in both groups (-0.13 in whites and -0.186 in African Americans). Possible reasons for this observed difference are presented in the Discussion. Unadjusted

leukoaraiosis showed a strong negative correlation with the majority of the cognitive traits in both groups, ranging from -0.072 to -0.298 in whites and from -0.094 to -0.259 in African Americans. This pattern is expected, since lower values of leukoaraiosis and higher cognitive function are both strongly associated with lower age.

The overall pattern of correlations among adjusted cognitive trait pairs were similar to that of unadjusted cognitive trait pairs. In both groups, the strongest correlation was between the two measures from the RAVLT (0.755 in whites, 0.733 in African Americans), and the only non-significant correlation observed was between Stroop difference and RAVLT delayed recall. For cognitive traits, the remaining correlations ranged from 0.09 (Stroop difference and RAVLT total learning) to 0.363 (COWA FAS and COWA animals) in whites and from 0.127 (Stroop difference and COWA animals) to 0.399 (COWA FAS and COWA animals) in African Americans. The majority of the correlations among cognitive traits were moderate (in the range of 0.1 to 0.3), were highly significant, and were similar across groups. Leukoaraiosis showed negative and non-significant correlations with the majority of the cognitive traits in both groups, with correlations ranging from -0.001 to -0.088. Only one marginally significant correlation between leukoaraiosis and a cognitive trait (RAVLT total learning, correlation=-0.088) was observed in whites, and no significant correlations between these traits were observed in African Americans.

Correlation coefficient estimated biometrically in SOLAR

The patterns observed in the correlations estimated biometrically in SOLAR closely match those estimated with the Pearson's correlation coefficient (Tables 4.11-4.12). Again, the strongest correlation in the adjusted traits was between the two RAVLT measures (0.755 in whites, 0.729 in African Americans), and the weakest correlations were between leukoaraiosis and all cognitive traits (ranging from -0.001 to -0.083). In general, multiple measures from the same test exhibited stronger correlations than measures across tests, which is intuitive since measures from the same test are assessing different but closely related cognitive functions. Patterns of correlation in whites and African Americans were very similar, though whites generally tended to exhibit somewhat stronger correlations.

Genetic and environmental correlations estimated biometrically in SOLAR

In order to begin to understand the extent to which pleiotropic genetic effects may be contributing to each pair of traits, we used a biometrical approach to estimate genetic and environmental correlations (Tables 4.13-4.14). Overall, there were far more significant genetic correlations (pleiotropic effects) between trait pairs than environmental correlations, indicating that shared genetic effects were more common in these pairs of traits than shared environmental effects. For all estimates of genetic and environmental correlations, adjustment covariates were included in the biometric models.

The majority of significant genetic correlations observed were in whites. In whites, significant genetic correlations ranged from 0.263 (RAVLT total learning and DSST) to

0.918 (RAVLT total learning and RAVLT delayed recall). Other highly significant genetic correlations (p-value<0.001) were between DSST and Stroop color word (0.7) and between RAVLT total learning and COWA animals (0.55). Many of the pairs involving RAVLT, COWA, and DSST also showed significant genetic correlations, ranging from 0.263 to 0.476. Leukoaraiosis and RAVLT had a marginally significant negative genetic correlation (-0.28), indicating that genes shared between these two traits have opposite effects on the traits (for example, a certain genetic variation may increase leukoaraiosis volume while decreasing learning scores). No other evidence of pleiotropic effects was found between leukoaraiosis and cognitive measures.

In contrast to the relative abundance of genetic correlations between these measures, there were very few significant environmental correlations in whites. The most significant environmental correlation was between the trait pair that also had the highest genetic correlation, RAVLT total learning and RAVLT delayed recall (0.586), although the environmental correlation was substantially less than the genetic correlation. This indicates that for this trait pair, shared genetic effects have a stronger influence than shared environmental effects, though both contribute to the observed strong phenotypic correlation. The only other highly significant environmental correlation was between Stroop color word and Stroop difference (-0.427). This correlation is negative since low cognitive performance is indicated by a low score on Stroop color word but by a high score on Stroop difference. Only two other trait pairs exhibited even marginally significant environmental correlations in whites. Leukoaraiosis had a negative

environmental correlation with Stroop difference (-0.274) and the two measures from the COWA had a positive environmental correlation (0.293).

The overall patterns of genetic and environmental correlations in African Americans were strikingly similar to the patterns observed in whites. However, many of the genetic correlations in African Americans did not reach statistical significance due to larger standard errors in their estimates, and there was slightly more evidence of shared environmental effects. Two of the four highly significant genetic correlations observed in whites were also observed in African Americans. RAVLT total learning and RAVLT delayed recall were the most strongly genetically correlated (0.915) followed by DSST and Stroop color word (0.698). The only other significant genetic correlations were between the two measures of COWA (0.533) and between COWA FAS and Stroop color word (0.363). There were no significant genetic correlations between leukoaraiosis and any of the cognitive traits.

As with whites, the most highly significant environmental correlation was between the two measures of RAVLT (0.596) and between Stroop color word and Stroop difference (-0.541). The direction and magnitudes of the correlations for these traits were also the same in African Americans as they were in whites. The other strongly significant environmental correlations observed in African Americans were between DSST and RAVLT total learning (0.401) and DSST and COWA FAS (0.442). Four additional pairs of traits also exhibited marginally significant environmental correlations, including the

two COWA measures (0.313) that also showed a marginally significant environmental correlation in whites.

Genome-wide association

The quantile-quantile plots, Manhattan plots, and genome-wide inflation factors for the GWAS of the residual values for leukoaraiosis and cognitive traits are presented in Appendix 4.4. Overall, the quantile-quantile plots are suggestive of the p-value distribution that would be expected by chance alone.

Examining pleiotropy with GWAS

Percentage of shared SNP associations

To examine evidence of pleiotropy using results from the GWAS analysis, the percentage of SNPs that had nominal evidence of association (p-value ≤ 0.01) was calculated as a quantitative measure of the genetic similarity between trait pairs. The percentage of shared SNP associations was plotted against the absolute value of the genetic and environmental correlations estimated in SOLAR to examine the relationship between these two methods of assessing genetic similarity (Appendix 4.5). RAVLT delayed recall and RAVLT total learning had a clear outlying value for both the percentage of shared SNP associations as well as the genetic correlation (both were much larger values than observed in the other trait pairs), so this pair was removed from the analysis. To examine the extent to which genetic correlation explained the variability of the percentage of shared SNP associations beyond that explained by the environmental correlation in the remaining 15 trait pairs, the difference in R^2 between a linear model containing genetic

correlation and environmental correlation as predictors of the percentage of shared SNP associations was compared to a model containing only environmental correlation as a predictor. In whites, R^2 for the model with both genetic and environmental correlation was 0.688, while R^2 for the model with only environmental correlation was 0.081. Thus, genetic correlation explained an additional 60.7% of variability in the percentage of shared SNP associations. In African Americans, the R^2 values were 0.567 and 0.109 respectively, resulting in an additional 45.8% of variability in the percentage of shared SNP associations beyond environmental correlation.

The numbers and percentages of SNPs associated with each trait and shared SNP associations for each trait pair are presented in Tables 4.15-4.16. For single traits, the number of results with nominal evidence of significance (p-value ≤ 0.01) ranged from 0.98% (Stroop color word) to 1.09% (leukoaraiosis and RAVLT delayed recall) in whites and from 0.87% (Stroop difference) to 1.1% (RAVLT delayed recall) in African Americans. Of the eight traits analyzed in whites, four had more nominally associated SNPs at this significance level than expected by chance alone given the GWAS inflation factors for the traits, and only Stroop color word had fewer nominally associated SNPs than expected by chance alone. In African Americans, four traits had more nominally associated SNPs than expected by chance alone given the GWAS inflation factors, and three had fewer than expected by chance alone. This suggests that some of the traits show a small enrichment of significant genetic effects at this nominal significance level, including leukoaraiosis, RAVLT total learning, and COWA FAS in both whites and African Americans.

The percentage of shared SNP associations (p-value ≤ 0.01) in whites ranged from 0.4% (COWA animals and leukoaraiosis) to 16.06% (RAVLT delayed recall and RAVLT total learning). All of the remaining percentages of shared SNP associations were between 0.48% and 3.01%. In African Americans, the percentage of shared SNP association ranged from 0.33% (leukoaraiosis and Stroop difference) to 15.14% (RAVLT delayed recall and RAVLT total learning). The remaining percentages in African Americans were between 0.38% and 3.1%. The overall distribution of percent shared SNP associations was strikingly similar across the two groups.

Permutation testing

As an exploratory analysis to begin to characterize the utility of the percentage or number of shared SNP associations at a nominal significance level (α=0.01) as a metric for assessing pleiotropy, we performed permutation testing of these measures for four traits in whites (a total of six trait pairs) that exhibited a range of phenotypic and genetic correlations as measured by the biometrical approach in SOLAR. Appendix 4.6 is a summary review of the phenotypic, genetic, and environmental correlations of these trait pairs in whites.

The paired and unpaired approaches were used to generate the distribution of the number and percentage of shared SNP associations with the trait correlation structure preserved ("paired approach") or under the assumption of independence between the traits ("unpaired approach"). A significant empirical p-value for a pair of traits in the unpaired

approach would indicate that the correlation structure of the traits results in a greater number or percentage of shared SNP associations than by chance alone, suggesting that some aspect of the correlation structure of the traits (i.e. the total phenotypic or genetic correlation) results in a larger amount of shared SNP associations. A significant empirical p-value for the paired approach would indicate that the trait correlations alone are not leading to higher numbers and percentages of shared SNP associations, and would suggest that the genetic correlation (pleiotropy) is leading to larger amounts of shared association. We were especially interested in the permutation testing results from pairs of traits that exhibited a relatively high genetic correlation (for example, RAVLT total learning and COWA animals, ρ_g=0.55) but a smaller phenotypic correlation (ρ_p=0.311).

Permutation testing results for the paired and unpaired approaches are presented in Tables 4.17a-b (percentage of shared SNP associations) and 4.18a-b (number of shared SNP associations), and examples of histograms of the distributions of percentages and numbers of shared associated SNPs are presented in Appendix 4.7. None of the permutation tests from the paired approach showed a significant empirical p-value. Nominally significant empirical p-values ($0.1<p<0.05$) were observed in two pairs of traits from the unpaired approach in both the percentage and number of shared SNP associations. These two pairs of traits (RAVLT total learning and DSST, and RAVLT total learning and COWA animals) had the highest phenotypic correlations of the trait pairs tested (0.269 and 0.311, respectively), which indicates that it is likely that the phenotypic correlations were the most important factor leading to their relatively high percentage and number of shared SNP associations.

Discussion

It is well known that leukoaraiosis contributes to cognitive impairment and cognitive decline in individuals who have not yet progressed to dementia (Pantoni, 2007; De Groot, 2002; Schmidt, 2007), and leukoaraiosis progression is correlated with progression in cognitive decline (Kuller, 1998; Swan, 1998; Pantoni, 1997). Several studies have demonstrated an association between hypertension in midlife and cognitive dysfunction in later life (Elias, 1993; Launer, 2000), and it has been hypothesized that this is due to the cumulative effects of subclinical damage due to cerebrovascular disease (Swan, 1998; Knopman, 2001) with leukoaraiosis as one of the main mechanistic pathways implicated (Sierra, 2006). Several studies have demonstrated that cognitive decline and general cognition in midlife are also associated with hypertension, lending credence to the claim that later cognitive dysfunction and dementia are the clinical manifestations of a disease process in the brain that is cumulative throughout the lifespan (Knopman, 2001; Knecht, 2009; Knecht, 2008).

Hypertension and leukoaraiosis have been shown to have stronger effects on some cognitive and physical functions than others. In a sample of 1,702 participants from the Framingham Heart Study, Elias et al. (1998) showed that the cognitive areas most affected by hypertension are psychomotor speed, visual memory, learning, memory, and executive function. The association between hypertension and both immediate memory and attention were also demonstrated by Sleegers et al. (2007) in an extended pedigree of 780 individuals. Findings from the ARIC study showed that presence of hypertension at

baseline was associated with cognitive decline over a 6-year period in participants aged 47 to 70 at baseline (mean age = 56.8 years) (Knopman, 2001). In this study, cognitive decline was measured by delayed word recall, processing speed, and word fluency. They found that decline in both processing speed and word fluency scores were significantly more pronounced in hypertensive subjects, with processing speed showing the strongest association. In a follow-up paper, Alves de Moraes and colleagues (2002) reported that decline in processing speed was also significantly different between normotensive and untreated hypertensive subjects in the ARIC cohort, particularly in those with older ages. An association between decline in processing speed scores and hypertension was also demonstrated in the Cardiovascular Health Study (Haan, 1999). It has been hypothesized that processing speed was most strongly associated with hypertension because it was a timed test and thus may be more affected by subcortical lesions (Knopman, 2001). Leukoaraiosis itself appears to be more strongly associated with decreasing cognitive performance than memory and is also associated with a decline in motor performances such as gait disturbances (Pantoni, 2007; Schmidt, 2007).

Heritability of leukoaraiosis

Heritability of white matter hyperintensities on MRI was estimated to be 0.71 in study of male twins after adjustment for age and head size (Carmelli, 1998) and 0.55 in the Framingham Heart Study after adjustment for sex, age, age^2, and total cranial volume (Atwood, 2004). Turner et al. (2009) estimated the heritability of the logarithm of leukoaraiosis in GENOA participants as 0.49 in whites and 0.45 in African Americans after adjustment for age, sex, and brain volume. In an earlier publication, Turner et al.

(2004) showed that leukoaraiosis has a consistently high heritability even after adjustment for blood pressure. These high heritabilities imply that much of the inter-individual differences in variation are due to differences in genetics.

Heritability of cognitive function

Heritability of cognitive functioning has been conducted primarily in twin studies using factor analysis to identify a common factor to act as a proxy for overall cognitive function, and the findings point to a high heritability for this measure. McGue et al. (2002) estimated the heritability of general cognitive function as measured by five cognitive tasks comprised of fluency, digit span, and recall to be 0.70. Finkel et al. (1995) peformed quantitative genetic analysis on four measures of cognitive function (verbal, spatial, perceptual speed, and memory) and showed that heritability for a general cognitive factor was between 0.54 and 0.81 in two samples across several age groups of adults (from young to elderly).

Different areas of cognitive function have been shown to have different heritability estimates, as well as differences in the trajectories of changing heritability over the age span. Neale, Carmelli, and colleagues (1997) used multivariate analysis to detect a common pathway model which revealed a latent "executive control factor" that adequately summarized the results of four cognitive tests that were related only to executive control requiring perceptual-motor functions (sustained attention, visual perception, language, and short-term memory). Heritability of this executive control factor was estimated to be 0.70, and they found that processing speed had strongest

genetic influence while verbal fluency seemed to have the most difference in terms of genetic and environmental control compared to the other three tests (Carmelli, 2002). Other estimates of executive function range from 0.34 to 0.68 (Sleegers, 2007; Swan, 2002). Measures of memory generally seem to have a lower heritability than executive function. Memory as assessed by immediate and delayed recall of a word list has been estimated to be between 0.16 (Sleegers, 2007) and 0.4 (McGue, 2001; Plomin, 1994), with immediate recall showing a slightly higher heritability (0.24) than delayed recall (0.16) (Sleegers, 2007).

Bivariate variance component analysis in leukoaraiosis and cognitive traits

Carmelli (2002) used maximum likelihood nested modeling techniques to estimate the proportion of variance in leukoaraiosis and four measures of cognitive function due to genetic, shared environmental, and non-shared environmental effects in 142 pairs of elderly twins (Carmelli, 2002). They then used bivariate genetic analysis to quantify the genetic and environmental overlap between leukoaraiosis and cognitive function. Bivariate analysis between leukoaraiosis and executive control function, a single measure of executive function constructed by factor analysis of the four cognitive traits, showed that the total phenotypic correlation was -0.20, and that 70% of the total phenotypic correlation was accounted for by shared genes while 30% was accounted for by shared environments. While this is substantial, the contribution of overlapping genes (genes shared between the executive control factor and leukoaraiosis) to the genetic variance in executive function was only 8%. Other studies of the shared genetic components of cognitive traits tend to focus on change in cognition over time. Variance components

analysis of the relationship between cognitive change and perception speed in a study of 292 twins aged 40-84 revealed that 90% of the age-related variance and 70% of the genetic variance in cognitive function was shared with perception speed, demonstrating that there is a genetic component to processing speed which also influences general cognitive functioning (Finkel, 2000).

Pleiotropic mechanisms

Examining the way that pleiotropic mechanisms operate to affect complex human traits is an important area of research because it provides a means to obtain a more sophisticated understanding of the relationships within the biological pathways that underlie trait variation. The recent research indicating the existence of highly correlated and inter-connected transcriptional modules calls for a deeper exploration of the way that variation in gene expression patterns may affect multiple traits. This research also suggests that genetic variations that impact gene expression may affect a variety of traits that seem unrelated without knowledge of the underlying biology. A greater understanding of the underlying pleiotropic mechanisms contributing to human health and disease has the potential to allow for earlier identification of individuals at increased risk for disease, the development of more efficacious treatments, and the tailoring of particular treatments to people most likely to respond positively.

Characterization of pleiotropic mechanisms

Studies of pleiotropy in humans and animals have to grapple with the different ways that a single genetic variation can impact multiple traits. Hodgkin (1998) began to

characterize the different types of pleiotropy that operate in model organisms from the viewpoint of a geneticist, defining seven different types according to the molecular mechanism involved. This classification system may be informative as a starting point for classifying pleiotropy in humans, but it is most relevant to the study of evolutionary mechanisms in model organisms. However, the geneticist's view of pleiotropy has yet to be mapped onto the traditional modeling tools that epidemiologists use to describe causal relationships, such as directed acyclic graphs (DAGs).

In this study, biometrical modeling showed substantial evidence of pleiotropy underlying variation in leukoaraiosis and cognitive traits, particularly among pairs of cognitive traits. There are several underlying biological mechanisms that may have been contributing to the high degree of pleiotropy observed. While the evidence for shared genetic effects between leukoaraiosis and the cognitive traits did not reach the level of statistical significance in most instances, it may be possible these effects exist but were not significant in this sample due to both a small sample size and limited variation in leukoaraiosis due to the relatively young age of the sample. The evidence for pleiotropy among cognitive traits, however, warrants discussion for the potential pleiotropic mechanisms that may be functioning in this group of traits.

Perhaps the most straightforward explanation for shared genetic effects between leukoaraiosis and cognitive measures is that there is a direct casual relationship between variations in genes that affect hypertension and phenotypes of the downstream sequelae of hypertension (i.e., leukoaraiosis and ultimately cognition). In this case, illustrated as a

DAG with a single gene in Figure 4.1 (“pleiotropic mechanism A”), shared genetic effects are due only to the effects of the gene on hypertension itself. Genes involved in vasculature function, salt regulation, and cholesterol transport are examples of genes that may contribute to pleiotropy in this manner. Another possibility is that a genetic variation affects the development of leukoaraiosis independent of hypertension, and that increased leukoaraiosis has a direct causal effect on changes in cognition (Figure 4.1, “pleiotropic mechanism B”). Genes that play a role in regulating the extent of inflammation or response to oxidative stress after a period of ischemia may show evidence of shared genetic effects under this pleiotropic mechanism. A third mechanism that may be responsible for shared genetic effects between leukoaraiosis and cognitive measures is a genetic variation that affects both traits through different biological pathways (Figure 4.1, “pleiotropic mechanism C”). For example, genetic variation that leads to Alzheimer pathology, such as the *APOE* ε4 allele, may demonstrate pleiotropy through this mechanism. Alzheimer pathology is known to have a direct effect on cognition independent of leukoaraiosis, and there may also be a synergistic effect between Alzheimer pathology and development of leukoaraiosis, which would lead to detection of shared genetic effects between leukoaraiosis and cognitive traits.

While there was a moderate degree of shared genetic effects observed among leukoaraiosis and cognitive traits, a much larger degree of shared genetic effects was observed among pairs of cognitive traits. Pleiotropic mechanisms A or B may partially account for these shared genetic effects, as well as a fourth mechanism illustrated in Figure 4.1 (“pleiotropic mechanism D”) in which genetic variation affects aspects of

cognition independent from the leukoaraiosis pathway. Genes that function in this manner may include those that regulate neurocognitive development, contribute to cognitive aging, and affect general information processing capabilities through a variety of biological pathways including those related to axon and synapse formation and function.

Distinguishing between pleiotropic mechanisms

Multiple variable adjustment modeling could be used to distinguish between the different mechanisms of pleiotropy that may be operating in this study. For example, to distinguish between pleiotropic mechanisms A and B, leukoaraiosis could be adjusted for hypertension status or a quantitative of blood pressure measurement. If the SNP of interest were still associated with leukoaraiosis after adjustment, it would indicate that pleiotropic mechanism B is more likely to be acting than mechanism A. Likewise, to distinguish between mechanisms B and C, cognitive measures could be adjusted for leukoaraiosis volume. If the SNP of interest were still associated with the cognitive trait after adjustment for leukoaraiosis, this would provide evidence for mechanism C.

It is also important to verify whether relationships are due to potentially causal mechanisms or simply to confounding as a result of correlations among traits. For example, if a SNP of interest and leukoaraiosis both had independent causal effects on a cognitive trait (mechanism D), the SNP may show association with leukoaraiosis due to confounding. In this case, it would be necessary to adjust leukoaraiosis for the cognitive trait and test whether there was still an association between the SNP and leukoaraiosis

after adjustment. If there is still an association, it indicates that mechanism C is operating, but absence of an association would indicate that mechanism D is operating.

Lessons from animal studies of pleiotropy

The most extensive examinations of pleiotropy to date have been done in animal models. These studies provide insight into the range of pleiotropic influences across different types of traits, the complexity and extent of epistatic pleiotropy, and new ways of relating biometrical measures of pleiotropy to molecular mechanisms through systems biology approaches to transcriptomic analysis. For example, Kenney-Hunt and colleagues (2006) examined the genetic architecture and pleiotropic patterning of body size traits in mice, and found differing degrees of pleiotropy among categories of traits. Overall, 97 quantitative trait loci (QTLs) were identified for four bone length traits, four organ weight traits, and necropsy weight. Thirty-five of these QTLs demonstrated pleiotropy, with the majority of the pleiotropy (~85%) observed among traits of a single category (for example, among bone length traits), but a substantial amount (~15%) was also observed across trait categories. A greater degree of pleiotropy existed among bone length traits than organ weight traits, demonstrating that pleiotropic loci may have more ubiquitous effects in more strongly correlated, developmentally related traits. This observation is supported by other murine studies (Leamy, 2002; Wolf, 2006). Patterns of pleiotropy observed in the mouse model explains why higher levels of genetic correlation are found among functionally and developmentally related traits in mice and other organisms (Wolf, 2006). These patterns of genetic correlation have implications for trait

evolution, and have been proposed to reflect a history of selection for genetic integration of functionally related traits (Cheverud, 2004).

Extensions to the classic quantitative genetic theory of pleiotropy have recently been developed for assessing epistatic pleiotropy in mouse models (Wolf, 2005). In these initial studies, epistatic pleiotropy accounted for only a moderate amount of the genetic covariance between traits such as limb bone length and organ weight but was a widespread phenomena. Consistent with the theory of selection for genetic integration of functional trait groups, limb bone lengths display a higher degree of epistatic pleiotropy than organ weights. A review of complex behavioral traits in Drosophila also indicates that large numbers of pleiotropic genes interact in an epistatic fashion to regulate behavioral traits (Mackay, 2009).

Systems biology approaches are in early development for studying pleiotropic mechanisms influencing quantitative traits in model organisms (Mackay, 2009). This approach has led to the identification of biologically relevant modules of highly correlated transcripts, including modules that are associated with tissue-specific gene expression, transcription factor binding sites, and a variety of gene ontology categories (Magwire, 2010; Jumbo-Lucioni, 2010; Ayroles, 2009). This finding of the modular organization of genes and proteins extends to a variety of model organism systems including C. elegans (Zou, 2008; Gunsalus, 2005) and yeast (Ihmels, 2002; Han, 2004). Another consistent finding is that there is a significant amount of overlap of common transcripts between modules associated with different traits (Jumbo-Lucioni, 2010;

Ayroles, 2009). Zou et al. (2008) propose that due to the presence of pleiotropic genes, gene modules must overlap instead of being separate from one another, and that pleiotropic genes may act as the "connectors" between modules.

Limitations

While this study offers preliminary evidence of pleiotropy among measures of cognitive function and leukoaraiosis in both whites and African Americans, it has several limitations. The relatively small sample size and the limited variability of the leukoaraiosis phenotype are perhaps the largest limitations. While the sample size was adequate to detect large heritabilities, the large standard errors for many of the genetic correlations indicated that the sample size may not have been adequate to reliably estimate genetic correlations. The lack of power was compounded for estimates of genetic correlation between leukoaraiosis and each cognitive trait due to the limited variability (and thus co-variability) of leukoaraiosis in this relatively young sample. The inadequate sample size also resulted in lack of power to identify significant SNP associations in GWAS despite the relatively large heritabilities of the examined traits. This lack of power in GWAS made examining the utility of using percentage of shared SNP associations as a metric for evaluating pleiotropy inconclusive. However, our finding that percentage of shared SNP associations had a much stronger relationship to absolute genetic correlation than absolute environmental correlation was consistent with the other published study that used percentage of shared SNP associations as a metric of pleiotropy in human traits (Karasik, 2010).

A further limitation was our ability to perform a comprehensive assessment of the number and percentage of shared SNP associations expected by chance along using permutation testing. The computational burden of performing enough permutation GWAS to generate accurate estimates and distributions of the number and percentage of shared SNP associations expected (say, 1,000 permutations) prevented us from using this approach on all of the trait pairs and limited the conclusions that can be drawn from the testing that we did perform. A more computationally feasible approach for estimating the percentage of shared SNP associations expected by chance alone for a pair of traits involves using simulating SNPs (Karasik, 2010), and this method may be an alternative for evaluating percentage of shared SNP associations as a metric in the future with GENOA data. While our permutation testing was only exploratory to begin to get a sense of the relationship between percentage of shared SNP associations and biometrical evidence of pleiotropy, we believe that it is important to examine these relationships in order to better understand the extent to which SNPs identified through GWAS are contributing to genetic correlation.

Most studies of the genetic and environmental factors associated with leukoaraiosis have had samples composed of individuals who have already experienced clinical endpoints such as stroke or severe cognitive decline. The relatively young age range of our sample provided the unique opportunity to examine the relationships between leukoaraiosis and cognitive phenotypes in asymptomatic individuals, at a time when preventive treatment would be most effective. However, the young age of our sample also imposed constraints, including the limited variability in the leukoaraiosis phenotype. In addition, it has been

shown that heritabilities of cognitive traits tend to vary with age (Alves de Moraes, 2002; Knopman, 2001; Mattay, 2008). Since this study was cross-sectional, we did not have the ability to examine how heritabilities and genetic correlations change over time. Differential heritabilities across age groups also implies that there may be SNPs that show age-related changes in penetrance with respect to cognitive traits. Thus, a further limitation of this study was that we were only able to detect associations with SNPs that exert their influence in the age range of our sample.

Future directions

One of the most important first steps for research into pleiotropic mechanisms in humans is to further evaluate the relationship between measured SNPs and genetic correlations between pairs of traits. One approach that has potential to be useful in this area is to adapt the method used by Yang et al. (2010) used to estimate the variation in height explained by simultaneous modeling of all measured SNPs. This method could be modified to estimate the amount of co-variation between a pair of traits that is explained by measured SNPs. Though this method would not identify which specific SNPs are accounting for co-variation, it would provide information about whether genetic correlations are driven by SNPs that are measured on genotyping arrays.

Future directions for examining pleiotropy using GWAS results will include performing bioinformatic research on the SNPs that are identified as having an association with trait pairs, which may not be warranted until we have additional evidence that percentage of shared SNP associations is a useful measure of pleiotropy. One possible approach for

evaluating the contribution of shared associated SNPs to biometrically estimated genetic correlations is to perform principal components analysis on the shared associated SNPs for a pair of traits and add the first several principal components to the bivariate model in SOLAR. If the genetic correlation for the trait pair is significantly reduced, it indicates that the shared associated SNPs are accounting for a portion of the biometrically estimated genetic correlation. This approach does not identify specific SNPs that contribute to high genetic correlations, but it does allow us to begin to quantify the extent to which shared associated SNPs identified through GWAS contribute to genetic correlations.

Ultimately, given the extraordinarily rich genomic data and the difficulty of creating DAGs or other simple models for the metabolic processes that affect complex traits such as leukoaraiosis and cognitive measures, bioinformatic methods and systems biology approaches are needed to integrate the vast amount of information. Systems biology approaches have already begun to be used in studies of pleiotropy in model organisms (Mackay, 2009; Magwire, 2010; Jumbo-Lucioni, 2010; Ayroles, 2009) and of genetic architecture of single traits in humans (Zhong, 2010; Hsu, 2010). Whatever the methods used to explore pleiotropy, however, it is clear large consortia will be needed for the discovery and replication of the pleiotropic SNPs identified. Though this study provides only preliminary evidence of the existence of pleiotropy among leukoaraiosis and cognitive traits, it demonstrates that further research into the pleiotropy of these traits is warranted.

Table 4.1. Outcome measures and cognitive functions assessed with neurocognitive tests

Neurocognitive Test	Outcome Measure	Cognitive Functions
Rey's Auditory Verbal Learning Test (RAVLT)	RAVLT delayed recall	Learning Delayed memory Vulnerability to interference
Rey's Auditory Verbal Learning Test (RAVLT)	RAVLT total learning	Learning Immediate memory
Digit Symbol Substitution Test (DSST)	DSST	Response speed Visual attention Concentration
Controlled Oral Word Association Test (COWA)	COWA FAS	Language Verbal fluency (phonetic association)
Controlled Oral Word Association Test (COWA)	COWA animals	Language Category fluency (semantic association)
Stroop Color Word Test	Stroop color word	Concentration effectiveness Ability to shift perceptual sets in response to novel stimuli
Stroop Color Word Test	Stroop difference	Ability to shift perceptual sets in response to novel stimuli

Table 4.2. Descriptive statistics of outcome measures and adjustment variables in whites

Trait	N	Mean	Standard Deviation	Minimum	Maximum
Age, years	762	61.27	8.84	45	84
Leukoaraiosis Volume, cm^3	714	8.11	6.83	1.16	62
Ln(Leukoaraiosis+1)	714	2.06	0.50	0.77	4.14
RAVLT delayed recall	758	9.08	3.27	0	15
RAVLT total learning	759	47.71	9.82	21	71
DSST	758	50.18	12.43	13	86
COWA FAS	760	32.39	13.63	5	85
COWA animals	762	19.25	4.86	3	33
Stroop color word	740	34.54	9.30	7	64
Stroop difference	740	32.81	9.26	5	73

Trait	Category	N	Percentage
Education	0 (Less than HS)	40	5.2%
	1 (HS/GED)	329	43.2%
	2 (Some College)	246	32.3%
	3 (Grad/Professional)	147	19.3%
Gender	Male	319	41.9%
	Female	443	58.1%

Table 4.3. Descriptive statistics of outcome measures and adjustment variables in African Americans

Trait	N	Mean	Standard Deviation	Minimum	Maximum
Age, years	720	63.29	8.22	45	91
Leukoaraiosis volume, cm^3	574	9.56	9.89	2.04	126
Ln(Leukoaraiosis+1)	574	2.16	0.55	1.11	4.85
RAVLT delayed recall	708	6.80	3.36	0	15
RAVLT total learning	712	40.07	9.37	14	66
DSST	697	32.93	13.62	3	75
COWA FAS	687	28.63	11.71	2	73
COWA animals	716	14.92	4.47	4	33
Stroop color word	648	22.28	10.10	1	55
Stroop difference	648	33.25	11.83	2	77

Trait	Category	N	Percentage
Education	0 (Less than HS)	204	28.3%
	1 (HS/GED)	205	28.5%
	2 (Some College)	127	17.6%
	3 (Grad/Professional)	184	25.6%
Gender	Male	197	27.4%
	Female	523	72.6%

Table 4.4. Comparison of the outcome measures in whites and African Americans

	Whites		African Americans		T-test comparing white and AA samples
Trait	**N**	**Mean (±SD)**	**N**	**Mean (±SD)**	**P-value**
Leukoaraiosis volume, cm^3	714	8.11 (6.83)	574	9.56 (9.89)	0.0028
Ln (leukoaraiosis+1)	714	2.06 (0.05)	574	2.16 (0.55)	0.0002
RAVLT delayed recall	758	9.08 (3.27)	708	6.80 (3.36)	<2.2E-16
RAVLT total learning	759	47.71 (9.82)	712	40.07 (9.37)	<2.2E-16
DSST	758	50.18 (12.43)	697	32.93 (13.62)	<2.2E-16
COWA FAS	760	32.39 (13.63)	687	28.63 (11.71)	1.96E-08
COWA animals	762	19.25 (4.86)	716	14.92 (4.47)	<2.2E-16
Stroop color word	740	34.54 (9.30)	648	22.28 (10.10)	<2.2E-16
Stroop difference	740	32.81 (9.26)	648	33.25 (11.83)	0.4450

Table 4.5. Multivariable linear mixed model regression with adjustment covariates as fixed effects and family as the random effect in whites

Outcome	Age β(SD)	Gender β(SD)	Education β(SD)	Total Intracranial Volume β(SD)	R^2
	0.03***	-0.07	0.003	0.0007***	0.309
Ln (leukoaraiosis+1)	(0.002)	(0.04)	(0.02)	(0.0001)	
	-0.10***	-1.96***	0.66***		0.226
RAVLT delayed recall	(0.01)	(0.20)	(0.12)	NA	
	-0.36***	-6.50***	2.25***		0.307
RAVLT total learning	(0.03)	(0.56)	(0.34)	NA	
	-0.56***	-9.47***	2.70***		0.411
DSST	(0.04)	(0.65)	(0.39)	NA	
	-0.05	-5.35***	4.38***		0.122
COWA FAS	(0.05)	(0.88)	(0.53)	NA	
	-0.12***	-0.66*	1.42***		0.144
COWA animals	(0.02)	(0.31)	(0.19)	NA	
	-0.42***	-3.21***	1.69***		0.258
Stroop color word	(0.03)	(0.56)	(0.34)	NA	
	-0.08*	-1.82**	0.21		0.018
Stroop difference	(0.04)	(0.63)	(0.38)	NA	

* p-value 0.01 – 0.05, ** p-value 0.001 – 0.01, *** p-value < 0.001
NA indicates that the outcome was not adjusted for the covariate

Table 4.6. Multivariable linear mixed model regression with adjustment covariates as fixed effects and family as the random effect in African Americans

Outcome	Age β(SD)	Gender β(SD)	Education β(SD)	Total Intracranial Volume β(SD)	R^2
	0.03***	0.10	-0.02	0.001***	0.213
Ln(leukoaraiosis+1)	(0.003)	(0.05)	(0.02)	(0.0002)	
	-0.11***	-1.80***	0.61***		0.207
RAVLT delayed recall	(0.01)	(0.25)	(0.10)	NA	
	-0.35***	-4.92***	2.35***		0.281
RAVLT total learning	(0.04)	(0.67)	(0.27)	NA	
	-0.74***	-4.69***	5.51***		0.529
DSST	(0.05)	(0.81)	(0.32)	NA	
	-0.30***	-2.10*	4.58***		0.294
COWA FAS	(0.05)	(0.86)	(0.34)	NA	
	-0.18***	0.80*	1.17***		0.260
COWA animals	(0.02)	(0.32)	(0.13)	NA	
	-0.48***	0.62	1.89***		0.239
Stroop color word	(0.04)	(0.80)	(0.32)	NA	
	-0.16**	-4.30***	1.79***		0.074
Stroop difference	(0.06)	(1.04)	(0.41)	NA	

* p-value 0.01 – 0.05, ** p-value 0.001 – 0.01, *** p-value < 0.001
NA indicates that the outcome was not adjusted for the covariate

Table 4.7. Trait heritabilities in whites

Trait	h^2 (SE) Unadjusted	h^2 (SE) Adjusted	Proportion of Variance Explained by Adjustment Covariates	Percent Variation Due to Genetic Factors After Adjustment[a]
Ln(leukoaraiosis+1)	0.656 (0.09)***	0.529 (0.09)***	0.311	36.45
RAVLT delayed recall	0.602 (0.10)***	0.526 (0.10)***	0.222	40.92
RAVLT total learning	0.627 (0.09)***	0.516 (0.10)***	0.308	35.71
DSST	0.774 (0.09)***	0.843 (0.09)***	0.403	50.33
COWA FAS	0.441 (0.10)***	0.366 (0.10)***	0.139	31.51
COWA animals	0.503 (0.10)***	0.349 (0.10)***	0.152	29.60
Stroop color word	0.586 (0.09)***	0.429 (0.09)***	0.276	31.06
Stroop difference	0.302 (0.09)***	0.275 (0.09)***	0.023	26.87

[a] [(1 – Proportion of variance explained by adjustment covariates)*h^2]*100

For all adjusted traits, biometric models included age, sex, and education. The biometric model for ln(leukoaraiosis+1) also included TIV.

Null hypothesis of tests: $h^2 = 0$
*** p-value < 0.001

Table 4.8. Trait heritabilities in African Americans

Trait	h^2 (SE) Unadjusted	h^2 (SE) Adjusted	Proportion of Variance Explained by Adjustment Covariates	Percent Variation Due to Genetic Factors After Adjustment[a]
Ln(leukoaraiosis+1)	0.485 (0.14)***	0.432 (0.13)***	0.217	33.83
RAVLT delayed recall	0.494 (0.11)***	0.390 (0.11)***	0.203	31.08
RAVLT total learning	0.560 (0.11)***	0.440 (0.11)***	0.279	31.72
DSST	0.810 (0.10)***	0.556 (0.10)***	0.525	26.41
COWA FAS	0.710 (0.11)***	0.536 (0.11)***	0.300	37.52
COWA animals	0.551 (0.10)***	0.329 (0.10)***	0.260	24.35
Stroop color word	0.532 (0.10)***	0.440 (0.10)***	0.232	33.79
Stroop difference	0.135 (0.11)***	0.154 (0.10)*	0.075	14.25

[a] [(1 – Proportion of variance explained by adjustment covariates)*h^2]*100

For all adjusted traits, biometric models included age, sex, and education. The biometric model for ln(leukoaraiosis+1) also included TIV.

Null hypothesis of tests: $h^2 = 0$
* 0.01 < p-value < 0.05
** 0.001 < p-value < 0.01
*** p-value < 0.001

Table 4.9. Pearson's correlation coefficients between adjusted and unadjusted traits in whites

Trait	Ln (leuko+1)	RAVLT delayed recall	RAVLT total learning	DSST	COWA FAS	COWA animals	Stroop color word	Stroop difference
Ln(leukoaraiosis+1)	1	-0.252***	-0.298***	-0.287***	-0.072	-0.178***	-0.235***	-0.101**
RAVLT delayed recall	-0.064	1	0.816***	0.446***	0.280***	0.372***	0.372***	0.053
RAVLT total learning	-0.088*	0.755***	1	0.530***	0.351***	0.425***	0.437***	0.146***
DSST	-0.071	0.216***	0.278***	1	0.381***	0.332***	0.594***	0.263***
COWA FAS	-0.021	0.166***	0.236***	0.266***	1	0.437***	0.340***	0.156***
COWA animals	-0.042	0.260***	0.308***	0.171***	0.363***	1	0.303***	0.131***
Stroop color word	-0.001	0.183***	0.224***	0.423***	0.258***	0.156***	1	-0.130***
Stroop difference	-0.064	-0.01	0.09*	0.226***	0.126**	0.107**	-0.228***	1

Above diagonal: Pearson's correlations, ρ, for unadjusted traits
Below diagonal: Pearson's correlations, ρ, for adjusted traits
All adjusted traits were adjusted for age, sex, and education. Ln(leukoaraiosis+1) was additionally adjusted for TIV.

Null hypothesis of tests: $\rho = 0$
* $0.01 < \text{p-value} < 0.05$
** $0.001 < \text{p-value} < 0.01$
*** $\text{p-value} < 0.001$

Table 4.10. Pearson's correlation coefficients between adjusted and unadjusted traits in African Americans

Trait	Ln (leuko+1)	RAVLT delayed recall	RAVLT total learning	DSST	COWA FAS	COWA animals	Stroop color word	Stroop difference
Ln(leukoaraiosis+1)	1	-0.202***	-0.276***	-0.259***	-0.198***	-0.168***	-0.193***	-0.094*
RAVLT delayed recall	-0.004	1	0.789***	0.415***	0.301***	0.283***	0.277***	0.148***
RAVLT total learning	-0.069	0.733***	1	0.483***	0.387***	0.358***	0.302***	0.245***
DSST	-0.065	0.192***	0.221***	1	0.546***	0.486***	0.54***	0.304***
COWA FAS	-0.071	0.138***	0.196***	0.316***	1	0.54***	0.348***	0.309***
COWA animals	-0.025	0.149***	0.189***	0.252***	0.399***	1	0.359***	0.204***
Stroop color word	-0.059	0.152***	0.134**	0.331***	0.181***	0.168***	1	-0.186***
Stroop difference	0.006	0.056	0.134**	0.192***	0.223***	0.127**	-0.337***	1

Above diagonal: Pearson's correlations, ρ, for unadjusted traits
Below diagonal: Pearson's correlations, ρ, for adjusted traits
All adjusted traits were adjusted for age, sex, and education. Ln(leukoaraiosis+1) was additionally adjusted for TIV.

Null hypothesis of tests: $\rho = 0$
* $0.01 < \text{p-value} < 0.05$
** $0.001 < \text{p-value} < 0.01$
*** $\text{p-value} < 0.001$

Table 4.11. Biometrically derived correlation coefficients between adjusted and unadjusted traits in whites

	Ln (leuko+1)	RAVLT delayed recall	RAVLT total learning	DSST	COWA FAS	COWA animals	Stroop color word	Stroop difference
Ln(leukoaraiosis+1)	0.529*** (0.09)	-0.253	-0.295	-0.294	-0.072	-0.174	-0.232	-0.110
RAVLT delayed recall	-0.083	0.526*** (0.10)	0.816	0.442	0.274	0.373	0.376	0.052
RAVLT total learning	-0.063	0.755	0.516*** (0.10)	0.523	0.346	0.429	0.441	0.147
DSST	-0.064	0.213	0.269	0.843*** (0.09)	0.376	0.334	0.600	0.268
COWA FAS	-0.023	0.163	0.234	0.265	0.366*** (0.10)	0.440	0.345	0.160
COWA animals	-0.039	0.261	0.311	0.175	0.367	0.349*** (0.10)	0.309	0.134
Stroop color word	0.003	0.186	0.226	0.424	0.262	0.160	0.429*** (0.09)	-0.134
Stroop difference	-0.071	-0.010	0.089	0.228	0.128	0.107	-0.233	0.275*** (0.09)

Above diagonal: phenotypic correlations, $_p$, for unadjusted traits
Below diagonal: phenotypic correlations, $_p$, for adjusted traits
Diagonal: heritabilities from univariate polygenic analysis, h^2 (SE), for adjusted traits
For all adjusted traits, biometric models included age, sex, and education. The biometric model for ln(leukoaraiosis+1) also included TIV.

Null hypothesis of tests: $h^2 = 0$ (diagonal)
* 0.01 < p-value < 0.05
** 0.001 < p-value < 0.01
*** p-value < 0.001

Table 4.12. Biometrically derived correlation coefficients between adjusted and unadjusted traits in African Americans

	Ln (leuko+1)	RAVLT delayed recall	RAVLT total learning	DSST	COWA FAS	COWA animals	Stroop color word	Stroop difference
Ln(leukoaraiosis+1)	0.432*** (0.13)	-2.10	-0.281	-0.270	-0.213	-0.172	-0.197	-0.108
RAVLT delayed recall	-0.007	0.390*** (0.11)	0.790	0.416	0.307	0.285	0.275	0.153
RAVLT total learning	-0.079	0.729	0.440*** (0.11)	0.492	0.401	0.358	0.308	0.257
DSST	-0.075	0.188	0.220	0.556*** (0.10)	0.557	0.489	0.534	0.311
COWA FAS	-0.081	0.136	0.020	0.298	0.536*** (0.11)	0.540	0.347	0.311
COWA animals	-0.031	0.143	0.188	0.245	0.395	0.329*** (0.10)	0.357	0.204
Stroop color word	-0.056	0.145	0.132	0.332	0.170	0.161	0.440*** (0.10)	-0.189
Stroop difference	-0.021	0.046	0.139	0.194	0.219	0.129	-0.339	0.154* (0.10)

Above diagonal: phenotypic correlations, ρ_p, for unadjusted traits
Below diagonal: phenotypic correlations, ρ_p, for adjusted traits
Diagonal: heritabilities from univariate polygenic analysis, h^2 (SE), for adjusted traits
For all adjusted traits, biometric models included age, sex, and education. The biometric model for ln(leukoaraiosis+1) also included TIV.

Null hypothesis of tests: $h^2 = 0$ (diagonal)
* 0.01 < p-value < 0.05
** 0.001 < p-value < 0.01
*** p-value < 0.001

Table 4.13. Genetic and environmental correlations in whites

	Ln (leuko+1)	RAVLT delayed recall	RAVLT total learning	DSST	COWA FAS	COWA animals	Stroop color word	Stroop difference
	0.529***	0.078	0.178	-0.008	-0.092	-0.136	-0.066	-0.274*
Ln(leukoaraiosis+1)	(0.09)	(0.13)	(0.14)	(0.22)	(0.12)	(0.12)	(0.12)	(0.11)
RAVLT delayed	-0.233	0.526***	0.586***	-0.015	-0.072	0.185	0.087	-0.024
recall	(0.14)	(0.10)	(0.08)	(0.23)	(0.13)	(0.12)	(0.13)	(0.11)
RAVLT total	-0.280*	0.918***	0.516***	0.354	0.149	0.142	0.128	0.092
learning	(0.14)	(0.05)	(0.10)	(0.21)	(0.12)	(0.12)	(0.12)	(0.11)
	-0.092	0.329**	0.263*	0.843***	0.107	0.015	0.021	0.276
DSST	(0.12)	(0.11)	(0.11)	(0.09)	(0.20)	(0.20)	(0.21)	(0.18)
	0.062	0.476**	0.350*	0.418**	0.366***	0.293*	0.163	0.185
COWA FAS	(0.16)	(0.18)	(0.16)	(0.13)	(0.10)	(0.10)	(0.11)	(0.10)
	0.084	0.372*	0.550**	0.310*	0.495*	0.349***	0.023	0.157
COWA animals	(0.17)	(0.17)	(0.15)	(0.13)	(0.16)	(0.10)	(0.11)	(0.10)
	0.078	0.296	0.336*	0.700***	0.413*	0.371*	0.429***	-0.427***
Stroop color word	(0.15)	(0.15)	(0.14)	(0.10)	(0.16)	(0.17)	(0.09)	(0.09)
	0.230	0.010	0.091	0.280	0.007	-0.004	0.146	0.275***
Stroop difference	(0.19)	(0.19)	(0.19)	(0.15)	(0.23)	(0.24)	(0.24)	(0.09)

Above diagonal: environmental correlations, ρ_e (SE)
Below diagonal: genetic correlations, ρ_g (SE)
Diagonal: heritabilities from univariate polygenic analysis, h^2 (SE), for adjusted traits
For all adjusted traits, biometric models included age, sex, and education. The biometric model for ln(leukoaraiosis+1) also included TIV.

Null hypothesis of tests: $\rho_e = 0$ (above diagonal)
Null hypothesis of tests: $\rho_g = 0$ (below diagonal)
Null hypothesis of tests: $h^2 = 0$ (diagonal)
* $0.01 <$ p-value < 0.05, ** $0.001 <$ p-value < 0.01, *** p-value < 0.001

Null hypothesis of tests: $\rho_e = 1$ (above diagonal)
Null hypothesis of tests: $\rho_g = 1$ (below diagonal)
All tests for $\rho_e = 1$ and $\rho_g = 1$ had p-value ≤ 0.05.

Table 4.14. Genetic and environmental correlations in African Americans

	Ln (leuko+1)	RAVLT delayed recall	RAVLT total learning	DSST	COWA FAS	COWA animals	Stroop color word	Stroop difference
	0.432***	-0.161	-0.261	-0.232	-0.169	-0.127	-0.194	0.130
Ln(leukoaraiosis+1)	(0.13)	(0.15)	(0.15)	(0.16)	(0.16)	(0.14)	(0.14)	(0.12)
RAVLT delayed	0.215	0.390***	0.596***	0.327*	0.167	0.166	0.040	0.035
recall	(0.23)	(0.11)	(0.08)	(0.13)	(0.14)	(0.12)	(0.13)	(0.11)
RAVLT total	0.158	0.915*** †	0.440***	0.401**	0.322*	0.128	0.017	0.035
learning	(0.22)	(0.07)	(0.11)	(0.14)	(0.14)	(0.12)	(0.14)	(0.11)
	0.078	0.036	0.045	0.556***	0.442**	0.270*	-0.005	0.237
DSST	(0.17)	(0.17)	(0.16)	(0.10)	(0.14)	(0.12)	(0.15)	(0.11)
	0.010	0.102	0.075	0.179	0.536***	0.313*	-0.019	0.113
COWA FAS	(0.19)	(0.18)	(0.17)	(0.14)	(0.11)	(0.12)	(0.15)	(0.12)
	0.124	0.101	0.286	0.228	0.533**	0.329***	0.158	0.020
COWA animals	(0.23)	(0.23)	(0.20)	(0.17)	(0.17)	(0.10)	(0.12)	(0.10)
	0.128	0.293	0.278	0.698***	0.363*	0.167	0.440***	-0.541***
Stroop color word	(0.21)	(0.19)	(0.18)	(0.14)	(0.15)	(0.20)	(0.10)	(0.09)
	-0.422†	0.010	0.432†	0.164	0.507†	0.522†	0.099†	0.154*
Stroop difference	(0.35)	(0.33)	(0.29)	(0.25)	(0.26)	(0.36)	(0.30)	(0.10)

Above diagonal: environmental correlations, ρ_e (SE)
Below diagonal: genetic correlations, ρ_g (SE)
Diagonal: heritabilities from univariate polygenic analysis, h^2 (SE), for adjusted traits
For all adjusted traits, biometric models included age, sex, and education. The biometric model for ln(leukoaraiosis+1) also included TIV.

Null hypothesis of tests: $\rho_e = 0$ (above diagonal)
Null hypothesis of tests: $\rho_g = 0$ (below diagonal)
Null hypothesis of tests: $h^2 = 0$ (diagonal)
* $0.01 <$ p-value < 0.05, ** $0.001 <$ p-value < 0.01, *** p-value < 0.001

Null hypothesis of tests: $\rho_e = 1$ (above diagonal)
Null hypothesis of tests: $\rho_g = 1$ (below diagonal)
†p-value > 0.05
All other tests for $\rho_e = 1$ and $\rho_g = 1$ had p-value ≤ 0.05.

Table 4.15. Number (percentage) of associated SNPs in single traits and number (percentage) of shared associated SNPs between traits at α=0.01 in whites

	Ln (leuko+1)	RAVLT delayed recall	RAVLT total learning	DSST	COWA FAS	COWA animals	Stroop color word	Stroop difference
Ln(leukoaraiosis+1)	26189*** (1.09%)							
RAVLT delayed recall	282 (0.56%)	24249 (1.01%)						
RAVLT total learning	300 (0.59%)	6768 (16.06%)	24673** (1.03%)					
DSST	259 (0.50%)	583 (1.19%)	858 (1.74%)	25449*** (1.06%)				
COWA FAS	276 (0.54%)	432 (0.88%)	763 (1.55%)	593 (1.18%)	25294*** (1.05%)			
COWA animals	201 (0.40%)	746 (1.56%)	931 (1.93%)	496 (1.00%)	1394 (2.88%)	24422 (1.02%)		
Stroop color word	238 (0.48%)	404 (0.85%)	516 (1.08%)	1434 (3.01%)	854 (1.78%)	384 (0.81%)	23630††† (0.98%)	
Stroop difference	375 (0.75%)	240 (0.50%)	406 (0.83%)	593 (1.20%)	442 (0.90%)	269 (0.55%)	622 (1.31%)	24417 (1.02%)

Below diagonal: observed number (percentage) of shared significant associations at α=0.01
Diagonal: observed number (percentage) of significant associations for individual traits at α=0.01.
The total number of SNPs tested for association was 2,401,820.

Null hypothesis of tests: observed number of associations at α=0.01 ≥ expected number of associations at α=0.01 given genome-wide inflation factor (diagonal)
** 0.001 < p-value < 0.01
*** p-value < 0.001

Null hypothesis of tests: observed number of associations at α=0.01 ≤ expected number of associations at α=0.01 given genome-wide inflation factor (diagonal)
†† 0.001 < p-value < 0.01
††† p-value < 0.001

Table 4.16. Number (percentage) of associated SNPs in single traits and number (percentage) of shared SNP associations between traits at α=0.01 in African Americans

	Ln (leuko+1)	RAVLT delayed recall	RAVLT total learning	DSST	COWA FAS	COWA animals	Stroop color word	Stroop difference
Ln(leukoaraiosis+1)	22243*** (1.03%)							
RAVLT delayed recall	218 (0.48%)	23747*** (1.10%)						
RAVLT total learning	281 (0.62%)	6173 (15.14%)	23211*** (1.08%)					
DSST	207 (0.49%)	376 (0.86%)	493 (1.14%)	20491††† (0.95%)				
COWA FAS	258 (0.57%)	313 (0.67%)	462 (1.01%)	842 (1.97%)	23149*** (1.08%)			
COWA animals	232 (0.55%)	362 (0.83%)	368 (0.85%)	613 (1.52%)	1307 (3.10%)	20362††† (0.95%)		
Stroop color word	169 (0.38%)	371 (0.82%)	274 (0.61%)	882 (2.13%)	536 (1.21%)	386 (0.92%)	21823 (1.02%)	
Stroop difference	136 (0.33%)	171 (0.40%)	260 (0.62%)	289 (0.74%)	533 (1.29%)	315 (0.81%)	536 (1.34%)	18792††† (0.87%)

Below diagonal: observed number (percentage) of shared significant associations at α=0.01
Diagonal: observed number (percentage) of significant associations for individual traits at α=0.01.
The total number of SNPs tested for association was 2,150,041.

Null hypothesis of tests: observed number of associations at α=0.01 ≥ expected number of associations at α=0.01 given genome-wide inflation factor (diagonal)
** 0.001 < p-value < 0.01
*** p-value < 0.001

Null hypothesis of tests: observed number of associations at α=0.01 ≤ expected number of associations at α=0.01 given genome-wide inflation factor (diagonal)
†† 0.001 < p-value < 0.01
††† p-value < 0.001

Table 4.17a. Percentage of shared SNP associations at α=0.01 from permutation testing (unpaired approach)

		Observed Data		Permuted Data				
Trait 1	**Trait 2**	**Observed Percentage of Shared SNPs**	**Absolute Genetic Correlation**	**Minimum Percentage of Shared SNPs**	**Median Percentage of Shared SNPs**	**Maximum Percentage of Shared SNPs**	**SD of Percentage of Shared SNPs**	**Empirical P-value for Observed Data**
Ln(leukoaraiosis+1)	**RAVLT total learning**	0.59	0.280	0	0.28	3.38	0.67	0.264
Ln(leukoaraiosis+1)	**DSST**	0.50	0.092	0	0.13	10.94	1.03	0.229
Ln(leukoaraiosis+1)	**COWA animals**	0.40	0.084x	0	0.18	2.78	0.45	0.278
RAVLT total learning	**DSST**	1.74	0.263	0	0.20	3.42	0.67	0.069
RAVLT total learning	**COWA animals**	1.93	0.550	0	0.21	12.20	1.24	0.056
DSST	**COWA animals**	1.00	0.310	0	0.27	3.19	0.66	0.181

Table 4.17b. Percentage of shared SNP associations at α=0.01 from permutation testing (paired approach)

		Observed Data		Permuted Data				
Trait 1	**Trait 2**	**Observed Percentage of Shared SNPs**	**Absolute Genetic Correlation**	**Minimum Percentage of Shared SNPs**	**Median Percentage of Shared SNPs**	**Maximum Percentage of Shared SNPs**	**SD of Percentage of Shared SNPs**	**Empirical P-value for Observed Data**
Ln(leukoaraiosis+1)	**RAVLT total learning**	0.59	0.280	0	0.32	5.73	0.77	0.313
Ln(leukoaraiosis+1)	**DSST**	0.50	0.092	0	0.28	7.56	1.07	0.313
Ln(leukoaraiosis+1)	**COWA animals**	0.40	0.084	0	0.21	6.80	0.86	0.326
RAVLT total learning	**DSST**	1.74	0.263	0	1.12	9.74	1.65	0.326
RAVLT total learning	**COWA animals**	1.93	0.550	0	1.38	10.92	1.72	0.417
DSST	**COWA animals**	1.00	0.310	0	0.43	5.45	0.95	0.250

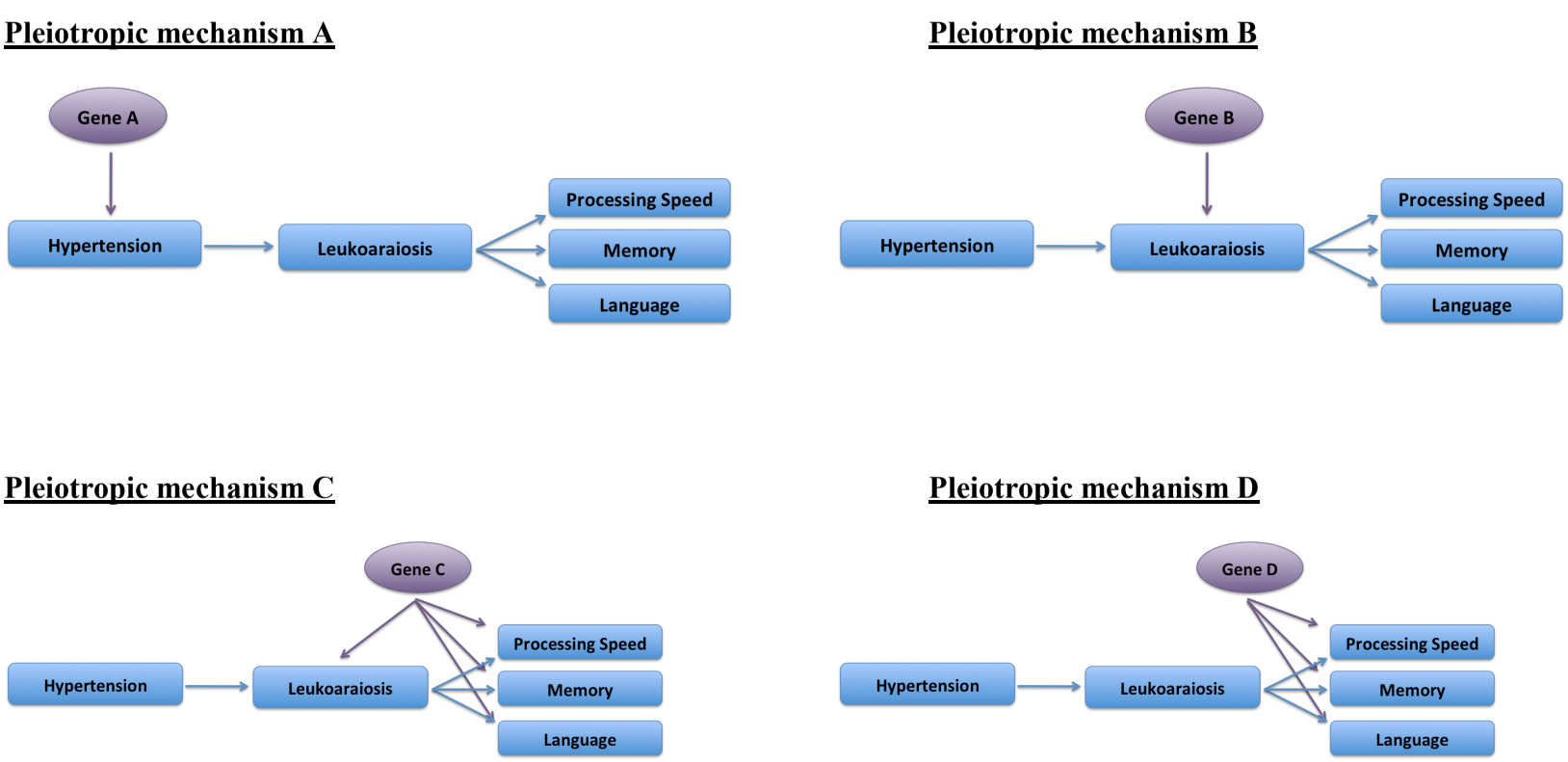
Table 4.18a. Number of shared SNP associations at α=0.01 from permutation testing (unpaired approach)

		Observed Data		Permuted Data				
Trait 1	**Trait 2**	**Observed Number of Shared SNPs**	**Absolute Genetic Correlation**	**Minimum Number of Shared SNPs**	**Median Number of Shared SNPs**	**Maximum Number of Shared SNPs**	**SD of Number of Shared SNPs**	**Empirical P-value for Observed Data**
Ln(leukoaraiosis+1)	**RAVLT total learning**	300	0.280	0	144	2258	326	0.229
Ln(leukoaraiosis+1)	**DSST**	259	0.092	0	144	3078	460	0.222
Ln(leukoaraiosis+1)	**COWA animals**	201	0.084	0	96	3891	207	0.250
RAVLT total learning	**DSST**	858	0.263	0	481	3561	332	0.056
RAVLT total learning	**COWA animals**	931	0.550	0	673	6524	627	0.056
DSST	**COWA animals**	496	0.310	0	193	2498	393	0.174

Table 4.18b. Number of shared SNP associations at α=0.01 from permutation testing (paired approach)

		Observed Data		Permuted Data				
Trait 1	**Trait 2**	**Observed Number of Shared SNPs**	**Absolute Genetic Correlation**	**Minimum Number of Shared SNPs**	**Median Number of Shared SNPs**	**Maximum Number of Shared SNPs**	**SD of Number of Shared SNPs**	**Empirical P-value for Observed Data**
Ln(leukoaraiosis+1)	**RAVLT total learning**	300	0.280	0	120	1683	320	0.292
Ln(leukoaraiosis+1)	**DSST**	259	0.092	0	48	4710	467	0.250
Ln(leukoaraiosis+1)	**COWA animals**	201	0.084	0	96	1354	447	0.285
RAVLT total learning	**DSST**	858	0.263	0	96	1634	774	0.326
RAVLT total learning	**COWA animals**	931	0.550	0	96	5861	862	0.354
DSST	**COWA animals**	496	0.310	0	144	2756	421	0.222

Figure 4.1. Pleiotropic mechanisms that may contribute to shared genetic effects among leukoaraiosis and cognitive traits

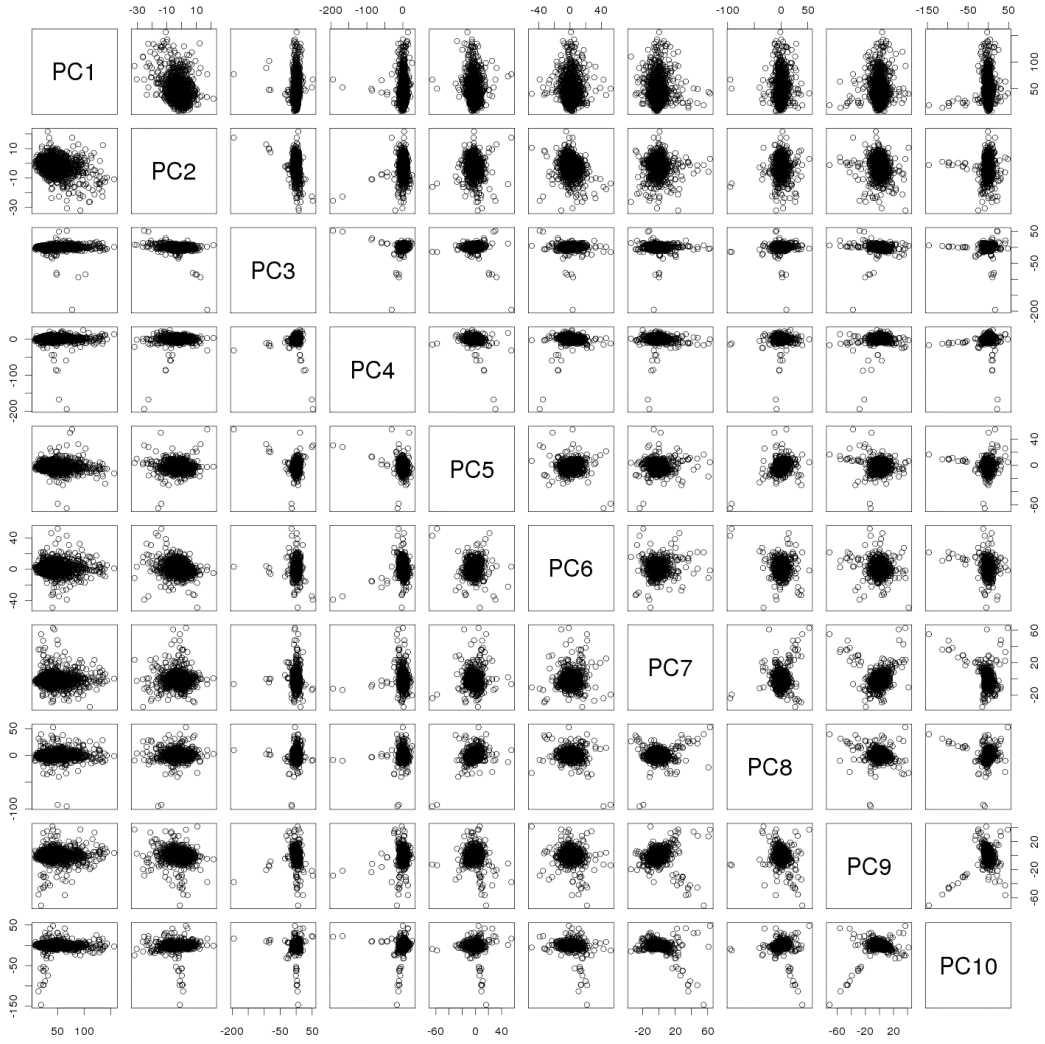


Appendix 4.1. Calculation of principal components in African Americans

Since GENOA is composed of sibships, we calculated principal components (PCs) used to control for population stratification in African Americans in an unrelated sample of individuals. First, we removed SNPs that had poor imputation quality as measured by the estimated r^2 between imputed and true genotypes (r^2<0.8) from MaCH output. Next, we obtained the maximum number of unrelated individuals in our total sample of 1624 individuals by selecting one sibling randomly from each sibship (N=644). In this sample, we calculated the first ten PCs on the set of 207,565 SNPs that were common to both genotyping platforms (Illumina 1M-Duo and Affymetrix 6.0) and were also in HapMap in order to ensure no missing values for SNPs. We then used the loading matrix for these PCs to calculate the PC values in the full sample. Next, outliers of more than 6 standard deviations on any of the ten PCs were removed to ensure that the PCs were not capturing variation due to poor quality genotyping or single individuals with a dramatically different admixture profile than the remainder of the sample. A total of 35 individuals were removed from the full sample.

Next, we again selected an unrelated sample of individuals by randomly selecting one individual from each sibship (N=638) and recalculated the first ten PCs in this sample. Finally, we used the loading matrix to calculate the first ten PCs in the final sample. A plot of these ten PCs in the full sample is shown below.

Scatterplot of the top ten principal components from genotype data in GENOA African Americans (N=1589)



Association between the top ten principal components from genotype data in GENOA African Americans and adjusted outcome measures

Trait	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Ln(leukoaraiosis+1)				***						
RAVLT delayed recall		**					*		*	
RAVLT total learning		**							*	
DSST										
COWA FAS			**						**	
COWA animals									**	
Stroop color word						*				
Stroop difference	**	*								

All traits were adjusted for age, sex, and education. Ln(leukoaraiosis+1) was additionally adjusted for TIV.

* 0.01 < p-value < 0.05
** 0.001 < p-value < 0.01
*** p-value < 0.001

Appendix 4.2. Biometrical models for estimation of heritability and genetic correlation

Inter-individual variation in many traits is the result of both genetic and environmental factors. Before undertaking research to identify specific genetic factors that contribute to a trait of interest, a typical first step is to examine how much of the observed variation in the trait is attributable to genetic factors and environmental factors. Quantitative genetic theory of polygenic inheritance provides a means of decomposing the variation in a trait to estimate heritability, which is defined as the proportion of variance in a trait due to variability in genetic factors. Heritability of a trait is a population-specific parameter that can change over time, and heritability estimates are influenced by several population and study features (Visscher, 2008). Population parameters that influence heritability estimates include allele frequencies and variation in environmental factors. Study-specific factors that influence heritability measures include the size and structure of the pedigrees, measurement error, and bias due to assortative mating and/or selection. Heritability estimates are useful for prioritizing traits with higher heritability for linkage and association studies, can provide context for interpreting the relative impact of genes on specific traits, and are a key parameter for assessing the usefulness of predicting genetic risk for disease. Estimating heritability is particularly important for providing knowledge about traits that are only measureable as a result of new technology, such as leukoaraiosis (Visscher, 2008).

The total phenotypic variance of a trait, σ^2_P, can be expressed as the sum of the underlying genotypic variance, σ^2_G, and the underlying environmental variance, σ^2_E, which includes both unmeasured environmental factors as well as any stochastic error and measurement error variances (Visscher, 2008). When estimating heritability, the underlying assumption is that there is no interaction between genetics and the environment. "Broad sense" heritability is defined as the ratio of the genetic and phenotypic variances: σ^2_G/σ^2_P. The genetic variance can be further partitioned into additive genetic effects, dominant genetic effects, and epistatic genetic effects due to interactions among genes: $\sigma^2_G = \sigma^2_A + \sigma^2_D + \sigma^2_I$. "Narrow sense" heritability is the ratio of the additive genetic effects to the total phenotypic variance: σ^2_A/σ^2_P, with the remainder of the genetic variance absorbed into the environmental variance component, σ^2_E. Narrow sense heritability is typically reported in the literature for several reasons: 1) animal and plant breeding studies have shown that the response to natural selection in fitness of organisms equals the additive genetic variance of fitness (Fisher's fundamental theorem of natural selection); thus narrow sense heritability is all that is needed to determining response to selection in breeding programs, and 2) with the exception of siblings, relatives typically share at most only one copy of alleles that are identical by descent, so non-additive effects that require sharing two copies generally do not contribute to their phenotypic resemblance (Visscher, 2008).

Genetic variance (heritability)

As described in the heritability section above, quantitative genetic modeling assumes that the total phenotypic variance of a trait is comprised of the sum of genetic variance and environmental variance.

$\sigma^2_P = \sigma^2_G + \sigma^2_E$
Where:
σ^2_P = total phenotypic variance
σ^2_G = total genetic variance
σ^2_E = total environmental variance (includes stochastic error and measurement error)
Assumption: no covariation between genetics and the environment

Broad sense heritability, H^2, is the proportion of variance due to all genetic factors.
$H^2 = \sigma^2_G/\sigma^2_P$

The genetic variance can be further partitioned into additive genetic effects, dominant genetic effects, and epistatic genetic effects due to interactions among genes:

$\sigma^2_G = \sigma^2_A + \sigma^2_D + \sigma^2_I$.
Where:
σ^2_G = total genetic variance
σ^2_A = additive genetic variance
σ^2_D = genetic variance due to dominance
σ^2_I = genetic variance due to epistatis (gene-gene interaction)

"Narrow sense" heritability, h^2, is the ratio of the additive genetic effects to the total phenotypic variance:
$h^2 = \sigma^2_A/\sigma^2_P$
When estimating h^2, the remainder of the genetic variance is absorbed into the environmental variance component, σ^2_E.

The heritability of a trait in a sample can be calculated using analysis of variance (ANOVA) method and observed phenotypic information from related individuals. Using this method, the total phenotypic variance is partitioned into variance between families (σ^2_B) and variance within families(σ^2_W). The proportion of total phenotypic variance due to within-family variance is the intra-class correlation coefficient: $r = \sigma^2_B/(\sigma^2_B+\sigma^2_W)$ (Mather, 1964). The intra-class correlation coefficient multiplied by the reciprocal of the expected amount of genetic information shared by the family members provides an estimate of heritability (Falconer, 1996). Since full siblings have approximately ½ of their genetic information identical by descent, an estimate of heritability for this kinship type is $h^2=2r$.

While the ANOVA method gives the sample statistic of heritability, a more robust procedure uses variance components to estimate heritability as a population parameter of the population from which the sample was drawn. The expected covariance of a trait between a pair of individuals is modeled as a function of the variance parameters and the expected correlation between the individuals for genetic effects (Sing, 1987). The additive genetic effects are unobservable and can only be modeled using the variance-covariance matrix of the trait, expressed as a function of identity-by-descent relationships

as expressed below. Here, shared residual (non-genetic) effects are assumed to be zero because the siblings are all adults and have reported living in separate households.

$$\mathrm{cov}(P_x, P_y) = 2\Phi_{x,y}V(G) + I_{x,y}V(R)$$

Where:
$\Phi_{x,y}$ = kinship coefficient for two individuals, *x* and *y*. The kinship coefficient is the probability that a randomly selected allele from each person at the same locus is identical-by-descent. The kinship coefficient is ½ for an individual (if *x*=*y*) and ¼ for full siblings.
$I_{x,y}$ = 1 if *x* and *y* are the same individual, 0 otherwise (identity matrix)
V(G) = deviation attributable to additive genetic factors
V(R) = deviation attributable to random residual effects (individual factors) including measurement error

In this study, SOLAR (Sequential Oligogenic Linkage Analysis Routines) (Almasy, 1998) was used to implement a variance component regression based on maximum likelihood estimation to partition phenotypic variance according to the following model:

$$y_i = \mu + \sum_{j=1}^{m} \beta_j x_{ij} + g_i + \varepsilon_i$$

Where:
$i = 1, 2, \ldots n$ (individuals)
$j = 1, 2, \ldots m$ (covariates)
y_i = outcome measure for the i^{th} individual
μ = population mean of *y*
x_{ij} = i^{th} value for the j^{th} covariate
β_j = regression coefficient associated with j^{th} covariate
g_i = additive genetic effect where $g_i \sim N(0, \sigma^2_g)$
ε_i = random residual effect where $\varepsilon_i \sim N(0, \sigma^2_e)$

The model is constrained so that $\sigma^2_g + \sigma^2_e = 1$. In this way, the heritability estimate (σ^2_g) is the heritability of the residual variance of the trait that is not accounted for by adjustment variables (the heritability of the trait after adjustment for measured covariates). Thus, the contribution of genetic effects to the total phenotypic variance when accounting for measured covariates is given by [(1-proportion of variance explained by covariates)*h^2]*100.

The heritability of the trait can be tested for significance by comparing the log-likelihood of the full model above to the log-likelihood of a model with σ^2_g constrained to 0.

$$\mathbf{H_o}: \sigma^2_g = 0, \qquad \mathbf{H_a}: \sigma^2_g \neq 0$$

The null distribution of the likelihood ratio test statistic is a 50:50 mixture of a χ^2_1 and a point mass at zero. This statistic is used to test the null hypothesis of $\sigma^2_g = 0$ (no variability is due to genetics).

Narrow sense heritabilities were estimated for the outcome variables (all cognitive traits and ln(leukoaraiosis+1)) both with and without adjustment covariates included in the biometric models.

Phenotypic, genetic, and environmental correlations

To examine the relationships among pairs of the cognitive testing and leukoaraiosis measures, two methods were used in Chapter 4 to obtain estimates of the correlation coefficient, Pearson's correlation coefficient and the estimate of phenotypic correlation using the SOLAR software package. In addition to estimating total phenotypic correlation among pairs of traits, the SOLAR software package simultaneously estimates the genetic and environmental correlations of the traits.

The correlation coefficient is defined as the covariance of two traits, X and Y, divided by the product of their variances.

$$\rho = \frac{\text{cov}(X,Y)}{\sigma_x \sigma_y}$$

Pearson's correlation

As a first pass at examining correlation, Pearson's correlation coefficient was estimated for each pair of traits, and the significance of the correlations were determined. This method of estimating correlation assumes independence among all observations, so does not take the sibship structure of the GENOA data into account; however, it provides a rough estimate as to the general relationships among the outcomes. In addition, this method gives a standard error for the point estimate of the correlation coefficient, providing the means to test for the significance of the correlation.

For a series of n measurements of X and Y written as x_i and y_i where $i = 1, 2, \ldots, n$, Pearson's correlation coefficient between trait X and trait Y is estimated as:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y}$$

Where:

$\bar{x}$ and $\bar{y}$ are the sample means of X and Y

s_x and s_y are the sample standard deviations of X and Y

r_{xy} is an estimate of the true correlation, ρ

$$\mathbf{H_o}: \rho = 0, \qquad \mathbf{H_a}: \rho \neq 0$$

Test statistic: $t = \dfrac{r}{\sqrt{\dfrac{1-r^2}{n-2}}}$

If the traits are approximately normally distributed, the null distribution of the test statistic follows a Student's t-distribution with N-2 degrees of freedom, and can be used to test the null hypothesis of $\rho = 0$ (no correlation between the traits).

Phenotypic, genetic, and environmental correlation coefficients estimated in SOLAR
The SOLAR software package uses maximum likelihood methods to simultaneously estimate heritabilities of trait pairs (as described above) and the phenotypic, genetic, and environmental correlations between pairs of traits according to the formula below (Almasy, 1998). The advantage of using SOLAR to estimate phenotypic correlation is that it uses the family relationships among the participants to perform estimations; thus, it properly accounts for the sibship structure in the GENOA data. The disadvantage is that due to the method of estimation, only a point estimate of the phenotypic correlation is given. Since the standard error for this parameter is not estimable, the significance of the phenotypic correlation cannot be tested. The phenotypic, genetic, and environmental correlations among all pairs of traits in both ethnic groups were estimated in SOLAR, both with and without adjustment covariates in the biometric models.

$$\rho_p = \rho_g \sqrt{h_1^2 h_2^2} + \rho_e \sqrt{(1 - h_1^2)(1 - h_2^2)}$$

Where:
ρ_p = phenotypic correlation between the traits
ρ_g = genetic correlation between the traits
ρ_e = environmental correlation between the traits
h_1^2 = heritability of trait 1
h_2^2 = heritability of trait 2

The genetic and environmental correlations between the traits estimated in SOLAR, ρ_g and ρ_e, can be tested for significance by comparing the log-likelihood of the model in which the parameter of interest is estimated to that of the model in which the parameter is fixed to 0. The test for pleiotropy, or evidence of shared genetic influences, is as follows:

$$\mathbf{H_o}: \rho_g = 0, \qquad \mathbf{H_a}: \rho_g \neq 0$$

The null distribution of the likelihood ratio test statistic is a χ^2_1. This statistic is used to test the null hypothesis of $\rho_g = 0$ (no genetic influences are shared between the traits). Rejection of the null hypothesis provides evidence of pleiotropy.

The genetic correlation can also be tested for evidence of complete pleiotropy (all genetic influences are shared between the pair of traits) according to the following test:

$$\mathbf{H_o}: \rho_g = 1, \qquad \mathbf{H_a}: \rho_g \neq 1$$

The null distribution of the likelihood ratio test statistic is a χ^2_1. This statistic is used to test the null hypothesis of $\rho_g = 1$ (all genetic influences are shared between the traits).

Rejection of the null hypothesis provides evidence that there are differences in the genetic influences on the traits.

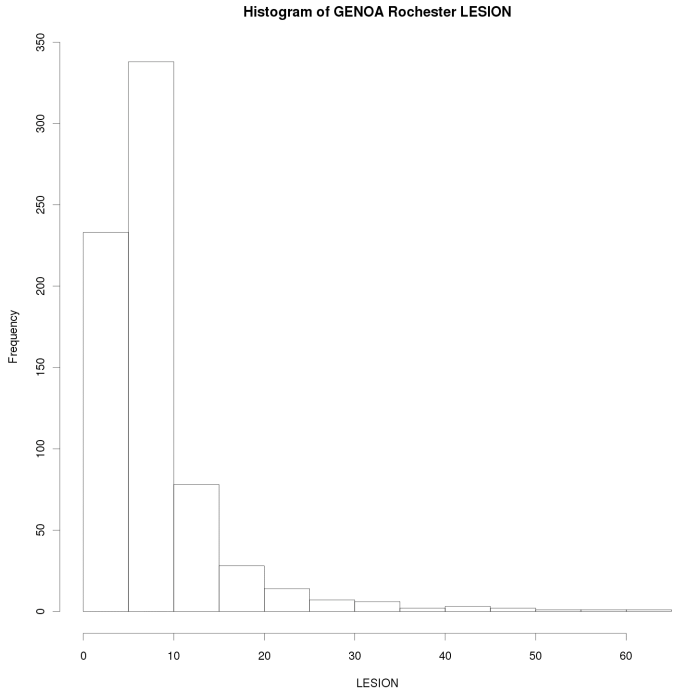
Finally, the presence of shared environmental influences can be determined by the following test:

$$\mathbf{H_o}: \rho_e = 0, \qquad \mathbf{H_a}: \rho_e \neq 0$$

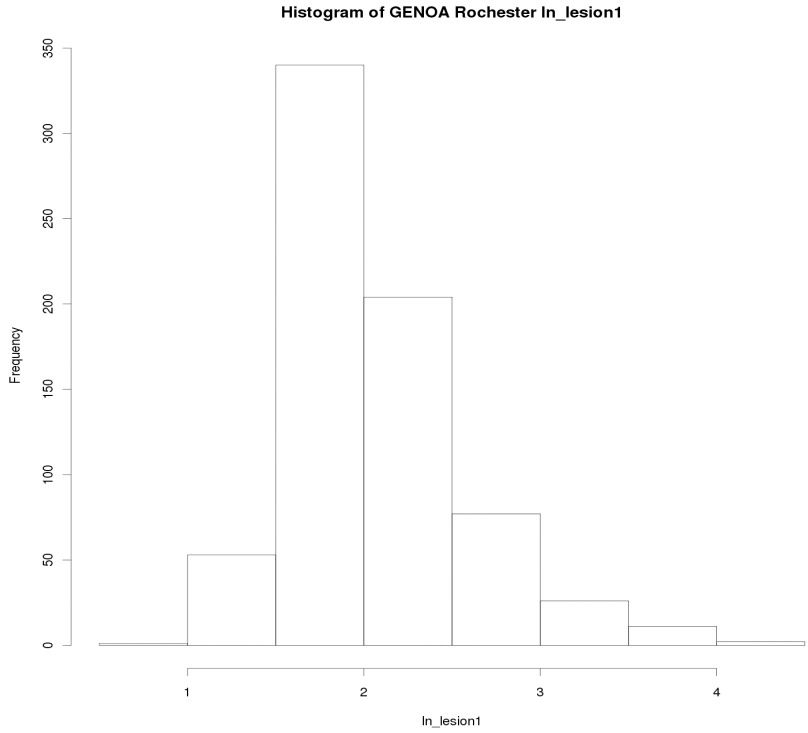
The null distribution of the likelihood ratio test statistic is a χ^2_1. This statistic is used to test the null hypothesis of $\rho_e = 0$ (no environmental influences beyond the adjustment variables are shared between the traits). Rejection of the null hypothesis provides evidence that there are shared environmental influences on the traits.

Appendix 4.3. Distributional plots of outcomes

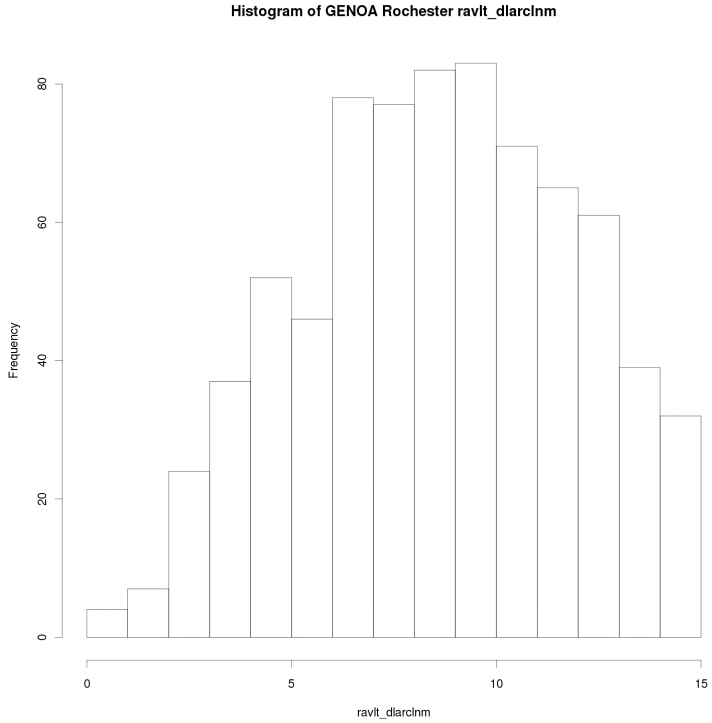
Distribution of leukoaraiosis volume (cm^3) in whites



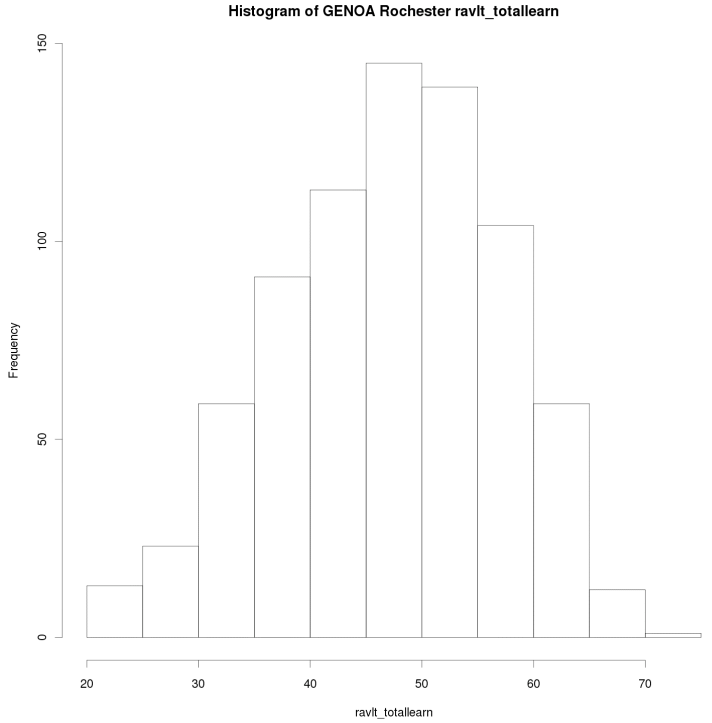
Distribution of the natural log of leukoaraiosis+1 in whites



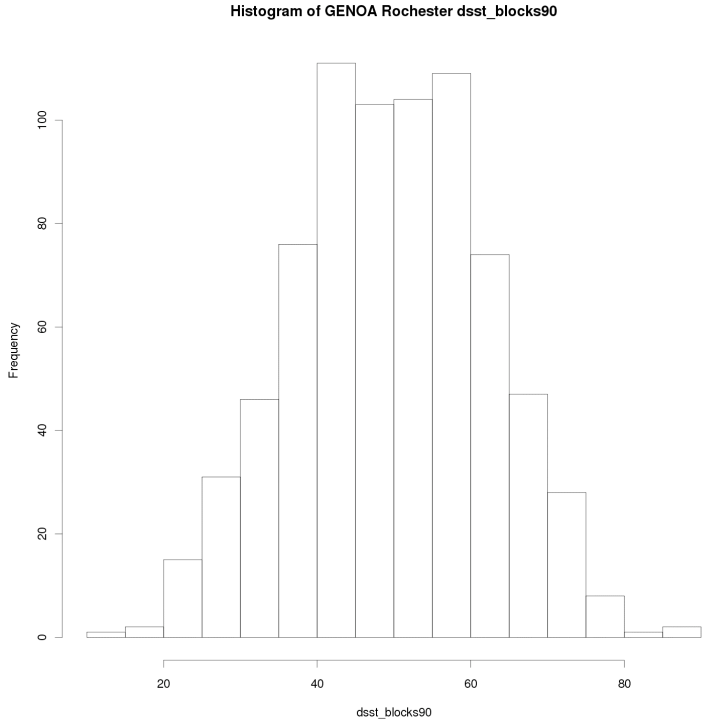
Distribution of RAVLT delayed recall in whites



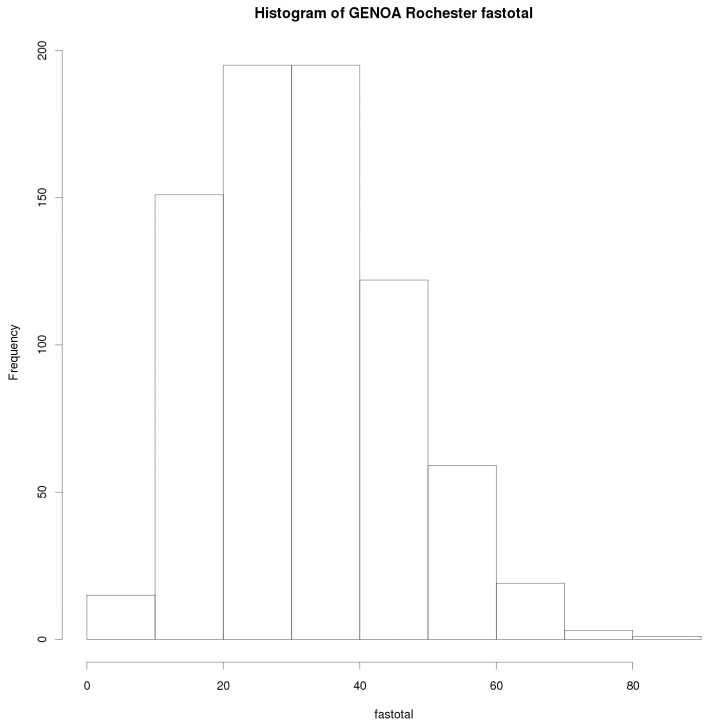
Distribution of RAVLT total learning in whites



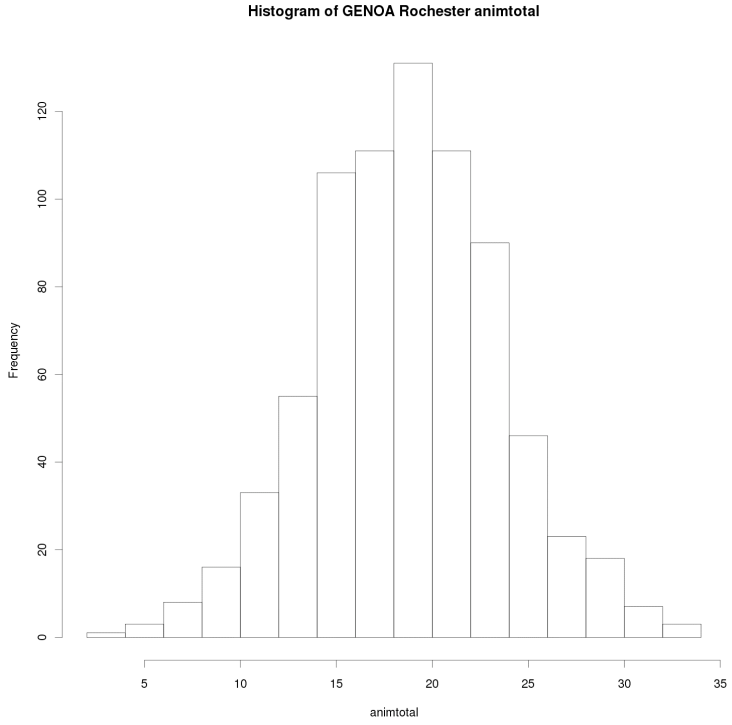
Distribution of DSST in whites



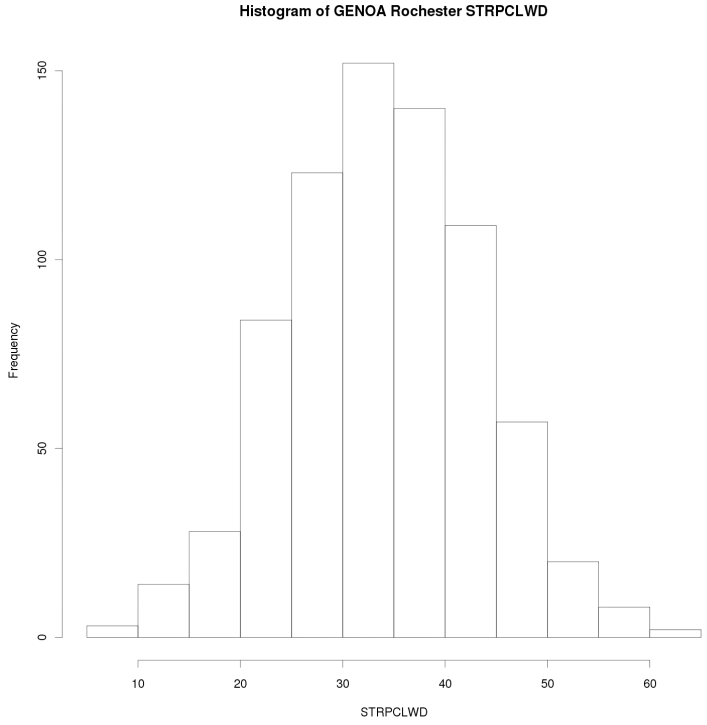
Distribution of COWA FAS in whites



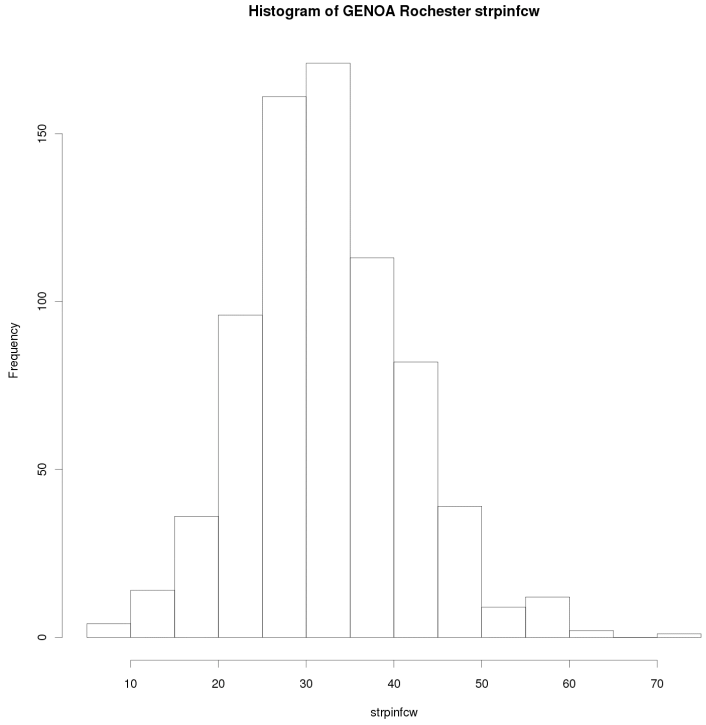
Distribution of COWA animals in whites



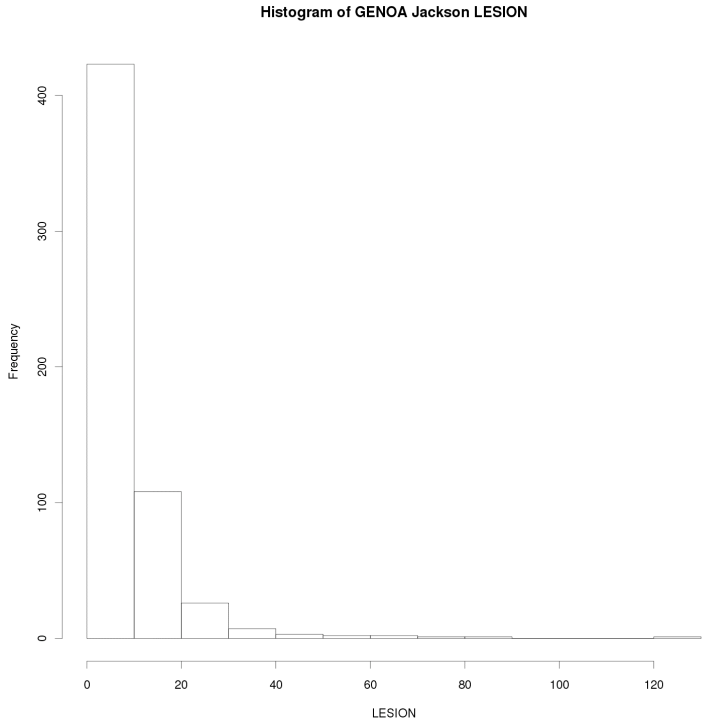
Distribution of Stroop color word in whites



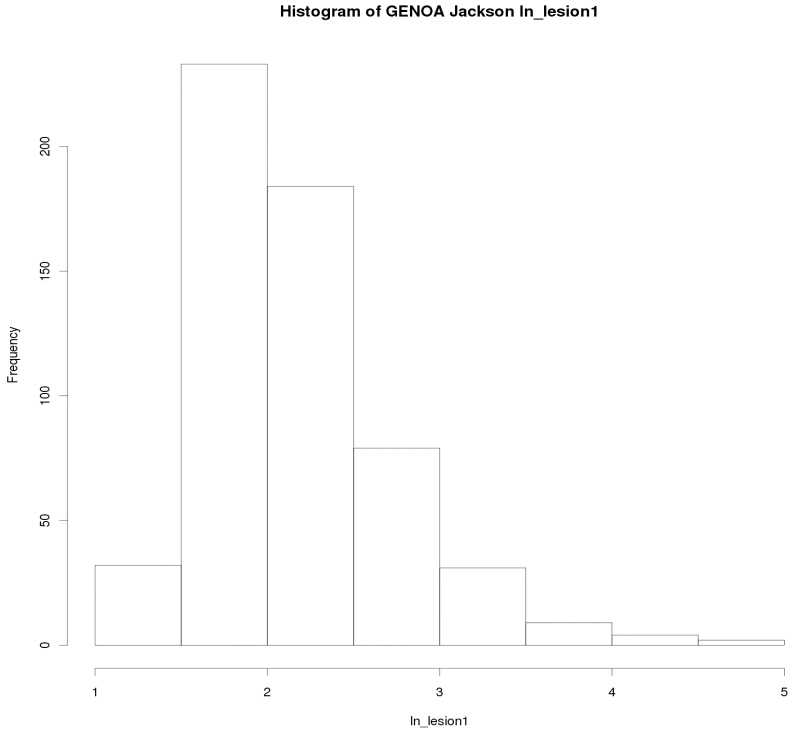
Distribution of Stroop difference in whites



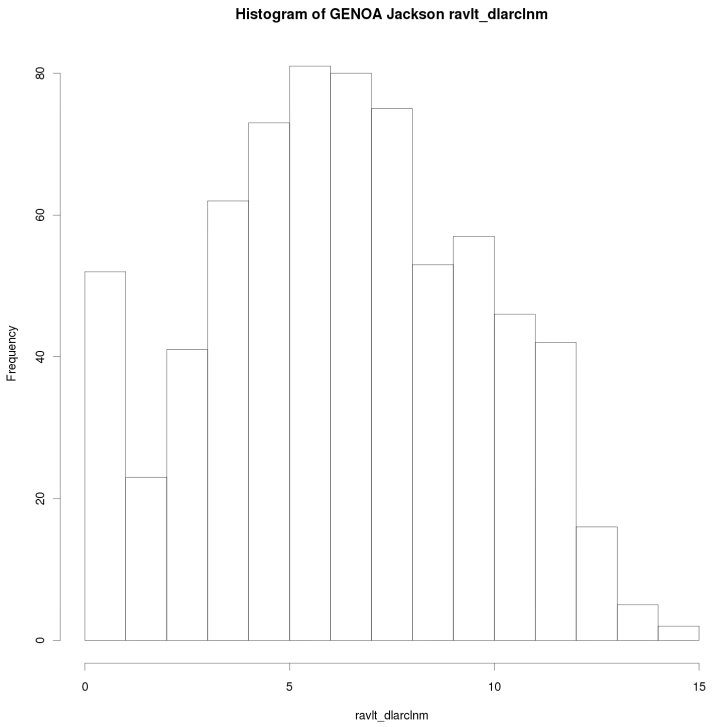
Distribution of leukoaraiosis volume (cm^3) in African Americans



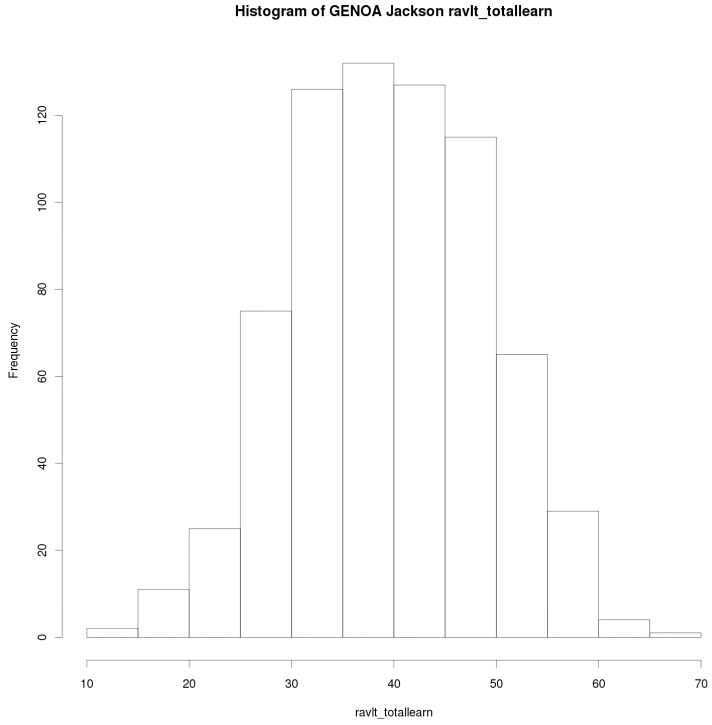
Distribution of the natural log of leukoaraiosis+1 in African Americans



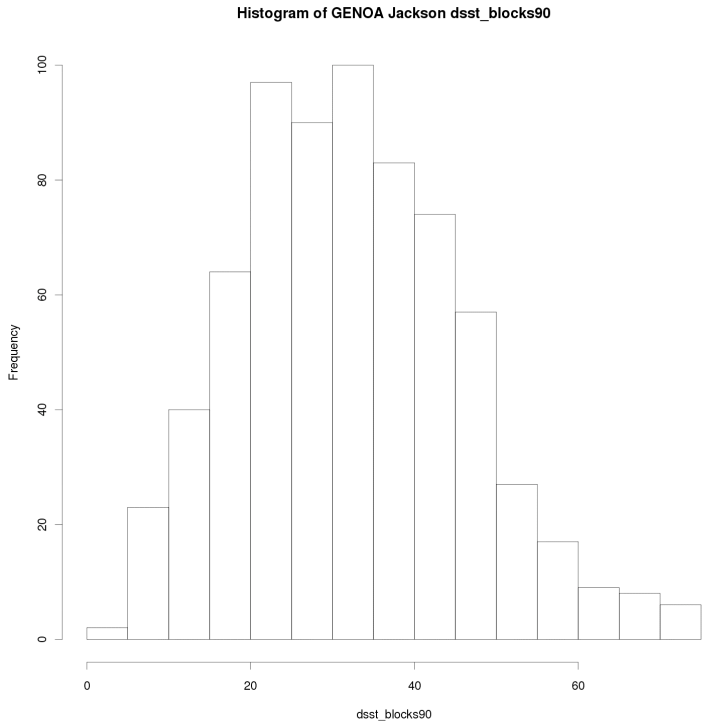
Distribution of RAVLT delayed recall in African Americans



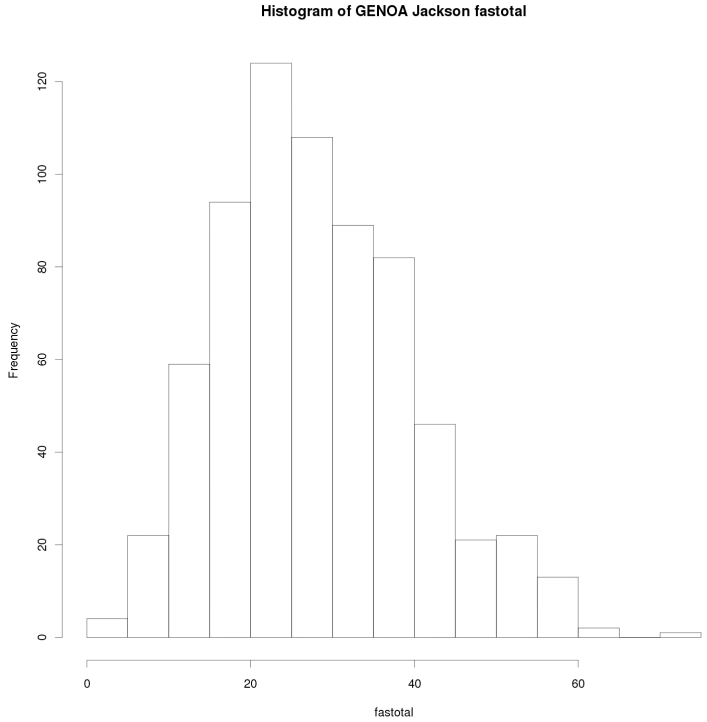
Distribution of RAVLT total learning in African Americans



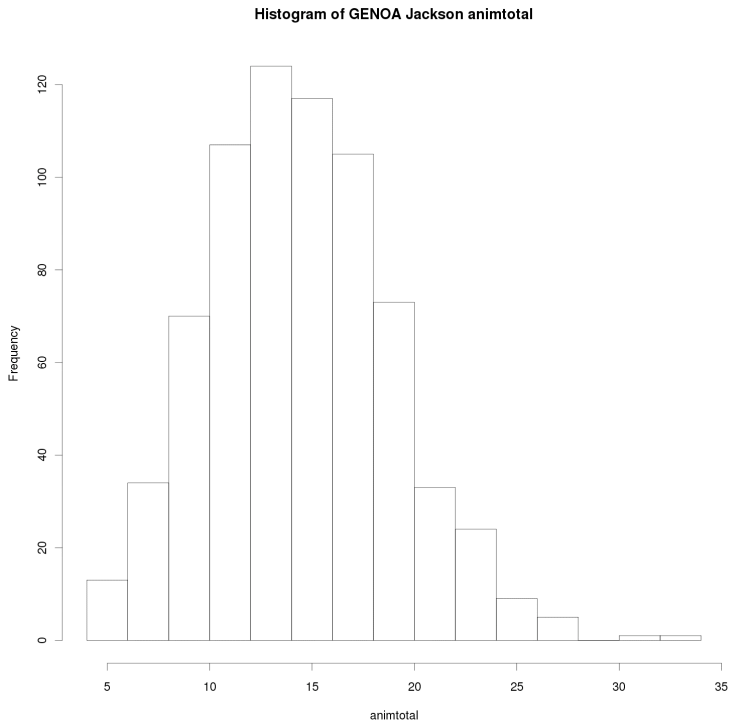
Distribution of DSST in African Americans



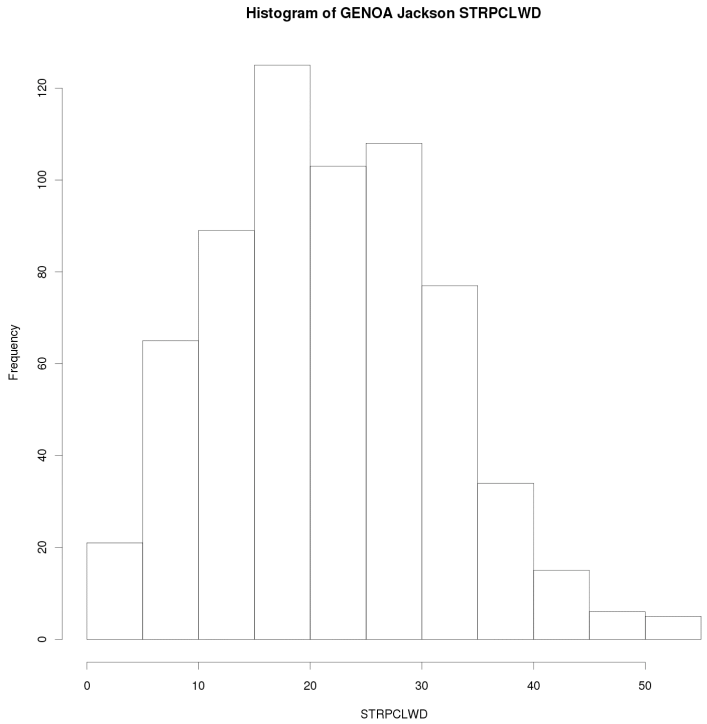
Distribution of COWA FAS in African Americans



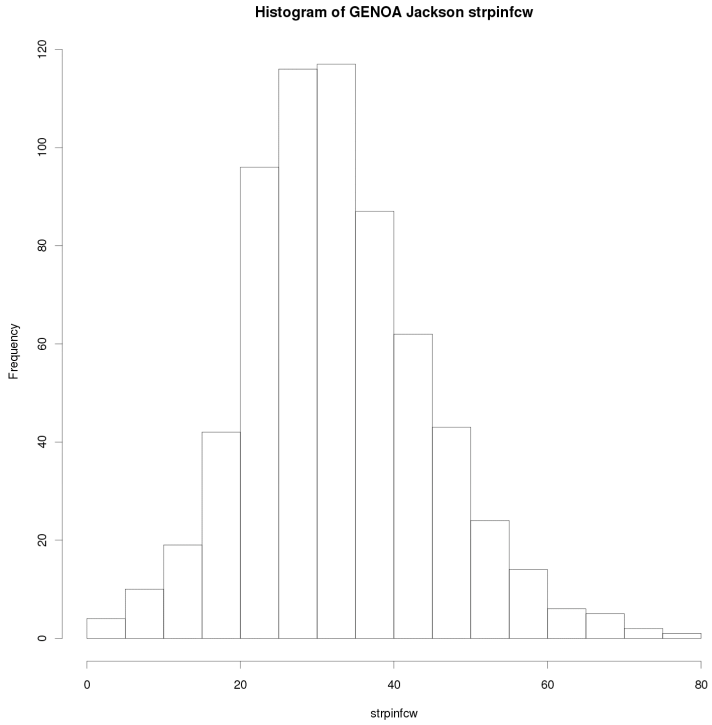
Distribution of COWA animals in African Americans



Distribution of Stroop color word in African Americans

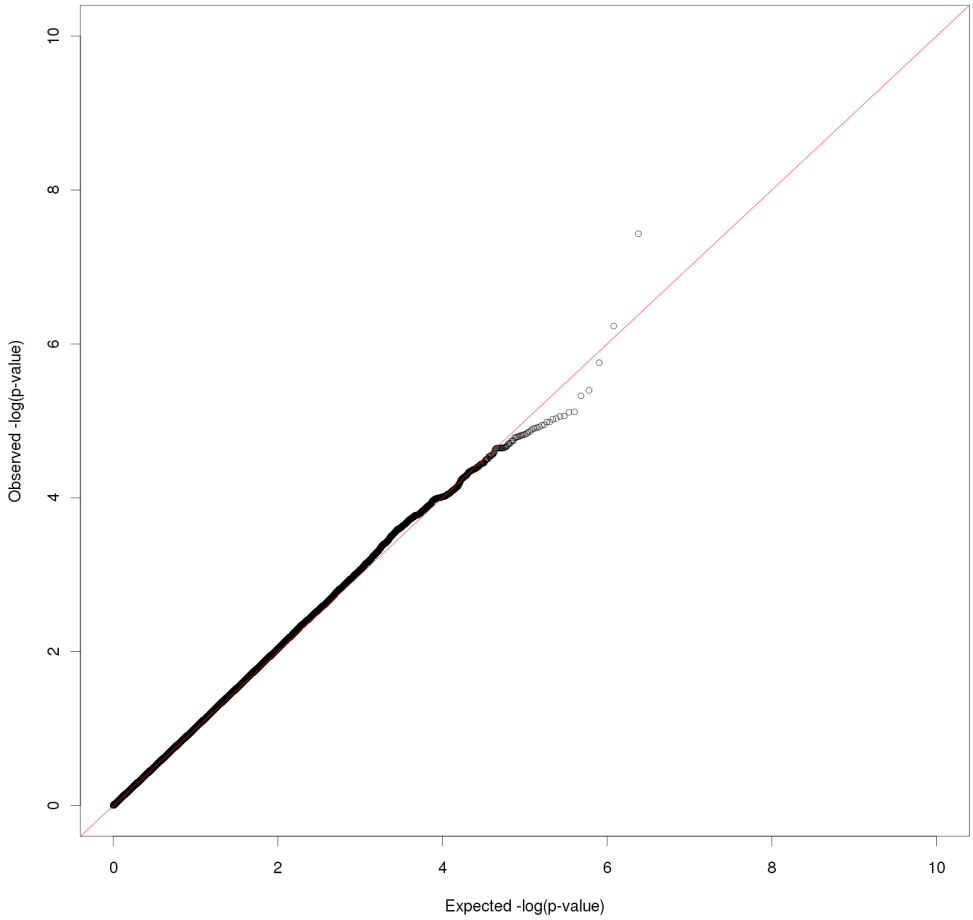


Distribution of Stroop difference in African Americans

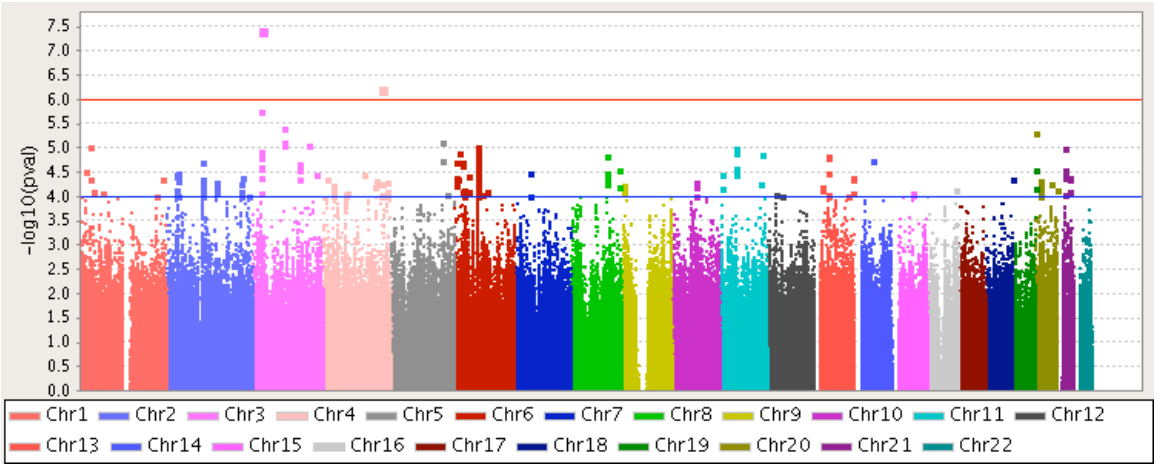


Appendix 4.4. Quantile-quantile plots, Manhattan plots, and genome-wide inflation factors for GWAS

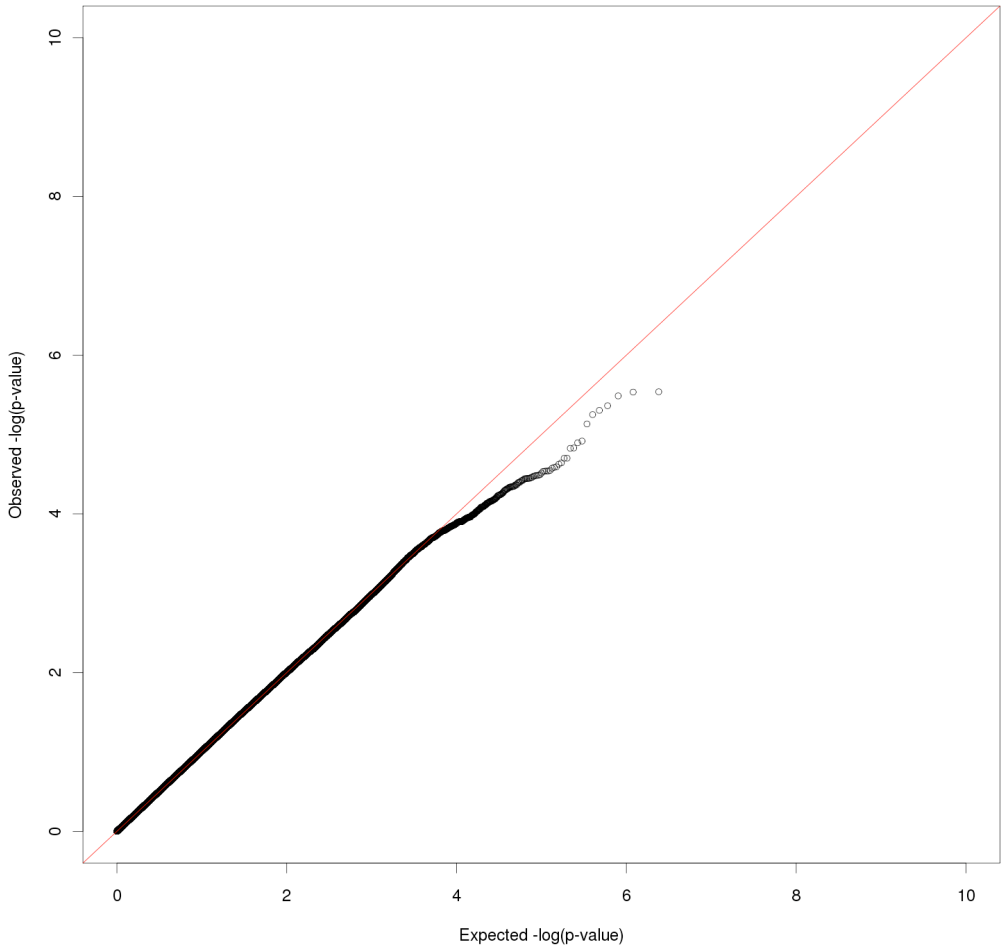
QQ plot for GWAS of Ln (leukoaraiosis+1) in whites



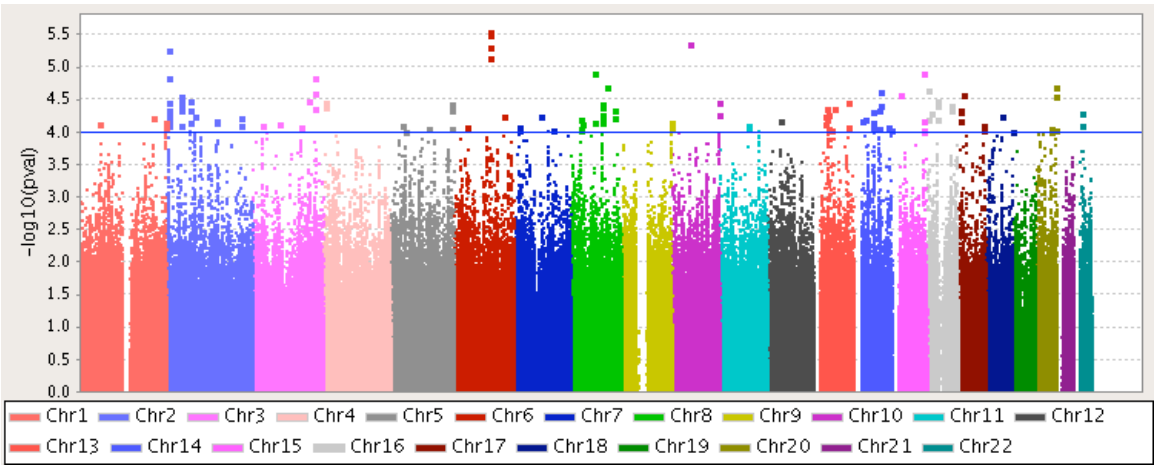
Manhattan plot for GWAS of Ln (Leukoaraiosis+1) in whites



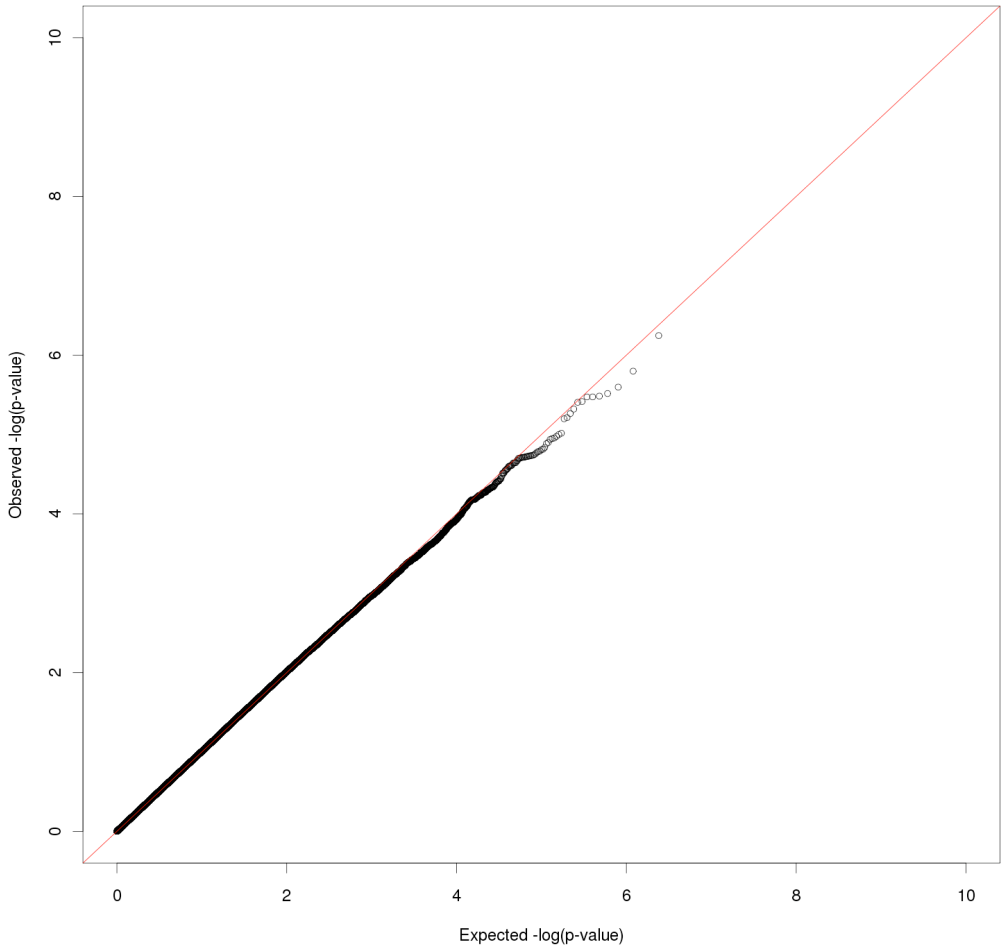
QQ plot for GWAS of RAVLT delayed recall in whites



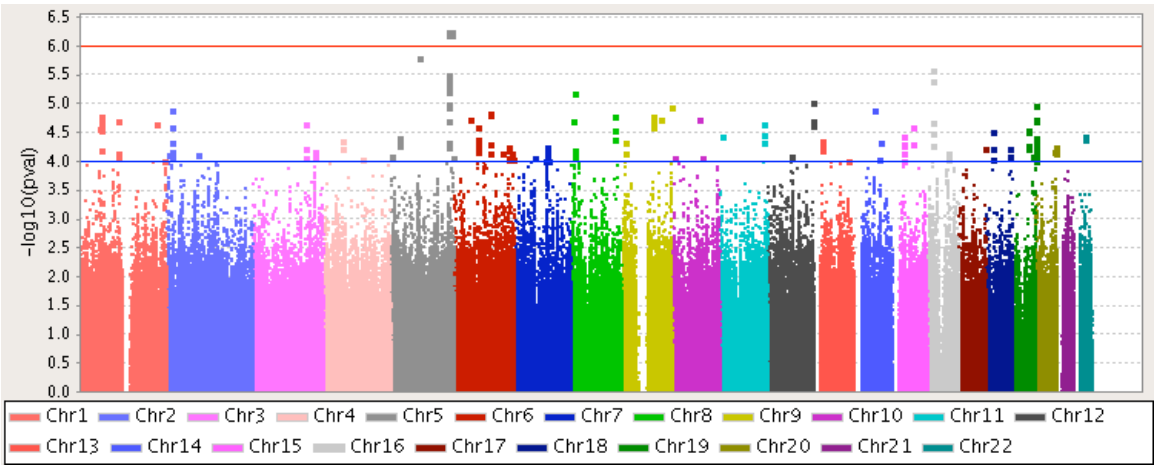
Manhattan plot for GWAS of RAVLT delayed recall in whites



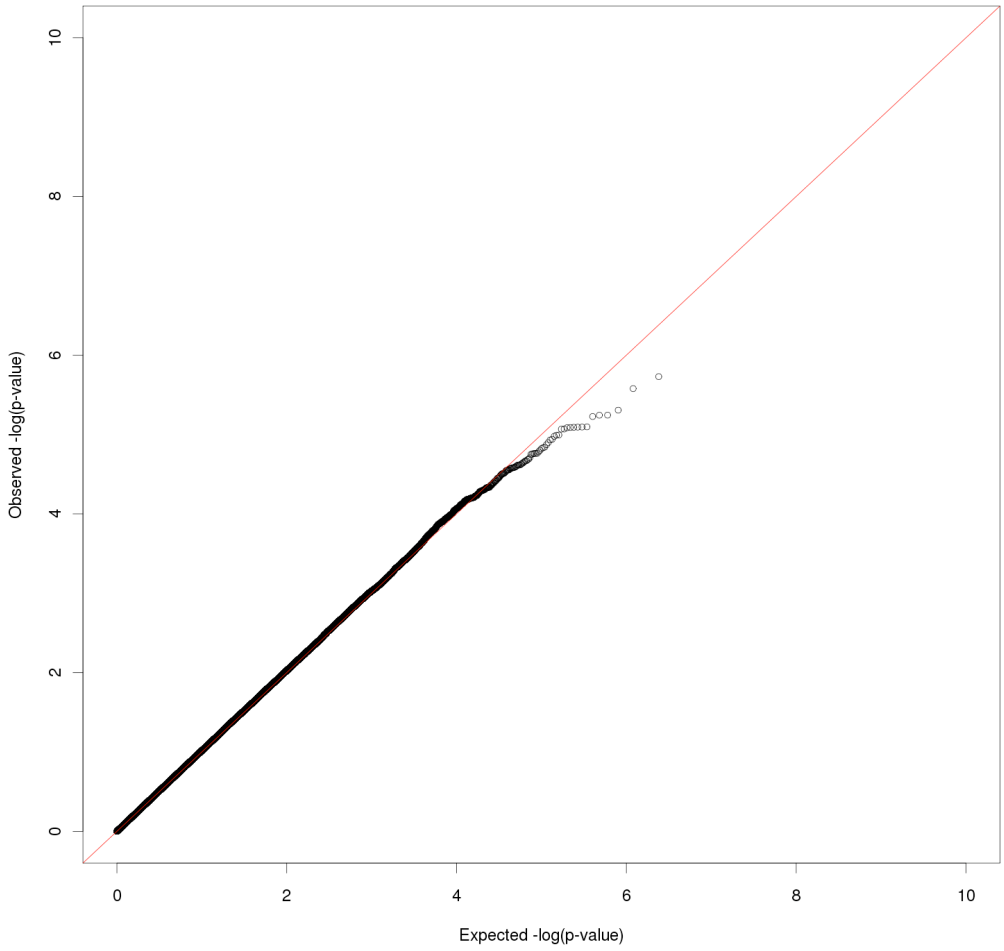
QQ plot for GWAS of RAVLT total learning in whites



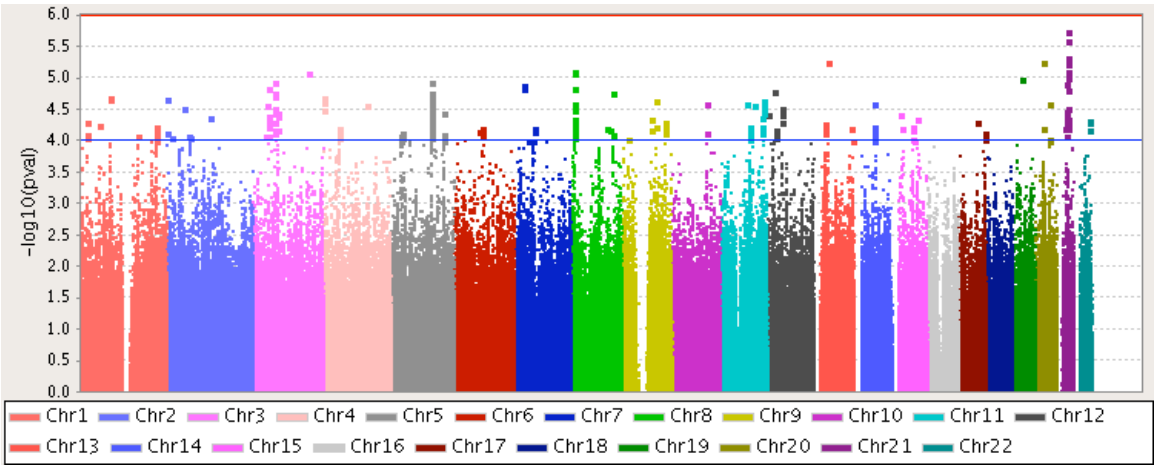
Manhattan plot for GWAS of RAVLT total learning in whites



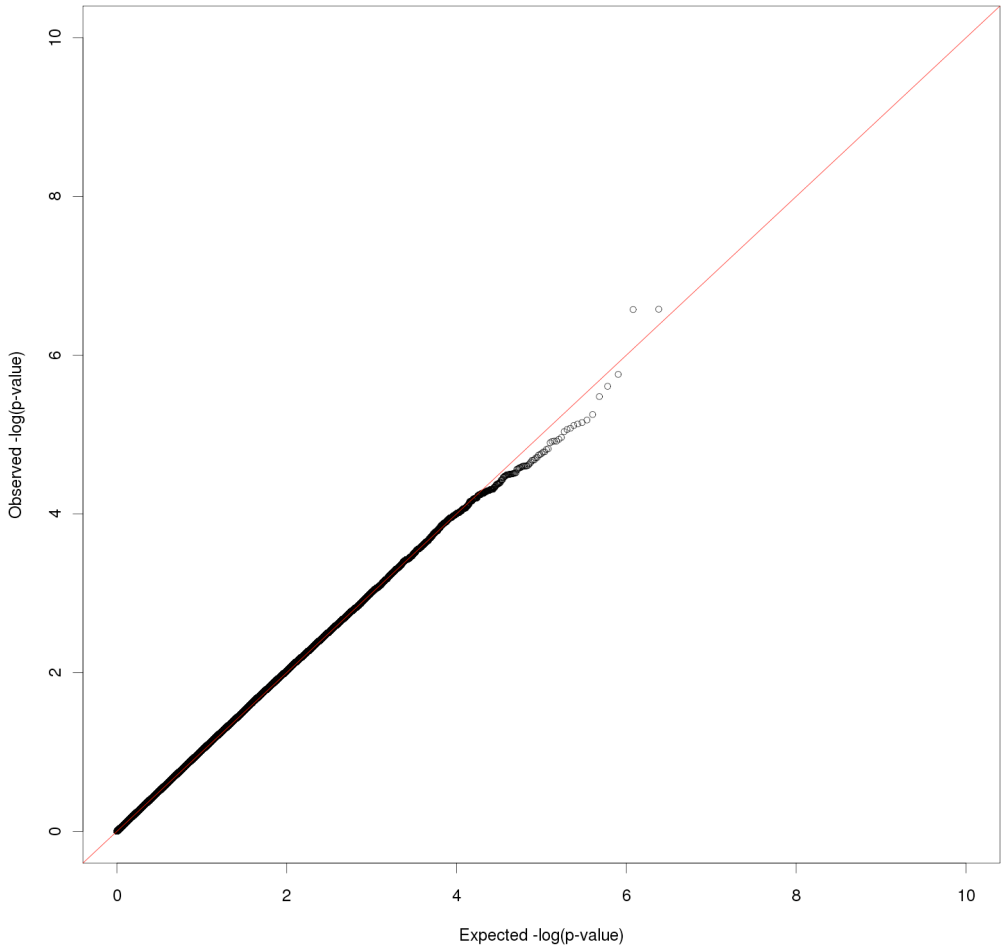
QQ plot for GWAS of DSST in whites



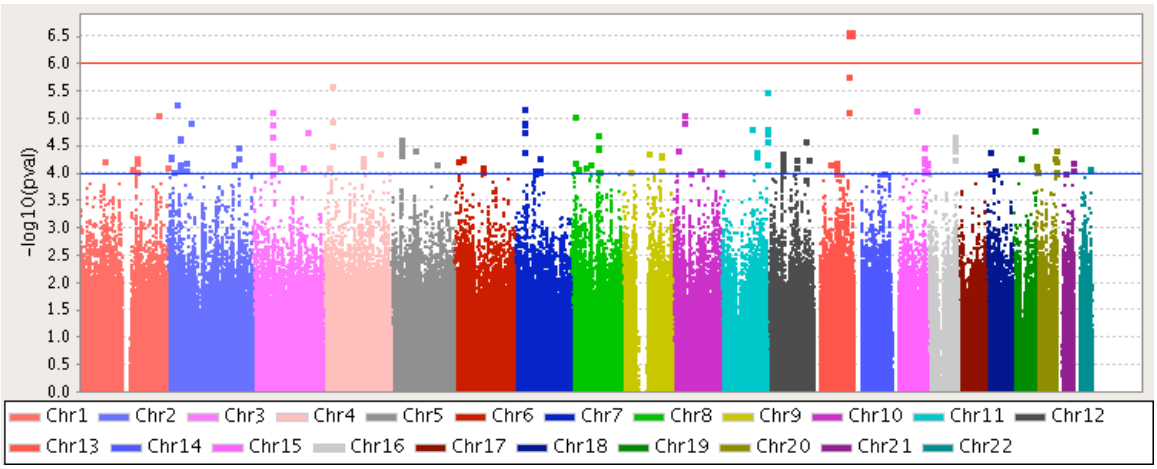
Manhattan plot for GWAS of DSST in whites



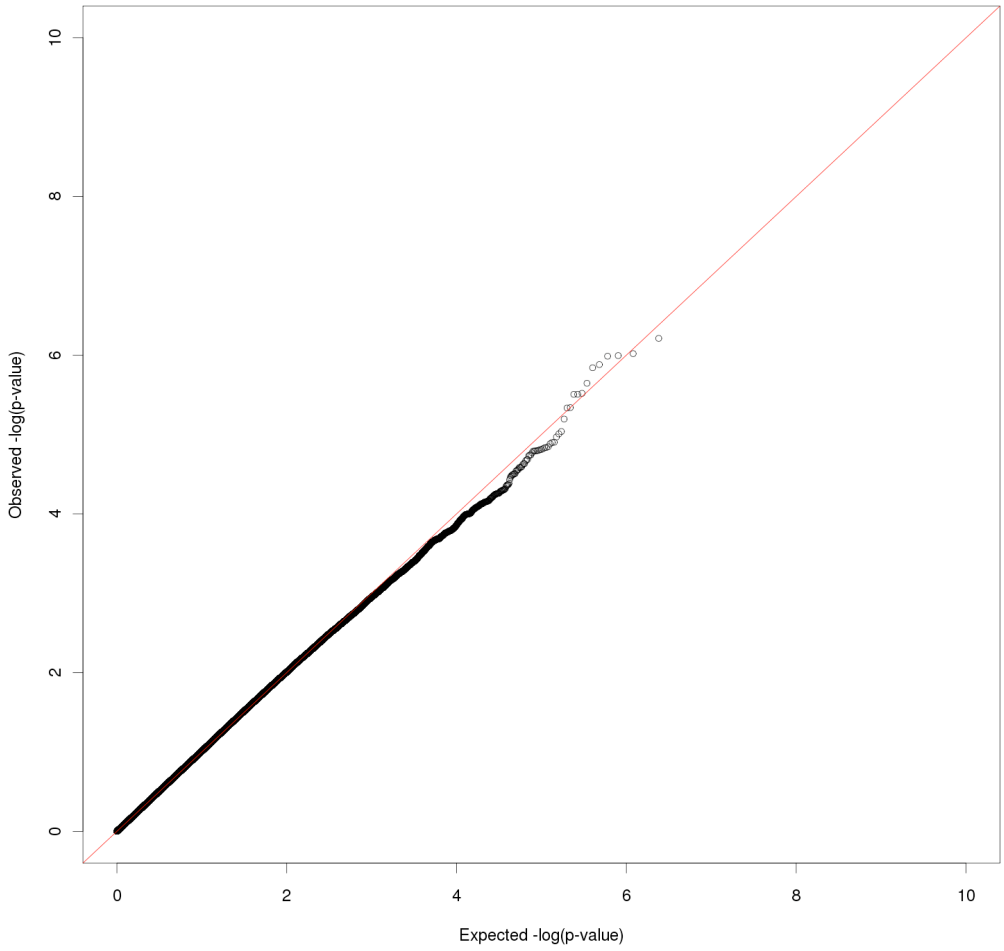
QQ plot for GWAS of COWA FAS in whites



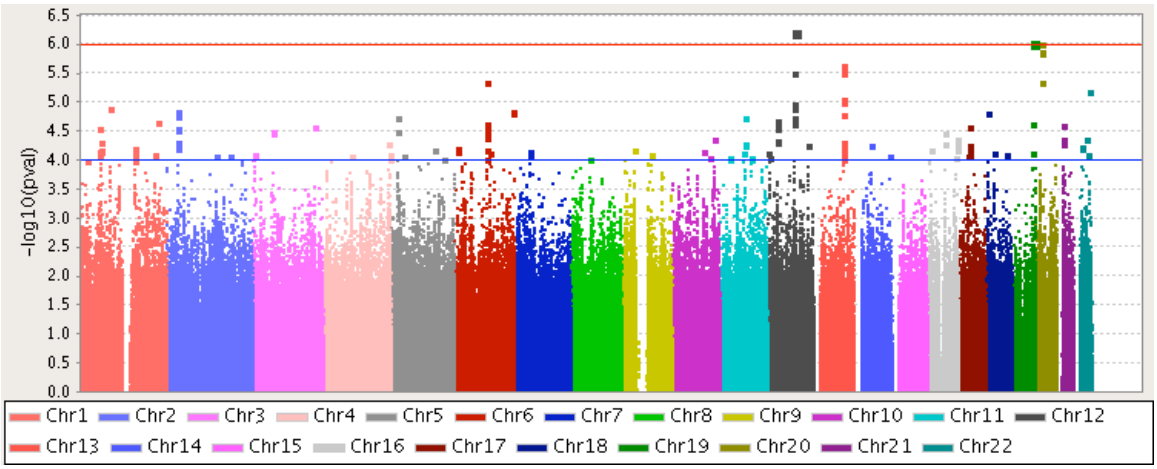
Manhattan plot for GWAS of COWA FAS in whites



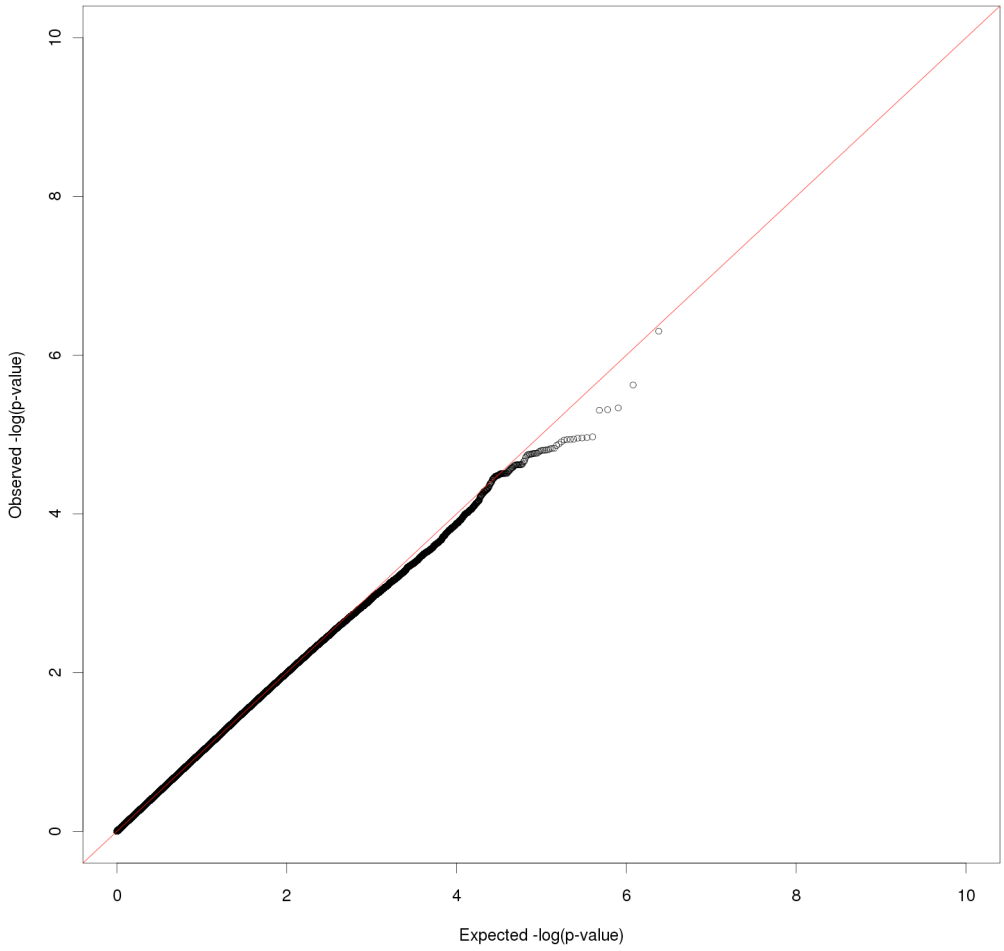
QQ plot for GWAS of COWA animals in whites



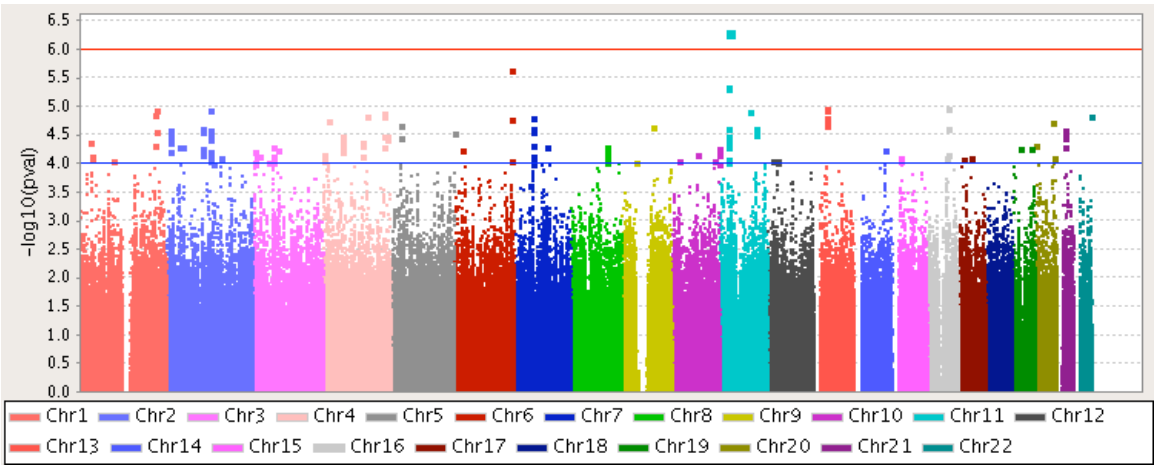
Manhattan plot for GWAS of COWA animals in whites



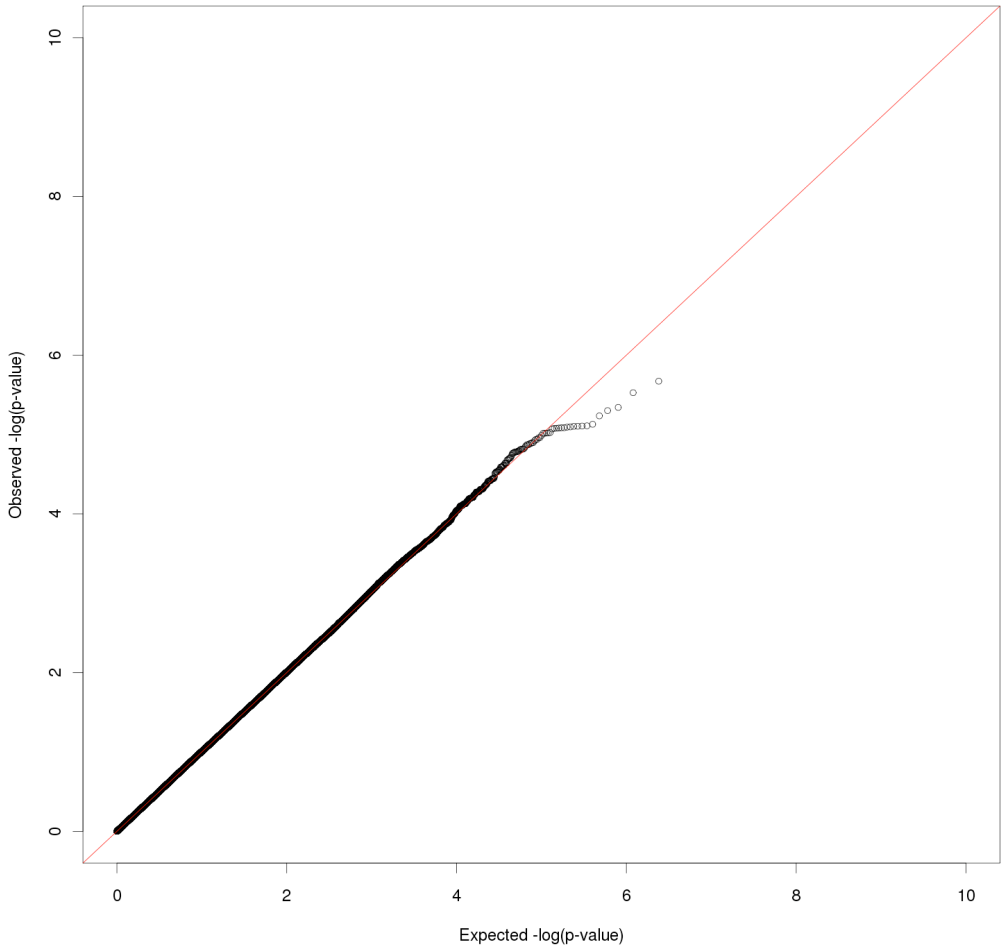
QQ plot for GWAS of Stroop color word in whites



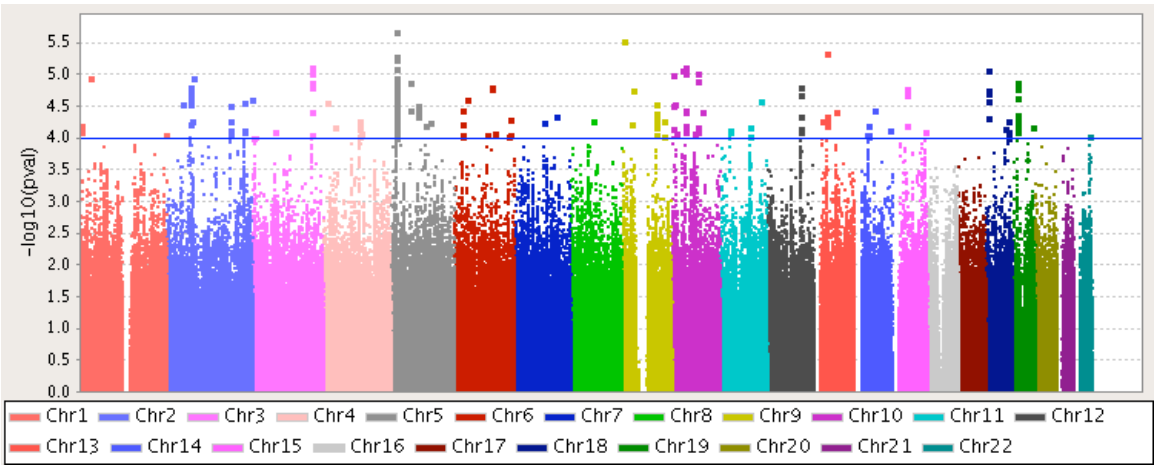
Manhattan plot for GWAS of Stroop color word in whites



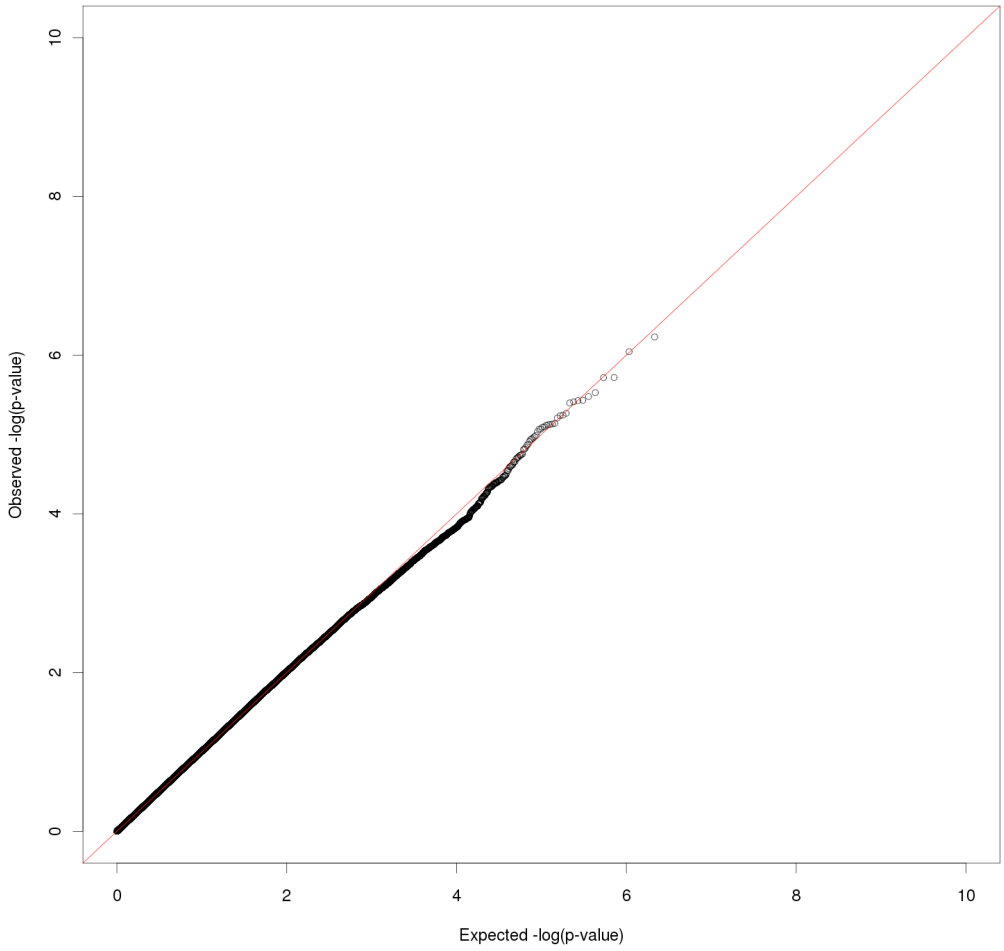
QQ plot for GWAS of Stroop difference in whites



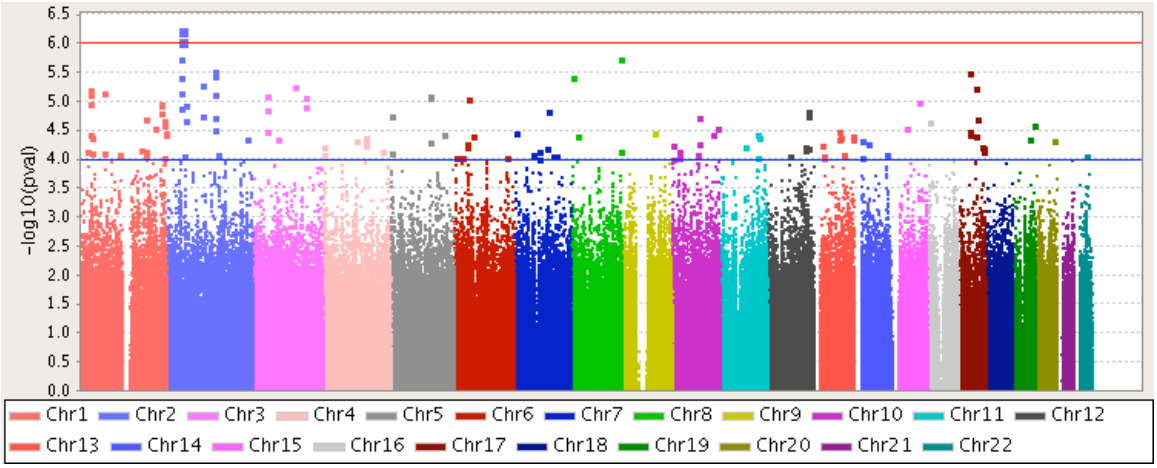
Manhattan plot for GWAS of Stroop difference in whites



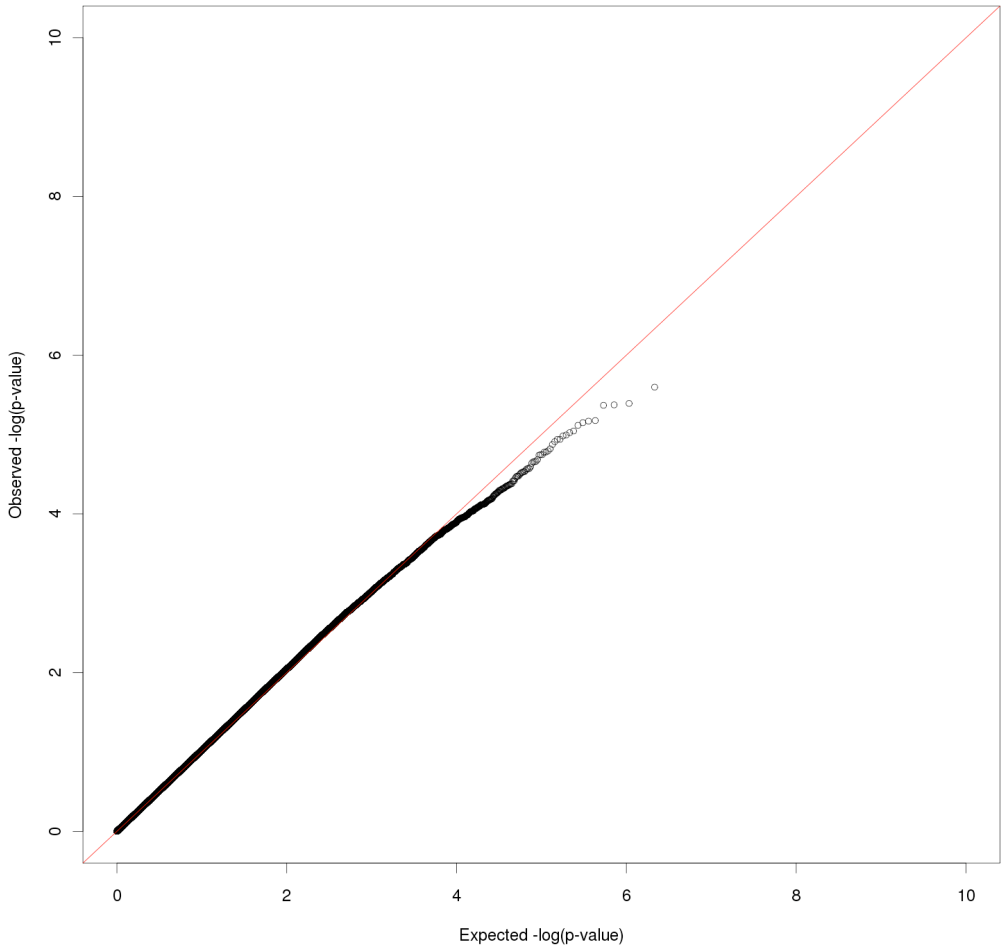
QQ plot for GWAS of Ln (leukoaraiosis+1) in African Americans



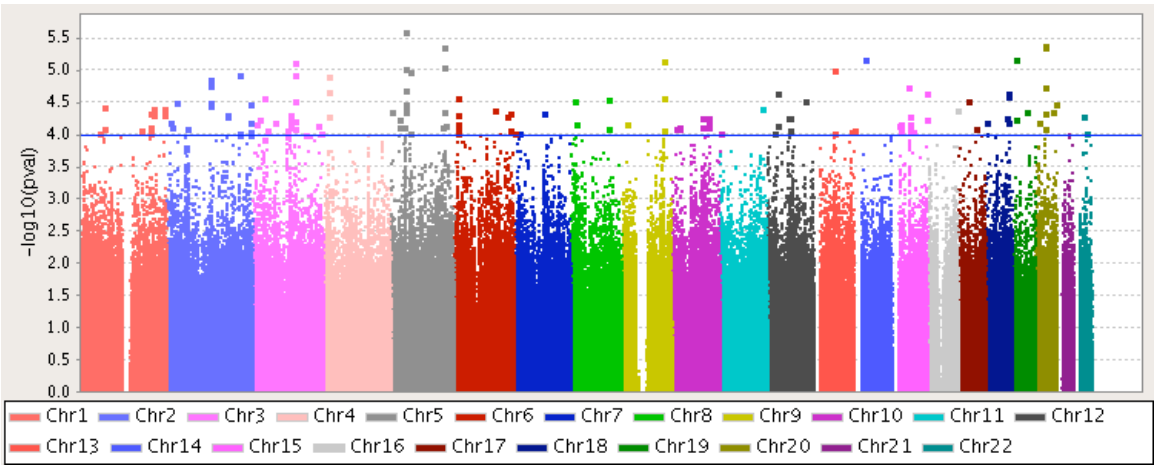
Manhattan plot for GWAS of Ln (leukoaraiosis+1) in African Americans



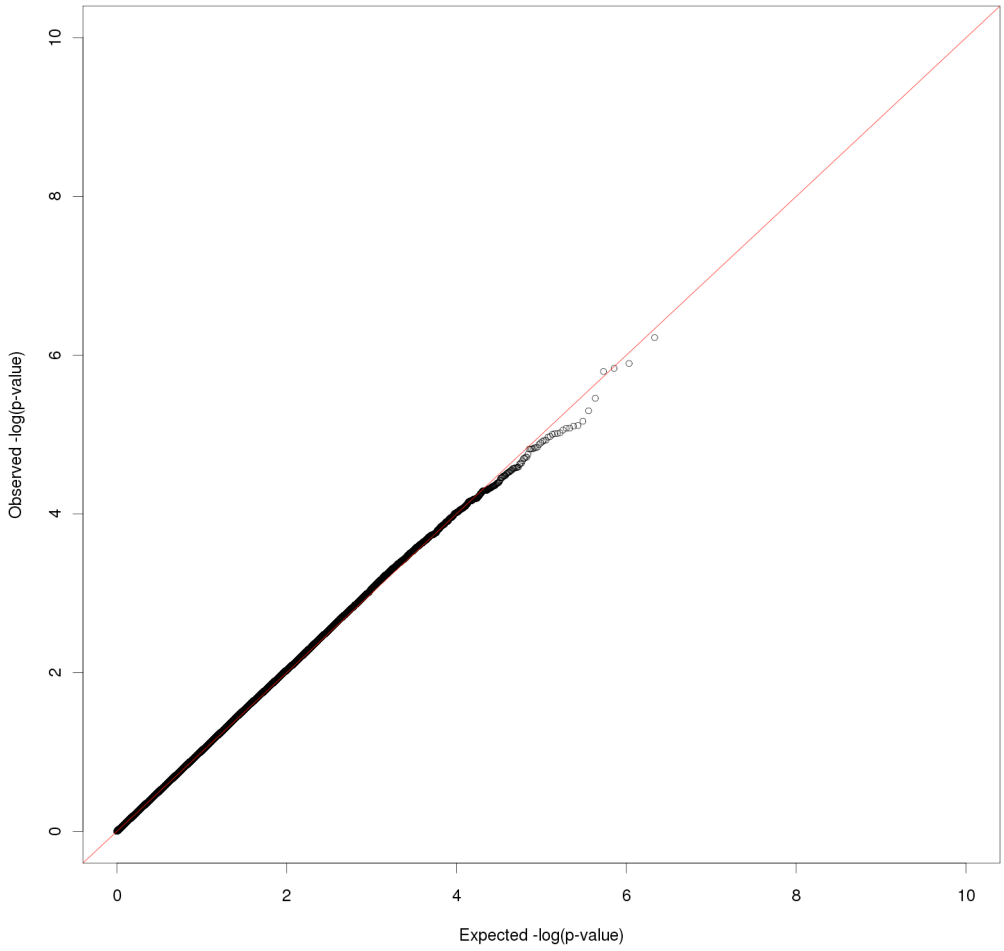
QQ plot for GWAS of RAVLT delayed recall in African Americans



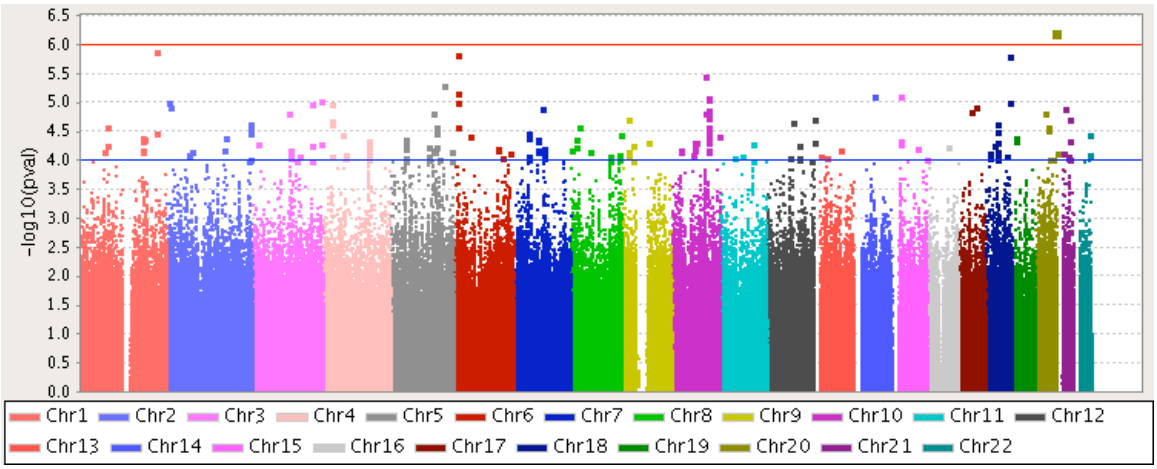
Manhattan plot for GWAS of RAVLT delayed recall in African Americans



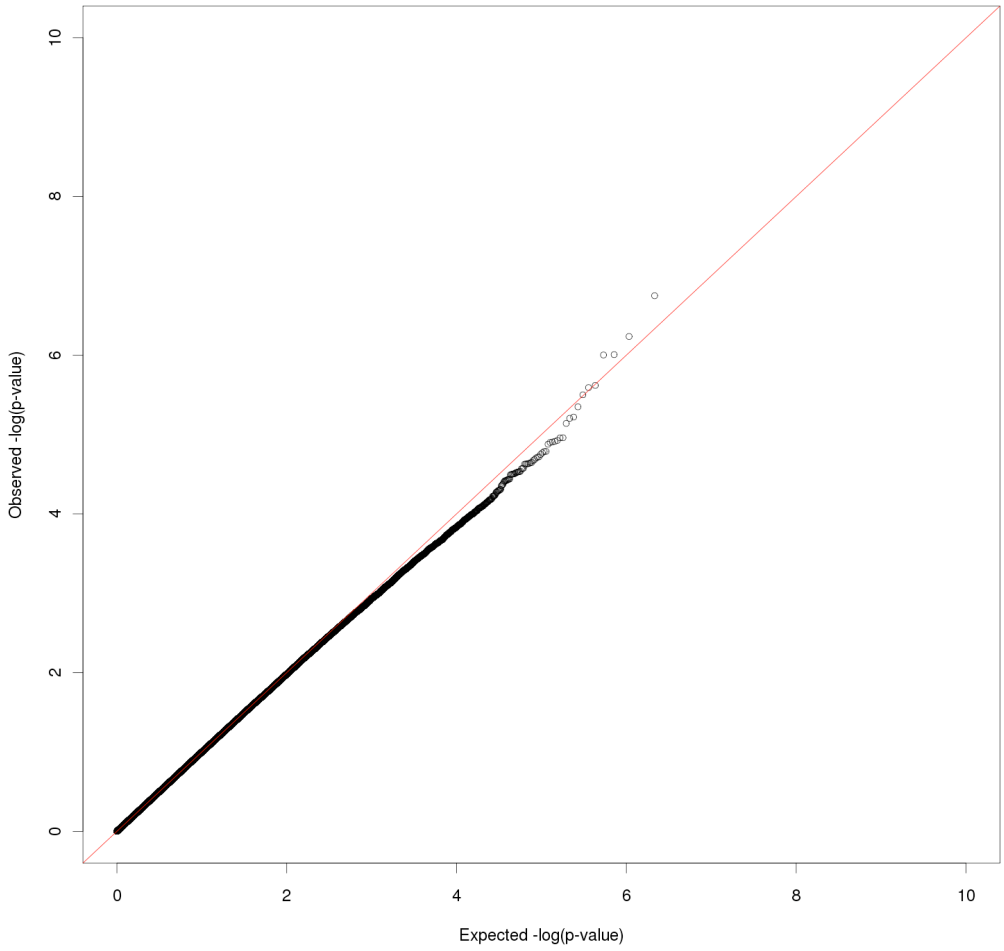
QQ plot for GWAS of RAVLT total learning in African Americans



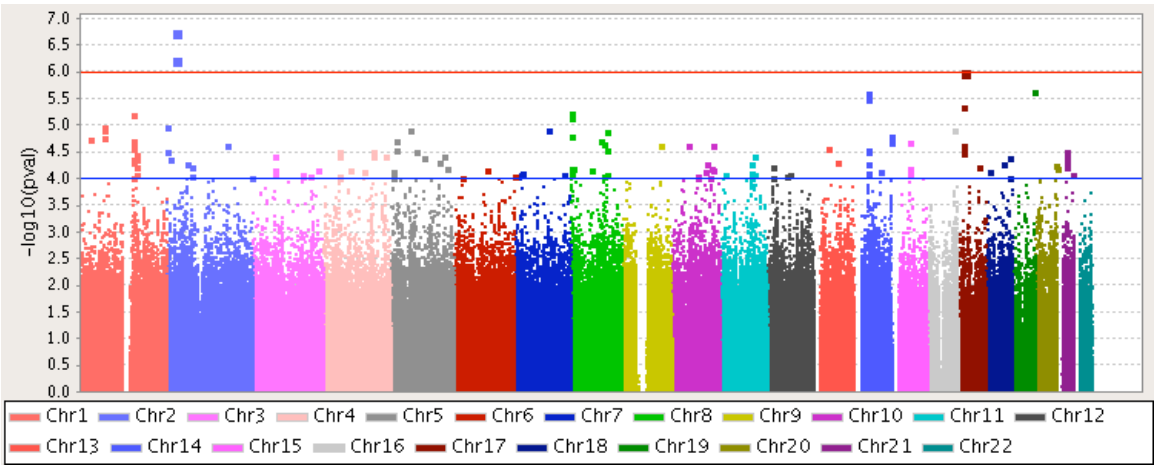
Manhattan plot for GWAS of RAVLT total learning in African Americans



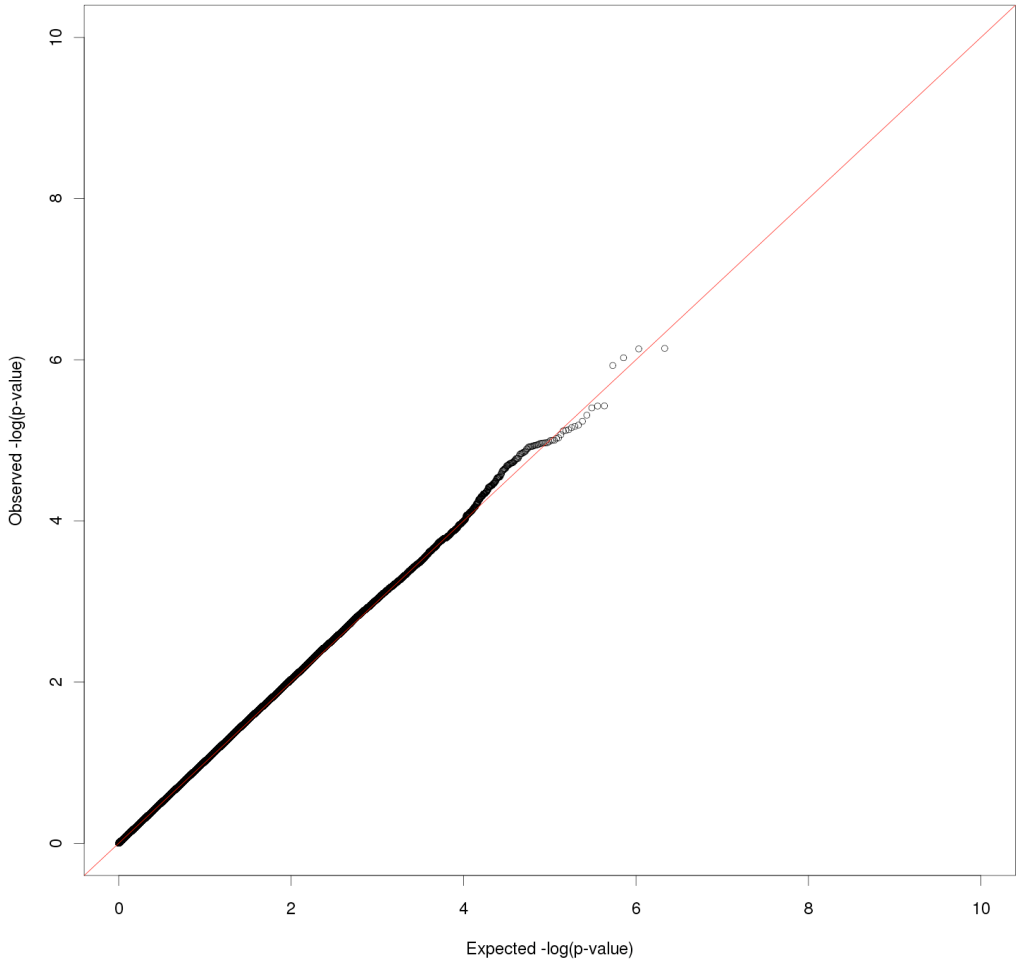
QQ plot for GWAS of DSST in African Americans



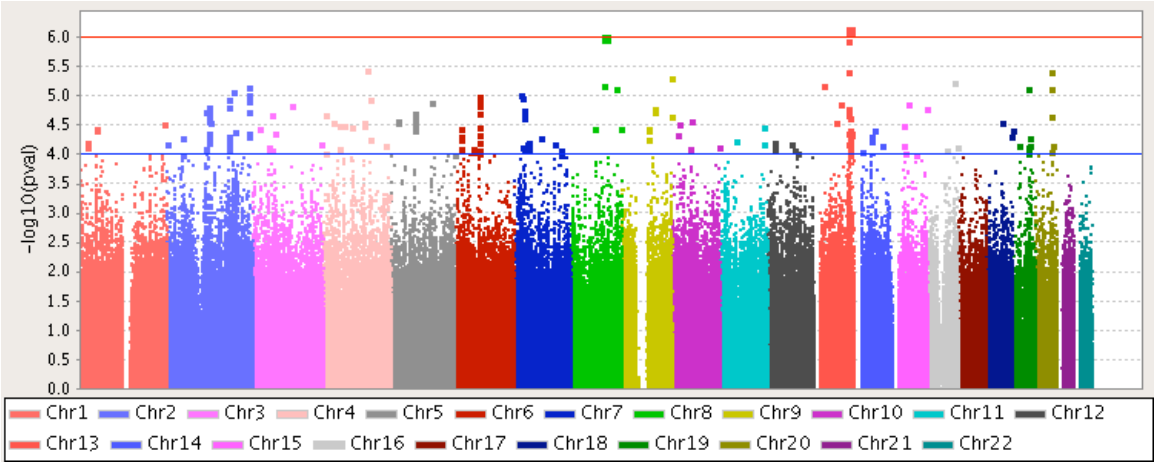
Manhattan plot for GWAS of DSST in African Americans



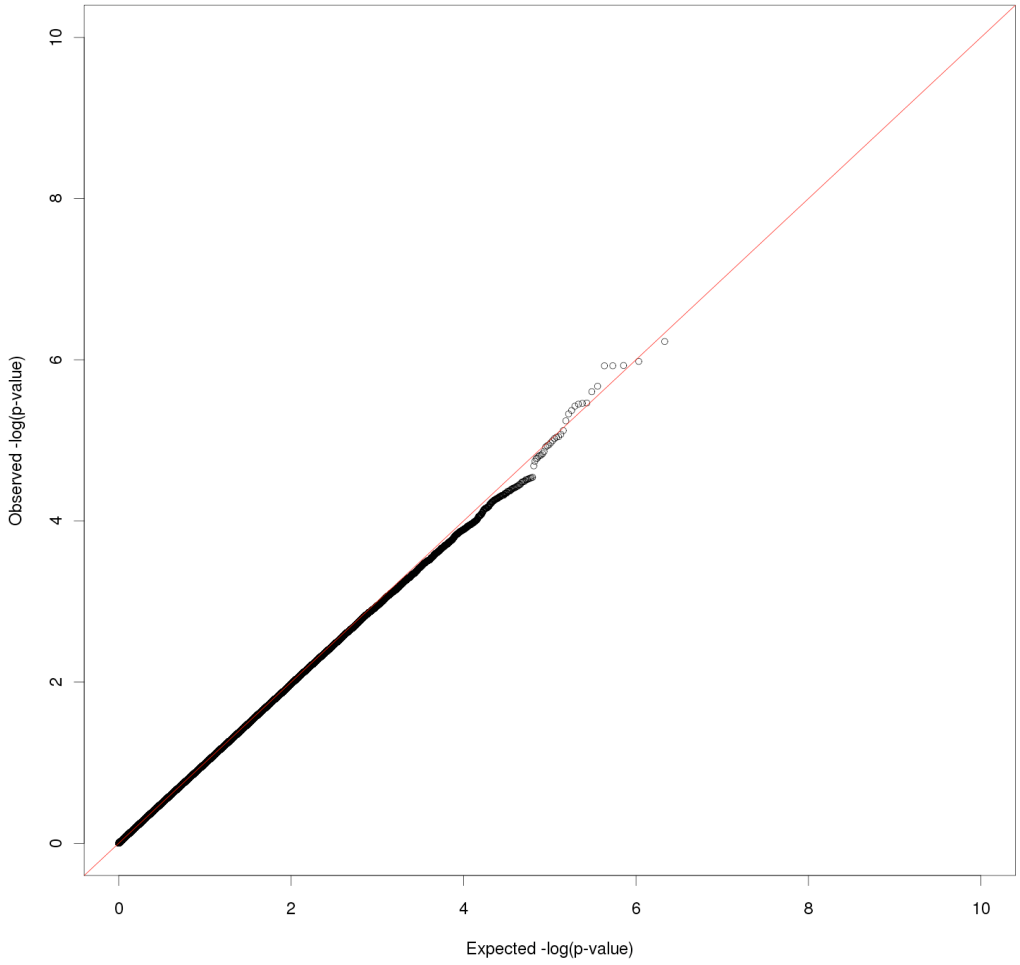
QQ plot for GWAS of COWA FAS in African Americans



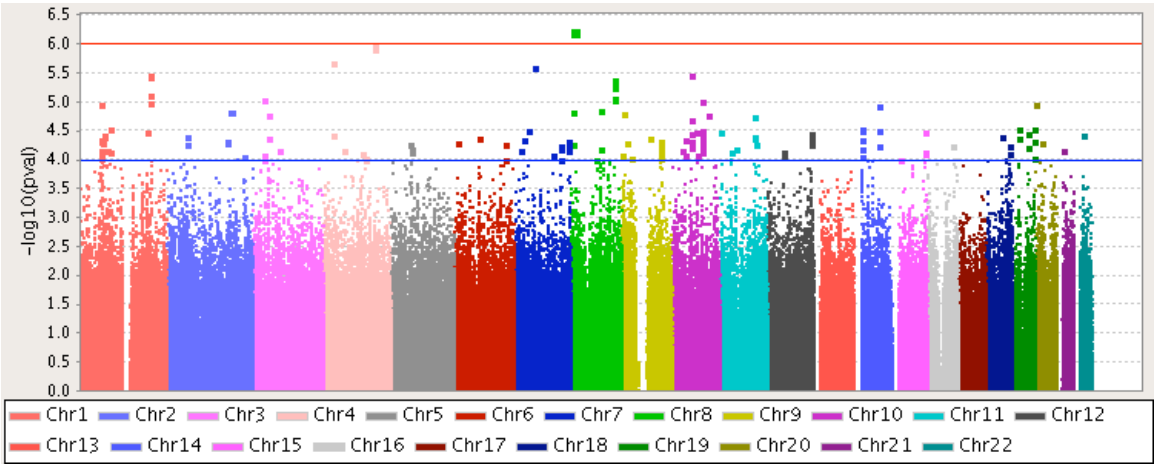
Manhattan plot for GWAS of COWA FAS in African Americans



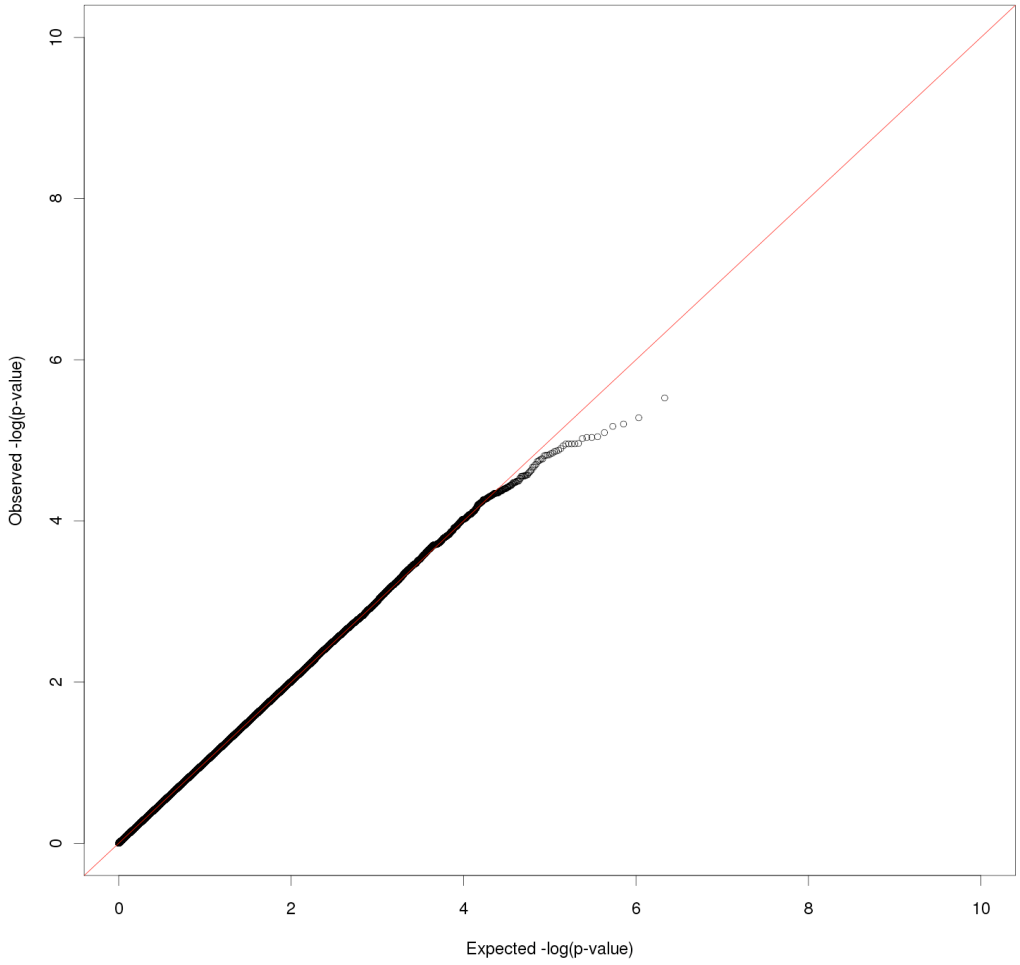
QQ plot for GWAS of COWA animals in African Americans



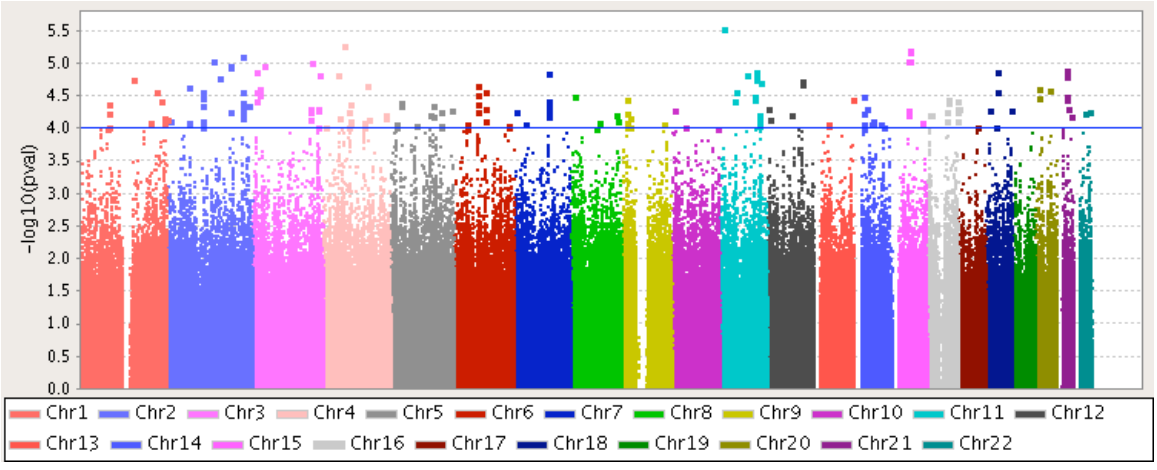
Manhattan plot for GWAS of COWA animals in African Americans



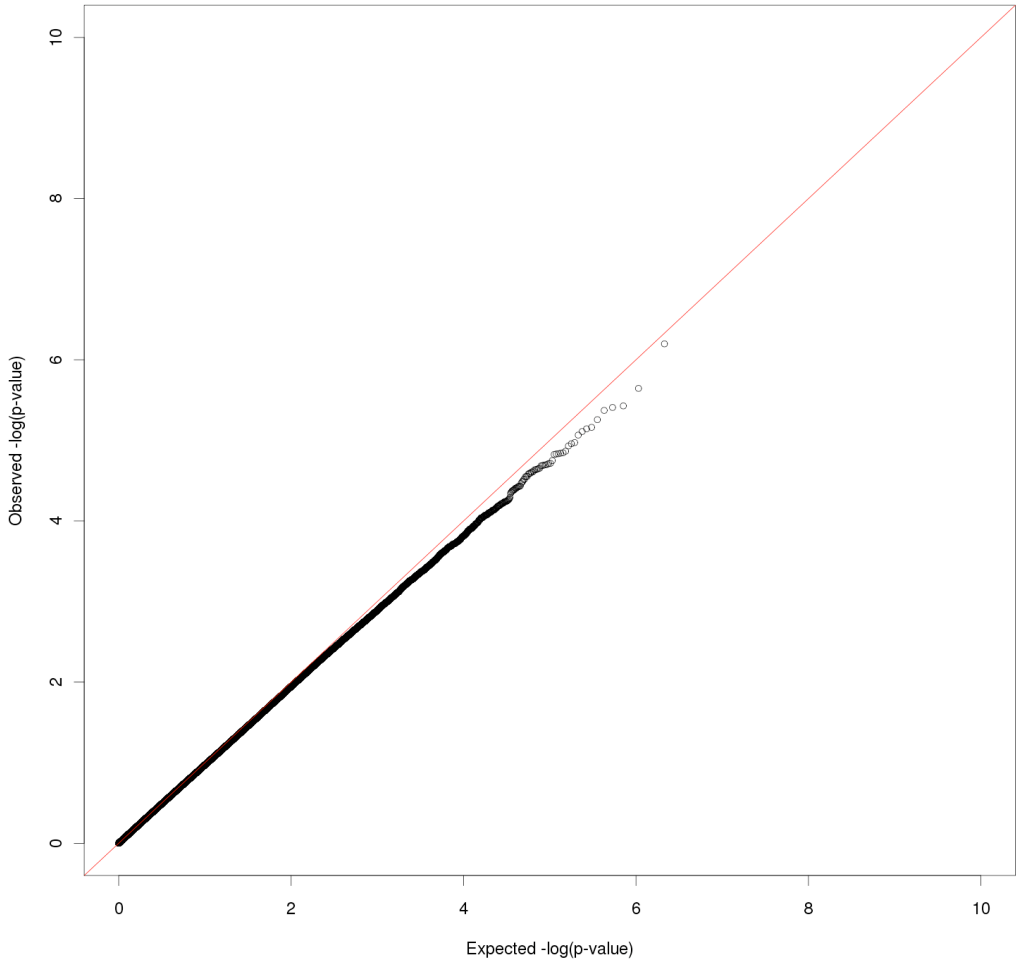
Q-Q plot for GWAS of Stroop color word in African Americans



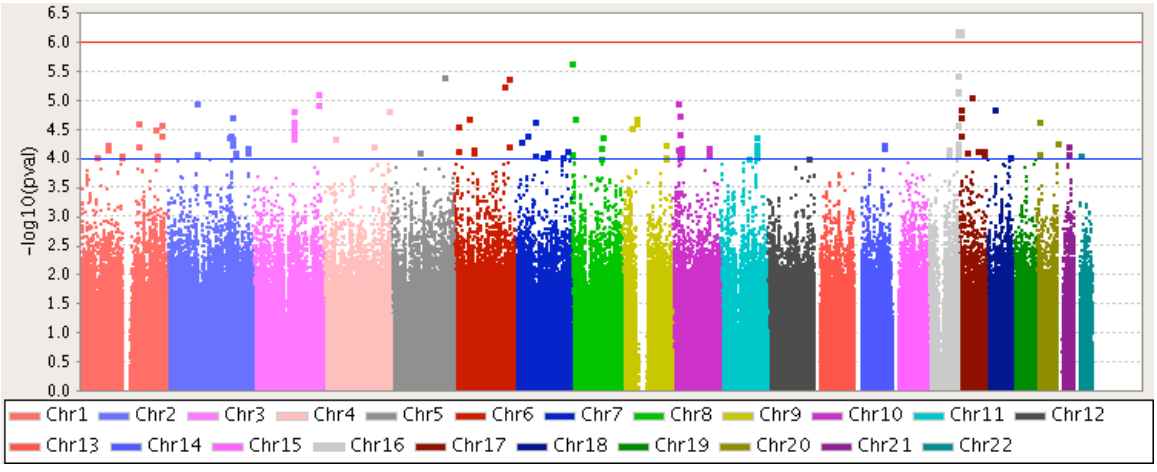
Manhattan plot for GWAS of Stroop color word in African Americans



Q-Q plot for GWAS of Stroop difference in African Americans



Manhattan plot for GWAS of Stroop difference in African Americans

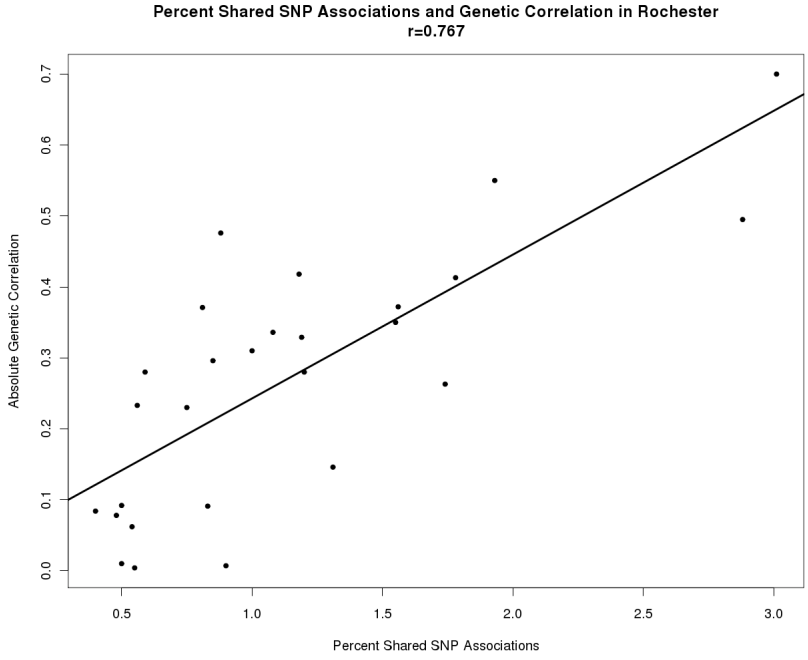


Genome-wide inflation factors for GWAS in whites and African Americans

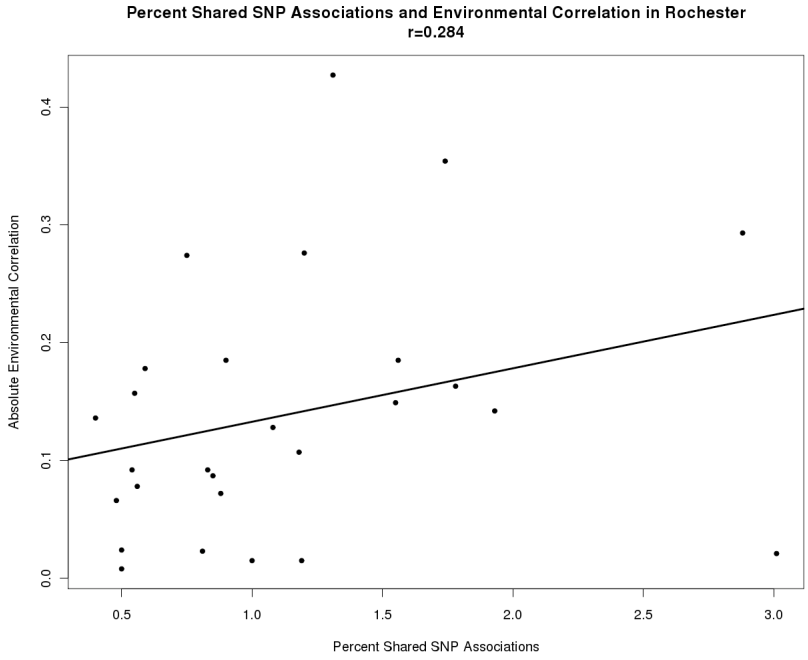
Trait	Whites	African Americans
Ln(leukoaraiosis+1)	1.008	1.011
RAVLT delayed recall	1.012	1.017
RAVLT total learning	1.011	1.019
DSST	1.017	1.003
COWA FAS	1.022	1.014
COWA animals	1.012	0.999
Stroop color word	1.006	1.007
Stroop difference	1.009	0.990

Appendix 4.5. Percentage of shared SNPs and genetic/environmental correlations

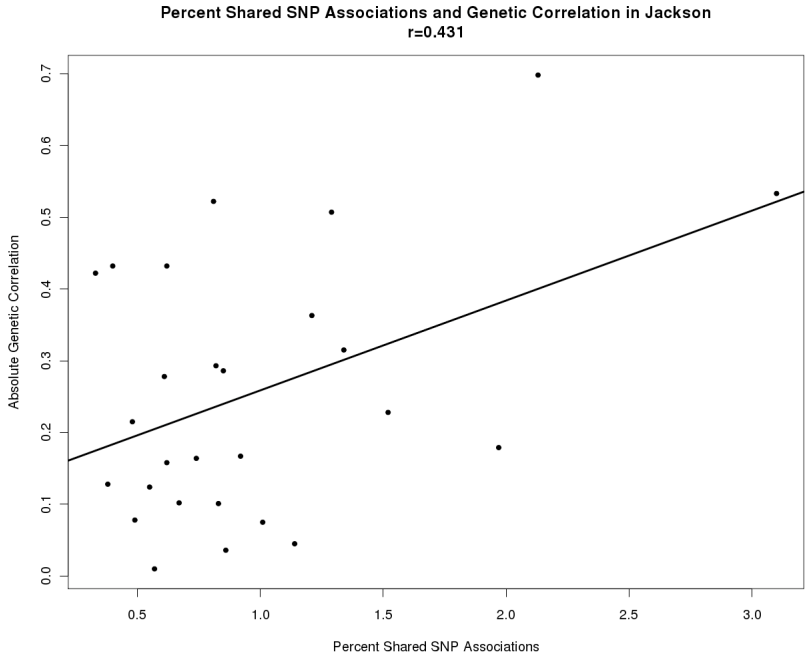
Percentage of shared SNPs and genetic correlation of trait pairs in whites



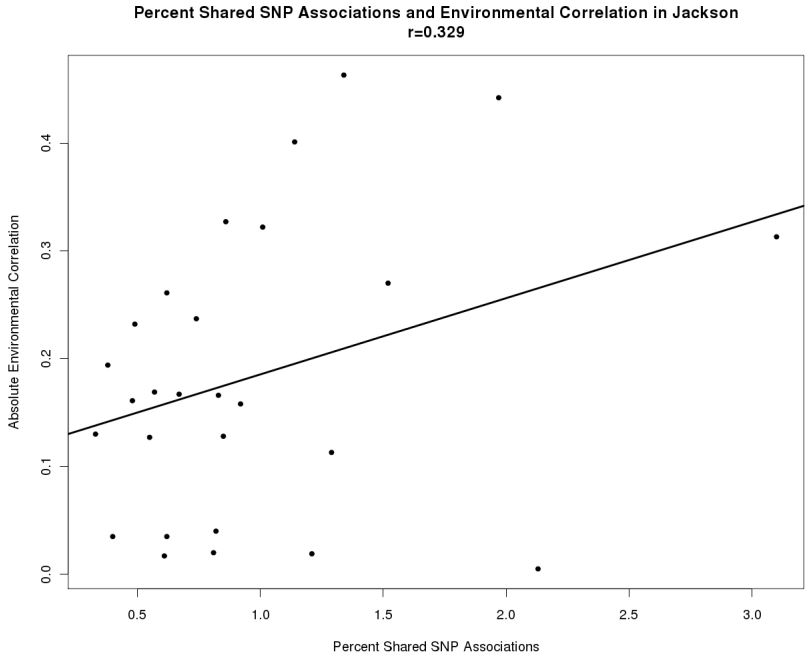
Percentage of shared SNPs and environmental correlation of trait pairs in whites



Percentage of shared SNPs and genetic correlation of trait pairs in African Americans



Percentage of shared SNPs and environmental correlation of trait pairs in African Americans



Appendix 4.6. Summary of evidence for pleiotropy among four traits selected for permutation GWAS

Phenotypic correlations

	Ln (leuko+1)	RAVLT total learning	DSST	COWA animals
Ln(leukoaraiosis+1)	0.529*** (0.09)	-0.295	-0.294	-0.174
RAVLT total learning	-0.063	0.516*** (0.10)	0.523	0.429
DSST	-0.064	0.269	0.843*** (0.09)	0.334
COWA animals	-0.039	0.311	0.175	0.349*** (0.10)

Above diagonal: phenotypic correlations, ρ_p, for unadjusted traits
Below diagonal: phenotypic correlations, ρ_p, for adjusted traits
Diagonal: heritabilities from univariate polygenic analysis, h^2 (SE), for adjusted traits
For all adjusted traits, biometric models included age, sex, and education. The biometric model for ln(leukoaraiosis+1) also included TIV.
Null hypothesis of tests: $h^2 = 0$ (diagonal)
*** p-value < 0.001

Genetic and environmental correlations

	Ln (leuko+1)	RAVLT total learning	DSST	COWA animals
Ln(leukoaraiosis+1)	0.529*** (0.09)	0.178 (0.14)	-0.008 (0.22)	-0.136 (0.12)
RAVLT total learning	-0.280* (0.14)	0.516*** (0.10)	0.354 (0.21)	0.142 (0.12)
DSST	-0.092 (0.12)	0.263* (0.11)	0.843*** (0.09)	0.015 (0.20)
COWA animals	0.084 (0.17)	0.550** (0.15)	0.310* (0.13)	0.349*** (0.10)

Above diagonal: environmental correlations, ρ_e (SE)
Below diagonal: genetic correlations, ρ_g (SE)
Diagonal: heritabilities from univariate polygenic analysis, h^2 (SE), for adjusted traits
For all adjusted traits, biometric models included age, sex, and education. The biometric model for ln(leukoaraiosis+1) also included TIV.
Null hypothesis of tests: $\rho_e = 0$ (below diagonal)
Null hypothesis of tests: $\rho_g = 0$ (above diagonal)
Null hypothesis of tests: $h^2 = 0$ (diagonal)
* 0.01 < p-value < 0.05
** 0.001 < p-value < 0.01
*** p-value < 0.001

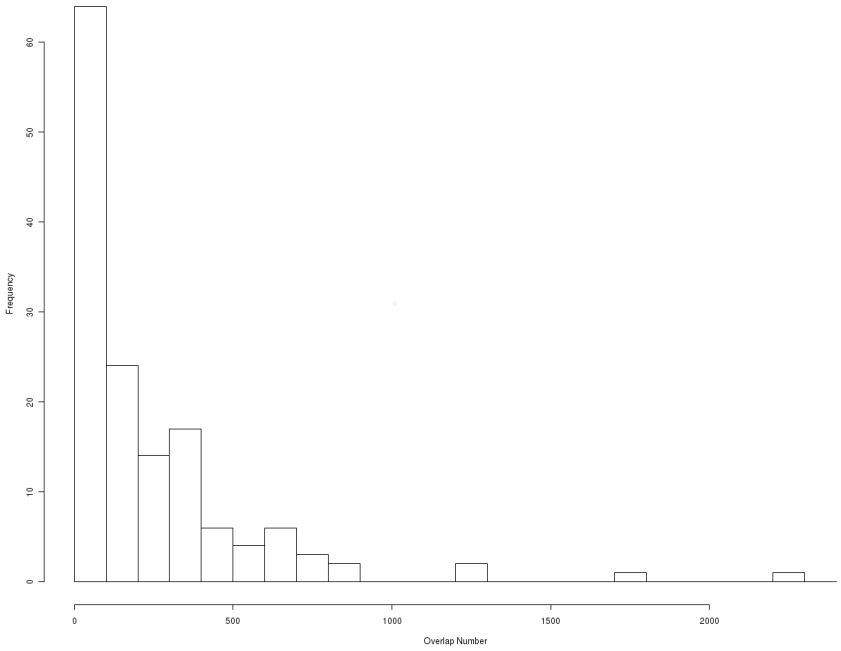
Percentage of shared SNP associations at α=0.01

	Ln (leuko+1)	RAVLT total learning	DSST	COWA animals
Ln(leukoaraiosis+1)	26189 (1.09%)			
RAVLT total learning	300 (0.59%)	24673 (1.03%)		
DSST	259 (0.50%)	858 (1.74%)	25449 (1.06%)	
COWA animals	201 (0.40%)	931 (1.93%)	496 (1.00%)	24422 (1.02%)

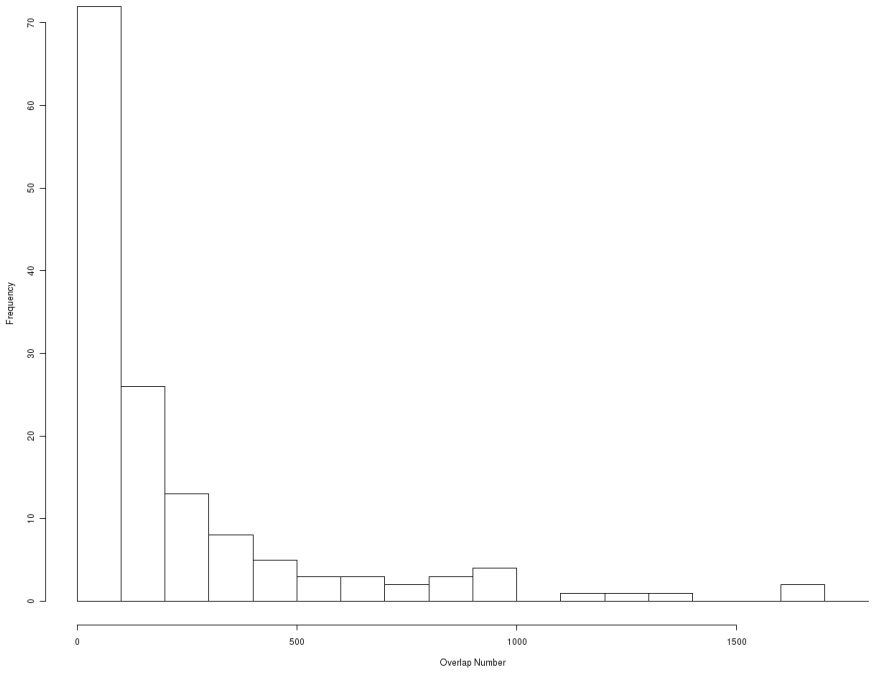
Below diagonal: observed number (percentage) of shared significant SNP associations at α=0.01
Diagonal: observed number (percentage) of significant associations for individual traits at α=0.01. The total number of SNPs tested for association was 2,401,820.

Appendix 4.7. Examples of histograms of permutation test results

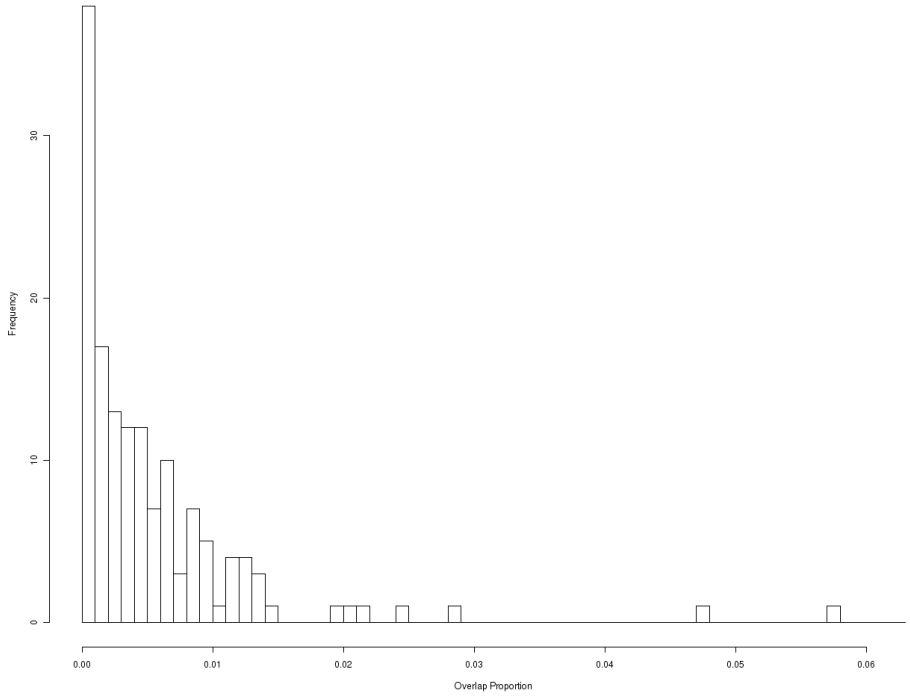
Histogram of number of shared associated SNPs for ln(leukoaraiosis+1) and RAVLT total learning from paired permutation tests



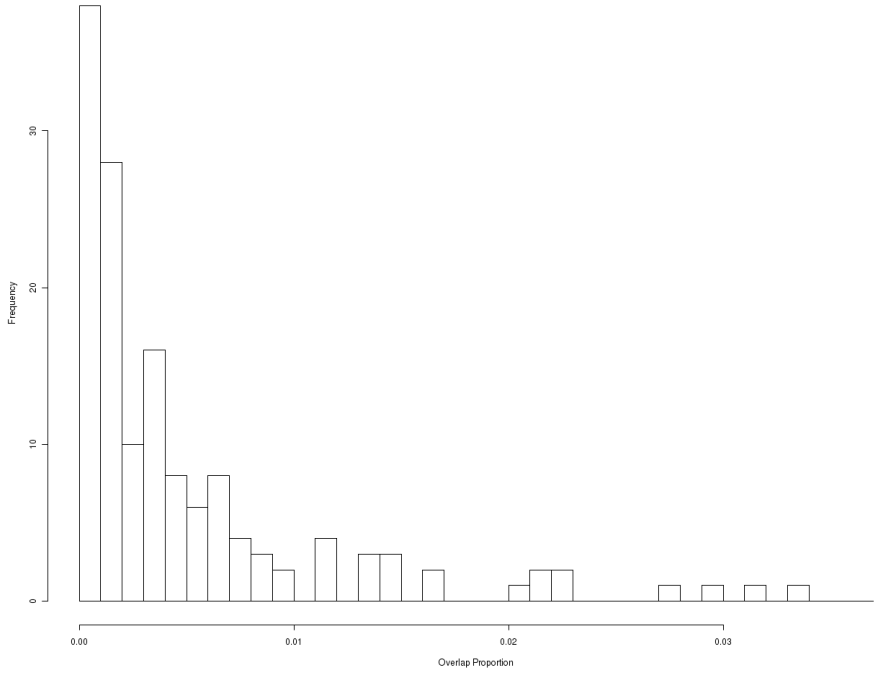
Histogram of number of shared associated SNPs for ln(leukoaraiosis+1) and RAVLT total learning from unpaired permutation tests



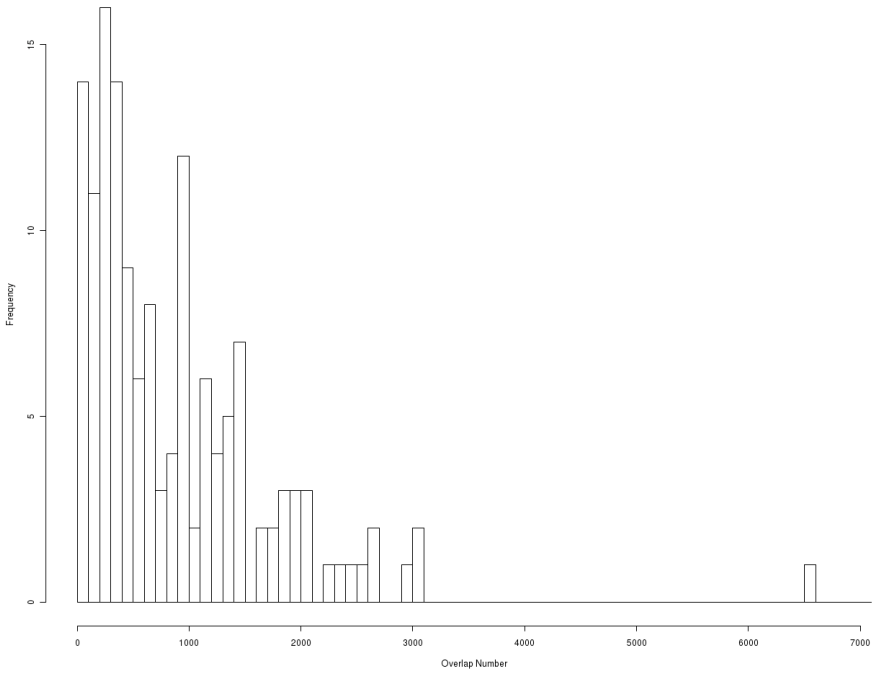
Histogram of percentage of shared associated SNPs for ln(leukoaraiosis+1) and RAVLT total learning from paired permutation tests



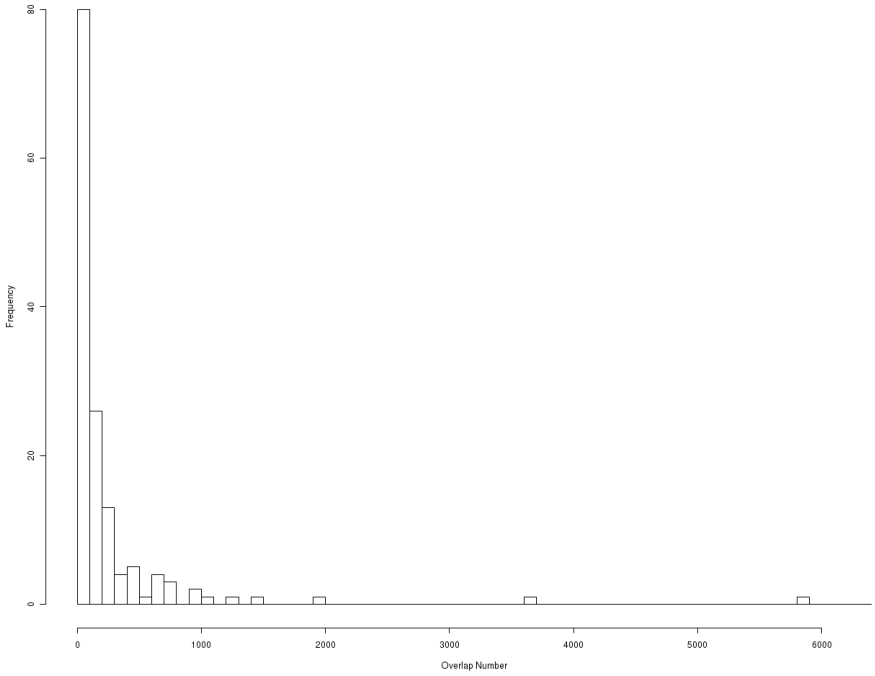
Histogram of percentage of shared associated SNPs for ln(leukoaraiosis+1) and RAVLT total learning from unpaired permutation tests



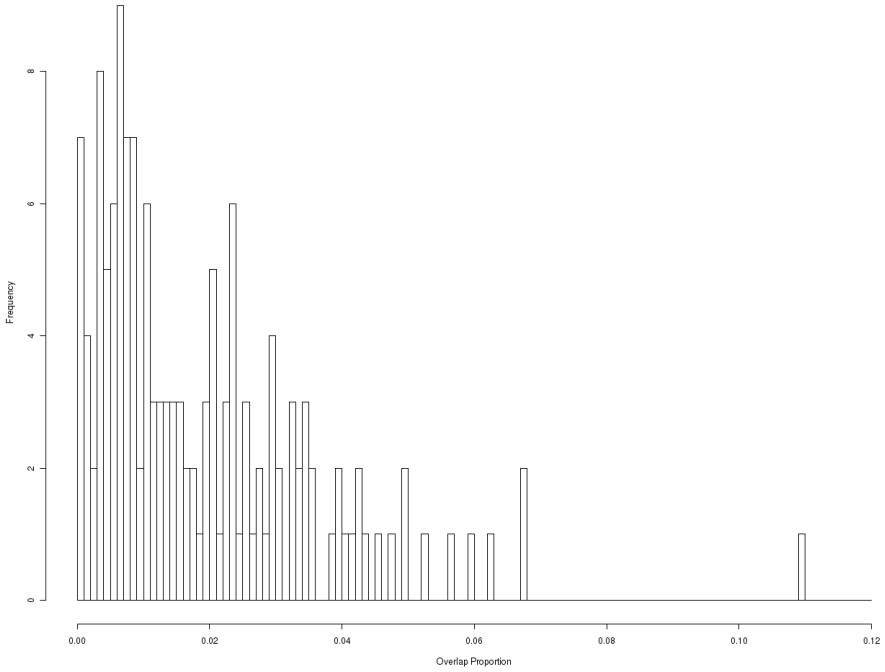
Histogram of number of shared associated SNPs for RAVLT total learning and COWA animals from paired permutation tests



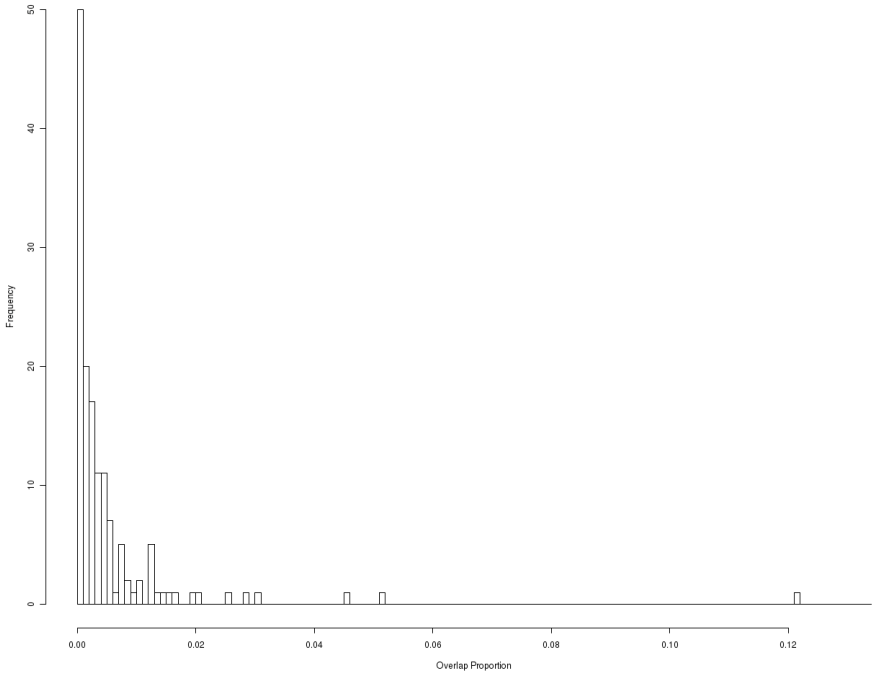
Histogram of number of shared associated SNPs for RAVLT total learning and COWA animals from unpaired permutation tests



Histogram of percentage of shared associated SNPs RAVLT total learning and COWA animals from paired permutation tests



Histogram of percentage of shared associated SNPs for RAVLT total learning and COWA animals from unpaired permutation tests



References

Affymetrix. (2007). Affymetrix® Genome-Wide Human SNP Nsp/Sty 6.0 User Guide.

Almasy, L., and Blangero, J. (1998). Multipoint Quantitative-Trait Linkage Analysis in General Pedigrees. *Am J Hum Genet* **62(5):** 1198-1211.

Alves de Moraes, S., Szklo, M., Knopman, D., and Sato, R. (2002). The Relationship between Temporal Changes in Blood Pressure and Changes in Cognitive Function: Atherosclerosis Risk in Communities (ARIC) Study. *Prev Med* **35(3):** 258-263.

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders.* Washington, D.C.

Atwood, L.D., Wolf, P.A., Heard-Costa, N.L., Massaro, J.M., Beiser, A., D'Agostino, R.B., and DeCarli, C. (2004). Genetic Variation in White Matter Hyperintensity Volume in the Framingham Study. *Stroke* **35(7):** 1609-1613.

Ayroles, J.F., Carbone, M.A., Stone, E.A., Jordan, K.W., Lyman, R.F., Magwire, M.M., Rollmann, S.M., Duncan, L.H., Lawrence, F., Anholt, R.R. et al. (2009). Systems Genetics of Complex Traits in Drosophila Melanogaster. *Nat Genet* **41(3):** 299-307.

Benton, A.L. (1994). Neuropsychological Assessment. *Annu Rev Psychol* **45:** 1-23.

Boustani, M., Peterson, B., Hanson, L., Harris, R., Lohr, K.N., and U.S. Preventive Services Task Force. (2003). Screening for Dementia in Primary Care: A Summary of the Evidence for the U.S. Preventive Services Task Force. *Ann Intern Med* **138(11):** 927-937.

Breteler, M.M. (2000). Vascular Risk Factors for Alzheimer's Disease: An Epidemiologic Perspective. *Neurobiol Aging* **21(2):** 153-160.

Carden, S.M., Boissy, R.E., Schoettker, P.J., and Good, W.V. (1998). Albinism: Modern Molecular Diagnosis. *Br J Ophthalmol* **82(2):** 189-195.

Carmelli, D., Reed, T., and DeCarli, C. (2002). A Bivariate Genetic Analysis of Cerebral White Matter Hyperintensities and Cognitive Performance in Elderly Male Twins. *Neurobiol Aging* **23(3):** 413-420.

Carmelli, D., DeCarli, C., Swan, G.E., Jack, L.M., Reed, T., Wolf, P.A., and Miller, B.L. (1998). Evidence for Genetic Variance in White Matter Hyperintensity Volume in Normal Elderly Male Twins. *Stroke* **29(6):** 1177-1181.

Cassidy-Bushrow, A.E., Bielak, L.F., Sheedy, P.F.,2nd, Turner, S.T., Kullo, I.J., Lin, X., and Peyser, P.A. (2007). Coronary Artery Calcification Progression is Heritable. *Circulation* **116(1):** 25-31.

Charlesworth, J., Kramer, P.L., Dyer, T., Diego, V., Samples, J.R., Craig, J.E., Mackey, D.A., Hewitt, A.W., Blangero, J., and Wirtz, M.K. (2010). The Path to Open-Angle Glaucoma Gene Discovery: Endophenotypic Status of Intraocular Pressure, Cup-to-Disc Ratio, and Central Corneal Thickness. *Invest Ophthalmol Vis Sci* **51(7):** 3509-3514.

Cheverud, J.M., Ehrich, T.H., Vaughn, T.T., Koreishi, S.F., Linsey, R.B., and Pletscher, L.S. (2004). Pleiotropic Effects on Mandibular Morphology II: Differential Epistasis and Genetic Variation in Morphological Integration. *J Exp Zool B Mol Dev Evol* **302(5):** 424-435.

Comuzzie, A.G., Blangero, J., Mahaney, M.C., Mitchell, B.D., Stern, M.P., and MacCluer, J.W. (1994). Genetic and Environmental Correlations among Skinfold Measures. *Int J Obes Relat Metab Disord* **18(6):** 413-418.

Daniels, P.R., Kardia, S.L., Hanis, C.L., Brown, C.A., Hutchinson, R., Boerwinkle, E., Turner, S.T., and Genetic Epidemiology Network of Arteriopathy study. (2004). Familial Aggregation of Hypertension Treatment and Control in the Genetic Epidemiology Network of Arteriopathy (GENOA) Study. *Am J Med* **116(10):** 676-681.

de Bakker, P.I., Ferreira, M.A., Jia, X., Neale, B.M., Raychaudhuri, S., and Voight, B.F. (2008). Practical Aspects of Imputation-Driven Meta-Analysis of Genome-Wide Association Studies. *Hum Mol Genet* **17(R2):** R122-8.

De Groot, J.C., De Leeuw, F.E., Oudkerk, M., Van Gijn, J., Hofman, A., Jolles, J., and Breteler, M.M. (2002). Periventricular Cerebral White Matter Lesions Predict Rate of Cognitive Decline. *Ann Neurol* **52(3):** 335-341.

Deary, I.J., Wright, A.F., Harris, S.E., Whalley, L.J., and Starr, J.M. (2004). Searching for Genetic Influences on Normal Cognitive Ageing. *Trends Cogn Sci* **8(4):** 178-184.

Debette, S., Bis, J.C., Fornage, M., Schmidt, H., Ikram, M.A., Sigurdsson, S., Heiss, G., Struchalin, M., Smith, A.V., van der Lugt, A. et al. (2010). Genome-Wide Association Studies of MRI-Defined Brain Infarcts: Meta-Analysis from the CHARGE Consortium. *Stroke* **41(2):** 210-217.

DeKosky, S.T., and Marek, K. (2003). Looking Backward to Move Forward: Early Detection of Neurodegenerative Disorders. *Science* **302(5646):** 830-834.

Dickstein, D.L., Walsh, J., Brautigam, H., Stockton, S.D.,Jr, Gandy, S., and Hof, P.R. (2010). Role of Vascular Risk Factors and Vascular Dysfunction in Alzheimer's Disease. *Mt Sinai J Med* **77(1):** 82-102.

Donnan, G.A., Fisher, M., Macleod, M., and Davis, S.M. (2008). Stroke. *Lancet* **371(9624):** 1612-1623.

Elias, M.F. (1998). Effects of Chronic Hypertension on Cognitive Functioning. *Geriatrics* **53(Suppl 1):** S49-52.

Elias, M.F., Wolf, P.A., D'Agostino, R.B., Cobb, J., and White, L.R. (1993). Untreated Blood Pressure Level is Inversely Related to Cognitive Functioning: The Framingham Study. *Am J Epidemiol* **138(6):** 353-364.

FBPP Investigators. (2002). Multi-Center Genetic Study of Hypertension: The Family Blood Pressure Program (FBPP). *Hypertension* **39(1):** 3-9.

Finkel, D., and Pedersen, N.L. (2000). Contribution of Age, Genes, and Environment to the Relationship between Perceptual Speed and Cognitive Ability. *Psychol Aging* **15(1):** 56-64.

Finkel, D., Pedersen, N.L., McGue, M., and McClearn, G.E. (1995). Heritability of Cognitive Abilities in Adult Twins: Comparison of Minnesota and Swedish Data. *Behav Genet* **25(5):** 421-431.

Fu, J.H., Lu, C.Z., Hong, Z., Dong, Q., Luo, Y., and Wong, K.S. (2005). Extent of White Matter Lesions is Related to Acute Subcortical Infarcts and Predicts further Stroke Risk in Patients with First Ever Ischaemic Stroke. *J Neurol Neurosurg Psychiatry* **76(6):** 793-796.

Golden, C.J. (1978). *Stroop Color and Word Test: A Manual for Clinical and Experimental Uses.* Shoelting Company: Wood Dale, IL.

Gudmundsson, J., Sulem, P., Steinthorsdottir, V., Bergthorsson, J.T., Thorleifsson, G., Manolescu, A., Rafnar, T., Gudbjartsson, D., Agnarsson, B.A., Baker, A. et al. (2007). Two Variants on Chromosome 17 Confer Prostate Cancer Risk, and the One in TCF2 Protects Against Type 2 Diabetes. *Nat Genet* **39(8):** 977-983.

Gunsalus, K.C., Ge, H., Schetter, A.J., Goldberg, D.S., Han, J.D., Hao, T., Berriz, G.F., Bertin, N., Huang, J., Chuang, L.S. et al. (2005). Predictive Models of Molecular Machines Involved in Caenorhabditis Elegans Early Embryogenesis. *Nature* **436(7052):** 861-865.

Haan, M.N., and Wallace, R. (2004). Can Dementia be Prevented? Brain Aging in a Population-Based Context. *Annu Rev Public Health* **25:** 1-24.

Haan, M.N., Shemanski, L., Jagust, W.J., Manolio, T.A., and Kuller, L. (1999). The Role of APOE epsilon4 in Modulating Effects of Other Risk Factors for Cognitive Decline in Elderly Persons. *JAMA* **282(1):** 40-46.

Han, J.D., Bertin, N., Hao, T., Goldberg, D.S., Berriz, G.F., Zhang, L.V., Dupuy, D., Walhout, A.J., Cusick, M.E., Roth, F.P. et al. (2004). Evidence for Dynamically

Organized Modularity in the Yeast Protein-Protein Interaction Network. *Nature* **430(6995):** 88-93.

Hodgkin, J. (1998). Seven Types of Pleiotropy. *Int J Dev Biol* **42(3):** 501-505.

Hsu, Y.H., Zillikens, M.C., Wilson, S.G., Farber, C.R., Demissie, S., Soranzo, N., Bianchi, E.N., Grundberg, E., Liang, L., Richards, J.B. et al. (2010). An Integration of Genome-Wide Association Study and Gene Expression Profiling to Prioritize the Discovery of Novel Susceptibility Loci for Osteoporosis-Related Traits. *PLoS Genet* **6(6):** e1000977.

Ihmels, J., Friedlander, G., Bergmann, S., Sarig, O., Ziv, Y., and Barkai, N. (2002). Revealing Modular Organization in the Yeast Transcriptional Network. *Nat Genet* **31(4):** 370-377.

Illumina. (2010). Genome-Wide DNA Analysis Beadchips.

Ivnik, R., Malec, J., and Smith, G. (1992). Mayo's Older Americans Normative Studies: Updated AVLT Norms for Ages 56 to 97. *The Clinical Neuropsychologist* **6(Supp):** 83-104.

Jack, C.R.,Jr, O'Brien, P.C., Rettman, D.W., Shiung, M.M., Xu, Y., Muthupillai, R., Manduca, A., Avula, R., and Erickson, B.J. (2001). FLAIR Histogram Segmentation for Measurement of Leukoaraiosis Volume. *J Magn Reson Imaging* **14(6):** 668-676.

Jack, C.R.,Jr, Twomey, C.K., Zinsmeister, A.R., Sharbrough, F.W., Petersen, R.C., and Cascino, G.D. (1989). Anterior Temporal Lobes and Hippocampal Formations: Normative Volumetric Measurements from MR Images in Young Adults. *Radiology* **172(2):** 549-554.

Jumbo-Lucioni, P., Ayroles, J.F., Chambers, M.M., Jordan, K.W., Leips, J., Mackay, T.F., and De Luca, M. (2010). Systems Genetics Analysis of Body Weight and Energy Metabolism Traits in Drosophila Melanogaster. *BMC Genomics* **11:** 297.

Kannel, W.B. (1995). Framingham Study Insights into Hypertensive Risk of Cardiovascular Disease. *Hypertens Res* **18(3):** 181-196.

Karasik, D., Hsu, Y.H., Zhou, Y., Cupples, L.A., Kiel, D.P., and Demissie, S. (2010). Genome-Wide Pleiotropy of Osteoporosis-Related Phenotypes: The Framingham Study. *J Bone Miner Res* **25(7):** 1555-1563.

Kenney-Hunt, J.P., Vaughn, T.T., Pletscher, L.S., Peripato, A., Routman, E., Cothran, K., Durand, D., Norgard, E., Perel, C., and Cheverud, J.M. (2006). Quantitative Trait Loci for Body Size Components in Mice. *Mamm Genome* **17(6):** 526-537.

Knecht, S., Wersching, H., Lohmann, H., Berger, K., and Ringelstein, E.B. (2009). How Much does Hypertension Affect Cognition?: Explained Variance in Cross-Sectional Analysis of Non-Demented Community-Dwelling Individuals in the SEARCH Study. *J Neurol Sci* **283(1-2):** 149-152.

Knecht, S., Wersching, H., Lohmann, H., Bruchmann, M., Duning, T., Dziewas, R., Berger, K., and Ringelstein, E.B. (2008). High-Normal Blood Pressure is Associated with Poor Cognitive Performance. *Hypertension* **51(3):** 663-668.

Knopman, D., Boland, L.L., Mosley, T., Howard, G., Liao, D., Szklo, M., McGovern, P., Folsom, A.R., and Atherosclerosis Risk in Communities (ARIC) Study Investigators. (2001). Cardiovascular Risk Factors and Cognitive Decline in Middle-Aged Adults. *Neurology* **56(1):** 42-48.

Kukull, W.A., and Bowen, J.D. (2002). Dementia Epidemiology. *Med Clin North Am* **86(3):** 573-590.

Kuller, L.H., Lopez, O.L., Jagust, W.J., Becker, J.T., DeKosky, S.T., Lyketsos, C., Kawas, C., Breitner, J.C., Fitzpatrick, A., and Dulberg, C. (2005). Determinants of Vascular Dementia in the Cardiovascular Health Cognition Study. *Neurology* **64(9):** 1548-1552.

Kuller, L.H., Shemanski, L., Manolio, T., Haan, M., Fried, L., Bryan, N., Burke, G.L., Tracy, R., and Bhadelia, R. (1998). Relationship between ApoE, MRI Findings, and Cognitive Function in the Cardiovascular Health Study. *Stroke* **29(2):** 388-398.

Launer, L.J., Ross, G.W., Petrovitch, H., Masaki, K., Foley, D., White, L.R., and Havlik, R.J. (2000). Midlife Blood Pressure and Dementia: The Honolulu-Asia Aging Study. *Neurobiol Aging* **21(1):** 49-55.

Leamy, L.J., Pomp, D., Eisen, E.J., and Cheverud, J.M. (2002). Pleiotropy of Quantitative Trait Loci for Organ Weights and Limb Bone Lengths in Mice. *Physiol Genomics* **10(1):** 21-29.

Lezak, M. (1995). *Neuropsychological Assessment.* Oxford University Press: New York.

Li, Y., and Abecasis, G.R. (2006). Mach 1.0: Rapid Haplotype Reconstruction and Missing Genotype Inference. *American Journal of Human Genetics* **S79:** 2290.

Liao, D., Cooper, L., Cai, J., Toole, J.F., Bryan, N.R., Hutchinson, R.G., and Tyroler, H.A. (1996). Presence and Severity of Cerebral White Matter Lesions and Hypertension, its Treatment, and its Control. the ARIC Study. Atherosclerosis Risk in Communities Study. *Stroke* **27(12):** 2262-2270.

Mackay, T.F. (2010). Mutations and Quantitative Genetic Variation: Lessons from Drosophila. *Philos Trans R Soc Lond B Biol Sci* **365(1544):** 1229-1239.

Mackay, T.F. (2009). The Genetic Architecture of Complex Behaviors: Lessons from Drosophila. *Genetica* **136(2):** 295-302.

Mackay, T.F., Heinsohn, S.L., Lyman, R.F., Moehring, A.J., Morgan, T.J., and Rollmann, S.M. (2005). Genetics and Genomics of Drosophila Mating Behavior. *Proc Natl Acad Sci USA* **102(Suppl 1):** 6622-6629.

Magwire, M.M., Yamamoto, A., Carbone, M.A., Roshina, N.V., Symonenko, A.V., Pasyukova, E.G., Morozova, T.V., and Mackay, T.F. (2010). Quantitative and Molecular Genetic Analyses of Mutations Increasing Drosophila Life Span. *PLoS Genet* **6(7):** e1001037.

Markus, H.S., Hunt, B., Palmer, K., Enzinger, C., Schmidt, H., and Schmidt, R. (2005). Markers of Endothelial and Hemostatic Activation and Progression of Cerebral White Matter Hyperintensities: Longitudinal Results of the Austrian Stroke Prevention Study. *Stroke* **36(7):** 1410-1414.

Mattay, V.S., Goldberg, T.E., Sambataro, F., and Weinberger, D.R. (2008). Neurobiology of Cognitive Aging: Insights from Imaging Genetics. *Biol Psychol* **79(1):** 9-22.

McGue, M., and Christensen, K. (2002). The Heritability of Level and Rate-of-Change in Cognitive Functioning in Danish Twins Aged 70 Years and Older. *Exp Aging Res* **28(4):** 435-451.

McGue, M., and Christensen, K. (2001). The Heritability of Cognitive Functioning in very Old Adults: Evidence from Danish Twins Aged 75 Years and Older. *Psychol Aging* **16(2):** 272-280.

Neale, M.C., and Miller, M.B. (1997). The use of Likelihood-Based Confidence Intervals in Genetic Models. *Behav Genet* **27(2):** 113-120.

O'Sullivan, M. (2008). Leukoaraiosis. *Pract Neurol* **8(1):** 26-38.

Pantoni, L., Poggesi, A., and Inzitari, D. (2007). The Relation between White-Matter Lesions and Cognition. *Curr Opin Neurol* **20(4):** 390-397.

Pantoni, L., and Garcia, J.H. (1997). Pathogenesis of Leukoaraiosis: A Review. *Stroke* **28(3):** 652-659.

Paternoster, L., Chen, W., and Sudlow, C.L. (2009). Genetic Determinants of White Matter Hyperintensities on Brain Scans: A Systematic Assessment of 19 Candidate Gene Polymorphisms in 46 Studies in 19,000 Subjects. *Stroke* **40(6):** 2020-2026.

Plomin, R., Pedersen, N.L., Lichtenstein, P., and McClearn, G.E. (1994). Variability and Stability in Cognitive Abilities are Largely Genetic Later in Life. *Behav Genet* **24(3):** 207-215.

Pohjasvaara, T., Mantyla, R., Ylikoski, R., Kaste, M., and Erkinjuntti, T. (2000). Comparison of Different Clinical Criteria (DSM-III, ADDTC, ICD-10, NINDS-AIREN, DSM-IV) for the Diagnosis of Vascular Dementia. National Institute of Neurological Disorders and Stroke-Association Internationale Pour La Recherche Et l'Enseignement En Neurosciences. *Stroke* **31(12):** 2952-2957.

R Core Development Team. (2008). R: A Language and Environment for Statistical Computing.

Raudenbush, S.W., and Bryk, A.S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods, 2nd Ed.* Sage Publications, Inc: Thousand Oaks, CA.

Rey, A. (1964). *L'Examen Clinique en Psychologie.* Presses Universitaires de France: 1964.

Roger, V.L., Go, A.S., Lloyd-Jones, D.M., Adams, R.J., Berry, J.D., Brown, T.M., Carnethon, M.R., Dai, S., de Simone, G., Ford, E.S., et al. (2011). Heart Disease and Stroke Statistics 2011 Update: A Report From the American Heart Association. *Circulation* **123**: e000-e000 (epub ahead of print).

Schmidt, R., Petrovic, K., Ropele, S., Enzinger, C., and Fazekas, F. (2007). Progression of Leukoaraiosis and Cognition. *Stroke* **38(9):** 2619-2625.

Schmidt, R., Scheltens, P., Erkinjuntti, T., Pantoni, L., Markus, H.S., Wallin, A., Barkhof, F., and Fazekas, F. (2004). White Matter Lesion Progression: A Surrogate Endpoint for Trials in Cerebral Small-Vessel Disease. *Neurology* **63(1):** 139-144.

Seshadri, S., DeStefano, A.L., Au, R., Massaro, J.M., Beiser, A.S., Kelly-Hayes, M., Kase, C.S., D'Agostino RB, S., Decarli, C., Atwood, L.D. et al. (2007). Genetic Correlates of Brain Aging on MRI and Cognitive Test Measures: A Genome-Wide Association and Linkage Analysis in the Framingham Study. *BMC Med Genet* **8(Suppl 1):** S15.

Sierra, C., and Coca, A. (2006). White Matter Lesions and Cognitive Impairment as Silent Cerebral Disease in Hypertension. *ScientificWorldJournal* **6:** 494-501.

Sing, C.F., Boerwinkle, E., Moll, P.P., and Templeton, A.R. (1987). *Characterization of Genes affecting Quantitative Traits in Humans.* In *Proceedings of the Second International Conference on Quantitative Genetics,* B. Weir, E.J. Eisen, M.M. Goodman, and G. Namkoong, eds. Sinauer Associates, Inc.: Raliegh, NC, pp. 250-269.

Slattery, M.L., Folsom, A.R., Wolff, R., Herrick, J., Caan, B.J., and Potter, J.D. (2008). Transcription Factor 7-Like 2 Polymorphism and Colon Cancer. *Cancer Epidemiol Biomarkers Prev* **17(4):** 978-982.

Sleegers, K., de Koning, I., Aulchenko, Y.S., van Rijn, M.J., Houben, M.P., Croes, E.A., van Swieten, J.C., Oostra, B.A., and van Duijn, C.M. (2007). Cerebrovascular Risk Factors do Not Contribute to Genetic Variance of Cognitive Function: The ERF Study. *Neurobiol Aging* **28(5):** 735-741.

Spreen, O., and Strauss, E. (1998). *Memory. A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary.* Oxford University Press: New York, pp. 104.

Staff, R.T., Murray, A.D., Deary, I.J., and Whalley, L.J. (2004). What Provides Cerebral Reserve? *Brain* **127(Pt 5):** 1191-1199.

Stroop, J. (1935). Studies of Inference in Serial Verbal Reactions. *J Exp Psychol* **18:** 643-662.

Swan, G.E., and Carmelli, D. (2002). Evidence for Genetic Mediation of Executive Control: A Study of Aging Male Twins. *J Gerontol B Psychol Sci Soc Sci* **57(2):** P133-43.

Swan, G.E., DeCarli, C., Miller, B.L., Reed, T., Wolf, P.A., Jack, L.M., and Carmelli, D. (1998). Association of Midlife Blood Pressure to Late-Life Cognitive Decline and Brain Morphology. *Neurology* **51(4):** 986-993.

Szolnoki, Z., and Melegh, B. (2006). Gene-Gene and Gene-Environment Interplay Represent Specific Susceptibility for Different Types of Ischaemic Stroke and Leukoaraiosis. *Curr Med Chem* **13(14):** 1627-1634.

The International HapMap Consortium. (2003). The International HapMap Project. *Nature* **426(6968):** 789-796.

Thomas, G., Jacobs, K.B., Yeager, M., Kraft, P., Wacholder, S., Orr, N., Yu, K., Chatterjee, N., Welch, R., Hutchinson, A. et al. (2008). Multiple Loci Identified in a Genome-Wide Association Study of Prostate Cancer. *Nat Genet* **40(3):** 310-315.

Turner, S.T., Fornage, M., Jack, C.R.,Jr, Mosley, T.H., Knopman, D.S., Kardia, S.L., Boerwinkle, E., and de Andrade, M. (2009). Genomic Susceptibility Loci for Brain Atrophy, Ventricular Volume, and Leukoaraiosis in Hypertensive Sibships. *Arch Neurol* **66(7):** 847-857.

Turner, S.T., Jack, C.R., Fornage, M., Mosley, T.H., Boerwinkle, E., and de Andrade, M. (2004). Heritability of Leukoaraiosis in Hypertensive Sibships. *Hypertension* **43(2):** 483-487.

Turner, S.T., and Boerwinkle, E. (2000). Genetics of Hypertension, Target-Organ Complications, and Response to Therapy. *Circulation* **102(20 Suppl 4):** IV40-5.

Ungewitter, E., and Scrable, H. (2009). Antagonistic Pleiotropy and p53. *Mech Ageing Dev* **130(1-2):** 10-17.

Valenzuela, M.J., and Sachdev, P. (2006). Brain Reserve and Cognitive Decline: A Non-Parametric Systematic Review. *Psychol Med* **36(8):** 1065-1073.

van Dijk, E.J., Breteler, M.M., Schmidt, R., Berger, K., Nilsson, L.G., Oudkerk, M., Pajak, A., Sans, S., de Ridder, M., Dufouil, C. et al. (2004). The Association between Blood Pressure, Hypertension, and Cerebral White Matter Lesions: Cardiovascular Determinants of Dementia Study. *Hypertension* **44(5):** 625-630.

Voruganti, V.S., Nath, S.D., Cole, S.A., Thameem, F., Jowett, J.B., Bauer, R., MacCluer, J.W., Blangero, J., Comuzzie, A.G., Abboud, H.E. et al. (2009). Genetics of Variation in Serum Uric Acid and Cardiovascular Risk Factors in Mexican Americans. *J Clin Endocrinol Metab* **94(2):** 632-638.

Wechsler, D. (1981a). *WAIS-R Manual.* Psychological Corporation: San Antonio, TX.

Wechsler, D. (1981b). *The Wechsler Adult Intelligence Scale-Revised.* The Psychological Corporation: New York.

Weir, B.S. (1996). *Genetic Data Analysis II: Methods for Discrete Population Genetic Data.* Sinauer Associates: Sunderland, MA.

Winckler, W., Weedon, M.N., Graham, R.R., McCarroll, S.A., Purcell, S., Almgren, P., Tuomi, T., Gaudet, D., Bostrom, K.B., Walker, M. et al. (2007). Evaluation of Common Variants in the Six Known Maturity-Onset Diabetes of the Young (MODY) Genes for Association with Type 2 Diabetes. *Diabetes* **56(3):** 685-693.

Wolf, J.B., Pomp, D., Eisen, E.J., Cheverud, J.M., and Leamy, L.J. (2006). The Contribution of Epistatic Pleiotropy to the Genetic Architecture of Covariation among Polygenic Traits in Mice. *Evol Dev* **8(5):** 468-476.

Wolf, J.B., Leamy, L.J., Routman, E.J., and Cheverud, J.M. (2005). Epistatic Pleiotropy and the Genetic Architecture of Covariation within Early and Late-Developing Skull Trait Complexes in Mice. *Genetics* **171(2):** 683-694.

Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W. et al. (2010). Common SNPs Explain a Large Proportion of the Heritability for Human Height. *Nat Genet* **42(7):** 565-569.

Zeggini, E., Scott, L.J., Saxena, R., Voight, B.F., Marchini, J.L., Hu, T., de Bakker, P.I., Abecasis, G.R., Almgren, P., Andersen, G. et al. (2008). Meta-Analysis of Genome-Wide Association Data and Large-Scale Replication Identifies Additional Susceptibility Loci for Type 2 Diabetes. *Nat Genet* **40(5):** 638-645.

Zhong, H., Yang, X., Kaplan, L.M., Molony, C., and Schadt, E.E. (2010). Integrating Pathway Analysis and Genetics of Gene Expression for Genome-Wide Association Studies. *Am J Hum Genet* **86(4):** 581-591.

Zhou, S., Stone, E.A., Mackay, T.F., and Anholt, R.R. (2009). Plasticity of the Chemoreceptor Repertoire in Drosophila Melanogaster. *PLoS Genet* **5(10):** e1000681.

Zou, L., Sriswasdi, S., Ross, B., Missiuro, P.V., Liu, J., and Ge, H. (2008). Systematic Analysis of Pleiotropy in C. Elegans Early Embryogenesis. *PLoS Comput Biol* **4(2):** e1000003.

Chapter 5

Conclusions and Future Directions

In this dissertation, we focused on elucidating multiple attributes of the genetic architecture of leukoaraiosis through three different types of studies. In the first study (Chapter 2), we utilized a candidate gene approach to understand the contribution of single gene effects on mean levels of leukoaraiosis as well as gene-risk factor and gene-gene interactions. In the second study (Chapter 3), we investigated genetic variation across the entire genome to identify chromosomal regions associated with additive effects on leukoaraiosis. In the final study (Chapter 4), we investigated the genetic correlations among leukoaraiosis and seven measures of cognitive function by using both measured (GWAS) and unmeasured (biometrical) genetic approaches. In Chapter 5, we first present a summary of the findings of this dissertation including an integration of findings across all three research studies and a discussion regarding the similarities and differences between whites and African Americans for GWAS results, heritabilities, and biometrically estimated genetic and environmental correlations. Next, we give an overview of the current state of knowledge of the genetic architecture of complex traits in humans and animals, noting how this dissertation research relates to and intersects with this growing body of knowledge. Finally, we offer promising avenues for furure research on the genetic architecture of leukoaraiosis and other complex traits.

Summary of Findings

Leukoaraiosis is a highly heritable trait that is not well predicted by any clinical measures or covariates except for hypertension and age. Although leukoaraiosis has a high heritability, strong evidence for genetic variation associated with leukoaraiosis is limited, and to date the only consistently replicated association is with the angiotensinogen-converting enzyme (*ACE*) insertion/deletion polymorphism. Meta-analysis of nine studies with a combined total of 2316 subjects found that the *ACE* deletion-deletion genotype had an association with leukoaraiosis with an OR of 1.95, 95% CI=1.09-3.48 (Paternoster, 2009). Although there has been much interest in the potential association between the apolipoprotein E (*APOE*) ε4 allele and increased leukoaraiosis due to the relationship between ε4 and both Alzheimer's disease (AD) and heart disease, most studies have found no association between ε4 and leukoaraiosis (Paternoster, 2009). In this section, we compare and discuss the results from the association studies of leukoaraiosis in GENOA whites and African Americans presented in Chapters 2, 3, and 4 and review the evidence for association between genetic variation in *ACE* and *APOE* in the GENOA cohorts.

Differences in samples and methods used in Chapters 2, 3, and 4 analyses

In order to compare the results across chapters in this dissertation, it is first necessary to delineate the differences in the samples and analysis methods used in each of the chapters. The majority of the differences in samples and methods are due to decisions made collaboratively with researchers analyzing leukoaraiosis and cognitive measures in other cohorts that we were interested in using for meta-analysis or for replication with

GENOA. For all of the analyses of leukoaraiosis, individuals were excluded if they had unusable MRIs or evidence of previously undetected stroke. The GWAS conducted in Chapter 4 also excluded individuals that were less than 45 years of age and individuals that had measures of leukoaraiosis but did not participate in cognitive testing, since the main focus of the Chapter 4 GWAS analysis was older-age decline in cognitive function. In addition, the sample varied across chapters depending on the genotype data available for analysis (candidate gene genotypes for Chapter 2, Affymetrix 6.0 genotypes for Chapter 3, and Affymetrix or Illumina genotypes for Chapter 4).

GENOA whites were included in the analyses conducted in Chapters 2, 3, and 4, while African Americans were included only in Chapters 3 and 4. We chose to perform the candidate gene study (Chapter 2) in GENOA whites and not African Americans because we had no means to control for population admixture in the African American sample at the time this research was conducted, since neither ancestry informative markers nor genome-wide genotypes had been collected for African Americans. A slightly different method was used to obtain principal components of genotypes to control for population admixture in African Americans in the Chapters 3 and 4 analyses. In Chapter 3, we used genotype data from the Affymetrix 6.0 platform only and obtained principal components from genome-wide genotypes using all of the SNPs on the Affymetrix platform and all of the individuals in our sample. In Chapter 4, we had genotype data from both the Affymetrix and Illumina platforms for African Americans. Therefore, we obtained principal components from genotme-wide SNP genotypes that were present on both of these platforms (approximately 200,000 SNPs), and we used a sample of unrelated

individuals in order to prevent artefacts from the sibship structure of our sample from affecting the principal components. We did not need to control for population admixture in whites because previous analyses with the Structure program (Pritchard, 2000) showed that there was very little African or Asian ancestral genotypes present in the whites.

The adjustment model for leukoaraiosis also varied slightly by chapter. In Chapter 2, we transformed leukoaraiosis using the natural logarithm plus 0.0001, while in the other chapters we used the natural logarithm plus 1. Adding a constant to the value of leukoaraiosis (either 0.0001 or 1) was done to prevent taking the natural logarithm of zero, and since the majority of leukoaraiosis values were substantially greater than zero, the effect of using two different constants is likely minimal. For all analyses, age and gender were used as adjustment covariates for leukoaraiosis, but education was also included as an adjustment covariate in Chapter 4 to retain consistency with adjustment of the other outcomes (cognitive measures), many of which have a strong association with education. Finally, a measure of head or brain size was also used as an adjustment covariate for leukoaraiosis in every analysis. Brain size was used in Chapter 2, while total intracranial volume was used in Chapter 3. Adjusting volume of leukoaraiosis by brain volume makes intuitive sense, since a larger volume of brain tissue would be affected in individuals with larger brains given the same percentage of affected tissue. However, the brain shrinks both with age and with increasing pathology of AD, so some researchers prefer to use a measure such as total intracranial volume which does not vary with age or disease process.

Association between ACE I/D and APOE ε4 polymorphisms in GENOA

Since the *ACE* I/D polymorphism has the only consistently replicated association with leukoaraiosis reported in the literature to date, we examined its association with the samples of GENOA whites and African Americans used in the Chapter 3 GWAS. In both whites and African Americans, the association between this polymorphism and the residual value of leukoaraiosis after adjustment for age, sex, total intracranial volume, and the top 10 principal components (in African Americans only) was not significant (p-value = 0.903 in whites and 0.935 in African Americans).

In Chapter 3, we chose to perform the meta-analysis of African Americans from GENOA and ARIC after removing individuals with at least one copy or with two copies of the *APOE* ε4 allele in order to reduce the genetic heterogeneity of the sample with respect to AD susceptibility, since ε4 has a strong association with the development of AD. Since the relationship between ε4 and leukoaraiosis is unclear, we conducted this analysis by treating ε4 both as a dominant allele (removing individuals with ε4/- genotypes) and as a recessive allele (removing only individuals with ε4/ε4 genotypes). In the full GENOA samples with candidate gene data, the allele frequency of ε4 was 0.145 in whites and 0.226 in African Americans, and the genotype frequencies were as follows: 0.019 ε4/ε4, 0.253 ε4/-, and 0.728 -/- in whites, and 0.045 ε4/ε4, 0.361 ε4/-, and 0.594 -/- in African Americans. The *APOE* ε4 allele showed no association with residual values of leukoaraiosis in the white and African American samples used in the Chapter 3 GWAS (p-value = 0.365 in whites and 0.380 in African Americans).

The removal of individuals with an ε4 genotype assumes that there is interaction between ε4 and other polymorphisms that affects variation in leukoaraiosis. Since conducting a test of interaction between ε4 and all other polymorphisms is not practical, a strategy for determining whether individuals with ε4 alleles should be removed from a GWAS analysis could be based on the premise that the most common types of interaction effects tend to also show at least marginally significant main effect associations (Kooperberg, 2008). Based on this, an appropriate strategy would be to first test for association between leukoaraiosis and ε4, select the genetic model that best fits the association results (treating ε4 as a dominant or recessive allele), and remove people from the sample accordingly if there is evidence of a main effect. For meta-analyses across studies in which the raw data are not shared among investigators, the association between leukoaraiosis and ε4 would need to be performed separately within each cohort, and an appropriate strategy for removing individuals should be discussed by collaborators based on the results of the association studies.

GWAS results for candidate gene variants

The candidate gene study in Chapter 2 focused on examining SNP main effects and context-dependent effects (both SNP-covariate interactions and SNP-SNP interactions) associated with leukoaraiosis in GENOA whites. Since this study was performed before genome-wide genotyping was common, we were unable to secure a replication sample that had both measures of leukoaraiosis as well as genotypes for the candidate gene SNPs that we examined. In the absence of a replication sample, we used a filtering approach to reduce false positive associations. Of the 1649 SNPs from 268 genes known or

hypothesized to be involved in arteriosclerosis and related pathways, we found four SNPs that were able to predict greater than half of a percent of variation in leukoaraiosis in testing samples using cross-validation, had consistent effects in two internal replication samples, and had a false positive rate < 0.3. These SNP main effects were found in factor III (*F3*) which encodes a blood clotting factor, KIT tyrosine kinase ligand (*KITLG*) which is involved in hematopoietic stem cell proliferation, calpain 10 (*CAPN10*) which has been implicated in pathways associated with obesity and diabetes, and matrix metalloproteinase 2 (*MMP2*) which encodes an extracellular matrix protein.

In Chapters 3 and 4, we were able to assay the entire genome for association with leukoaraiosis in both whites and African Americans, and the criteria for significance became much steeper due to the multiple testing issues involved in performing so many association tests. The p-values for the four significant associations in the Chapter 2 candidate gene study ranged from 0.0001 to 0.0032, which are not low enough to even be considered worthy of further investigation in a GWAS due to the large number of false positive associations that will be in this range of p-values when conducting millions of association tests. An investigation of the GWAS results for leukoaraiosis from whites and African Americans in the Chapter 4 GWAS analysis showed that the four SNPs from the candidate gene study and the SNPs in nearby chromosomal regions generally did not obtain a noteworthy level of significance in these studies.

The SNPs in the *F3* and *CAPN10* genes that showed association in Chapter 2, rs3917643 and rs7571442, were not in the set of imputed SNPs used in the GWAS in Chapter 4. No

SNPs within 50kb of either SNP had a p-value < 0.001 in the set of imputed results for either whites or African Americans, though one SNP with p-value = 0.002 was observed 40kb from rs7571442 in whites. The SNP in *KITLG* that showed association in Chapter 2, rs995029, also showed association in the Chapter 4 GWAS (p-value = 8.7x10-4) in whites, and four SNPs within 50kb of the original SNP had a p-value < 0.001 in this sample. However, this SNP showed no association in the GWAS conducted in African Americans (p-value=0.88), and no SNPs within 50kb of the original SNP showed association. The *MMP2* SNP that showed association in Chapter 2, rs0028731, did not show association in the Chapter 4 GWAS in either whites (p-value = 0.2) or African Americans (p-value = 0.8), although 12 SNPs within 50kb of the original SNP had p-values < 0.002 in African Americans.

Comparison of Chapter 3 and Chapter 4 GWAS results

Appendix 5.1 presents a graphical summary of the chromosomal regions showing the strongest association with leukoaraiosis for whites and for African Americans, by indicating regions of 10Mb that contain at least one SNP with moderate evidence of significance (p-value < $1x10^{-4}$) in either the Chapter 3 or Chapter 4 GWAS. For whites, there was a large degree of similarity in the findings in the two GWAS conducted for leukoaraiosis (most of the regions containing a moderately associated SNP in the Chapter 3 GWAS also had a moderately associated SNP in the Chapter 4 GWAS). Of the 49 regions showing moderate significance in the Chapter 3 GWAS, 34 (69.4%) also showed moderate significance in the Chapter 4 GWAS. This relatively high degree of similarity is expected, since the sample did not change very substantially between Chapter 3 and

Chapter 4 and the additional adjustment variable in Chapter 4, education, was not associated with leukoaraiosis.

For African Americans, there is a much larger discrepancy between the regions with moderately associated SNPs in GWAS from Chapters 3 and 4. Of the 61 regions showing moderate evidence of association in the Chapter 3 GWAS, only 22 (36.1%) also had moderate evidence of association in the Chapter 4 GWAS. This discrepancy between results of the two GWAS is likely due to two factors. First, the sample size of African Americans increased substantially from N=553 in Chapter 3 to N=720 in Chapter 4 due to the addition of the ARIC participants genotyped on the Affymetrix platform as well as individuals genotyped on the Illumina platform. This may have affected the results if the participants differed in levels of leukoaraiosis or the adjustment covariates compared to the participants gneotyped by GENOA on the Affymetrix platform. What likely had the largest effect on the GWAS results, however, was the difference in the methods for calculating the principal components of genotype data in Chapters 3 and 4.

Comparison of GWAS results in whites and African Americans

The final table in Appendix 5.1 shows chromosomal regions that had moderate evidence of association (at least one SNP with p-value < $1x10^{-4}$) in either the Chapter 3 or Chapter 4 GWAS for both whites and African Americans. Thirty-four 10Mb regions across the genome showed moderate evidence of association with leukoaraiosis in both whites and African Americans. To determine whether the SNPs in 10Mb regions showing association in whites and African Americans were close enough on the genome to be

marking variation in the same chromosomal location, instances in which there were SNPs within 1Mb of each other (one in whites and one in African Americans) were noted. Of the 34 10Mb regions that showed moderate evidence of association in both groups, 14 (41.2%) had SNPs within 1Mb of each other in whites and African Americans. This indicates that these regions are worthy of further follow-up and fine mapping to determine whether a specific gene or genetic variant may be responsible for the signal detected in this region in both ethnic groups. Of note, none of the regions that showed evidence of association in both whites and African Americans contained any of the four candidate genes with genetic variation associated with leukoaraiosis in Chapter 2, the *APOE* gene, or the *ACE* gene.

Overall conclusions from GWAS and potential clinical relevance of findings

Fourteen chromosomal regions show evidence of association with leukoaraiosis in both GENOA whites and African Americans. Further exploration through fine mapping strategies and careful evaluation of the linkage disequilibrium structures in whites and African Americans in these regions will be necessary to determine whether there is a true genetic signal or whether these signals are simply artefacts of the analysis that do not represent true genetic singal. For example, the region on chromosome 6 from 30Mb to 40Mb from the start of the chromosome showed association with leukoaraiosis in both whites and African Americans. This region harbors the Major Histocompatability (MHC) locus, which is known to vary substantially across populations and tends to lead to false positive associations when population stratification or admixture is accounted for completely.

Though follow-up analysis is warranted for several genomic regions identified in this dissertation as potential candidates for association with leukoaraiosis, the lack of strong signal and inconsistency among analyses in this dissertation and in the literature indicate that we currently have a very limited understanding of the genes and genetic variants within those genes that affect leukoaraiosis across a range of age and population groups. It is clear that until further studies confirm a relationship between specific genetic variants and leukoaraiosis, the application of findings from association studies to clinical practice can not take place. Even if further studies identify that genetic variants do, in fact, have a true association with leukoaraiosis, the clinical utility of these findings is likely to be limited given the recent research regarding the predictive ability of genetic variants associated with other complex disease traits.

A recent study by Ripatti et al. (2010) examined the predictive ability of a set of 13 SNPs that have shown replicated association with coronary heart disease (CHD). In this study, the authors constructed a genetic risk score (GRS) by summing the number of risk alleles (0, 1, or 2) at each of the 13 loci and used Cox proportional hazards modeling to examine the association between the GRS and time to coronary heart disease, cardiovascular disease, and myocardial infarction during a 10-year follow-up of 30,725 participants. The predictive ability of the GRS for CHD traits was on par with the predictive ability of traditional risk factors such as LDL cholesterol and systolic blood pressure, with hazard ratios in the range of 1.5 to 2.0 when comparing the top and bottom quintiles. Though the GRS was able to identify the 20% of individuals with 70% increased risk of a first CHD

event and was still associated with CHD events after adjustment for traditional risk factors and family history of CHD, including GRS in the models did not confer additional predictive ability beyond traditional risk factors when comparing area under the curve (0.872 vs. 0.871, p-value=0.35).

The results from Ripatti et al. (2010) mirror the behavior of GRS in the majority of the other complex disease traits reported in the literature. Combining replicated SNP associations, each explaining a very small proportion of variation in disease risk, into a GRS typically demonstrates predictive ability but does not improve risk prediction beyond measurable risk factors. One of the primary goals of constructing GRS is to identify individuals who may benefit from earlier or more intenstive intervention, but GRS will be of little clinical utility in this context if the predictive ability over traditional risk factors is not substantial (Sandhu, 2010). However, Sandhu et al. (2010) point out that for common traits with high heritability, such as leukoaraiosis, discovery of all or many of the loci that increase susceptibility may prove to have clinically useful implications, and that the identification of these variants may be possible with increased sample sizes and improving technologies and research strategies for GWAS.

Differences in heritabilities and genetic correlations in whites and African Americans

In Chapter 4, the heritabilities of leukoaraiosis and cognitive measures after adjustment for age, sex, education, and total intracranial volume (in leukoaraiosis only) showed a consistent pattern in whites and African Americans. DSST had the highest heritability (0.883 in whites, 0.556 in African Americans), Stroop difference had the lowest

heritability (0.275 in whites, 0.154 in African Americans), and leukoaraiosis and the other cognitive measures had mid-range heritabilities (0.35-0.53 in whites, 0.33-0.54 in African Americans). However, heritabilities showed a clear trend of being lower in African Americans, with COWA FAS being the only notable exception (heritability = 0.366 in whites, 0.536 in African Americans). A similar trend was observed for overall phenotypic correlations and genetic correlations, with African Americans showing a similar but weaker correlational structure.

There are several reasons that could account for the lower observed heritabilities in African Americans. Since heritability is the fraction of the total variability of the trait accounted for by additive genetic factors, lower heritabilities could result from greater trait variation (a larger denominator) or from a smaller contribution from additive genetic effects (a smaller numerator). Differences in either the numerator or the demoninator could be due to true population differences between whites and African Americans, or they could simply be artefacts of the GENOA samples, such as differences in age or family structure.

To explore the possibility that African Americans have a greater amount of variation in leukoaraiosis than whites, we used Levene's test for equality of variances (Levene, 1960) to test for homogeneity of the variances of leukoaraiosis before and after adjusment for covariates in whites and African Americans. This test was selected because it does not rely heavily on the assumption of normality. The mean, trimmed mean, or median can serve as the parameter of interest for Levene's test, but the median performs best when

the trait distribution is skewed (Brown, 1974). Using each of these parameters produced similar results, so only Levene's test using the median as the parameter of interest is reported here. This test detected significant heterogeneity in the variances of raw leukoaraiosis (p-value = 1.04×10^{-7}), ln(leukoaraiosis+1) (p-value = 3.54×10^{-6}), and ln(leukoaraiosis+1) after adjustment for age, sex, education, and total intracranial volume (p-value = 1.28×10^{-10}) between whites and African Americans. The evidence for heterogeneity in variation of leukoaraiosis between whites and African Americans actually increases after adjusting for covariates including age, and the ratio of the variances also increases (ratio of variances of ln(leukoaraiosis) = 0.565 before adjustment, 0.654 after adjustment). Therefore, the African American sample does have a larger degree of variation in this trait than whites, which is not due to age differences between the white and African American samples (mean age = 61.3 years in whites, 63.3 years in African Americans). This suggests that age differences alone are unlikely to account for the lower heritability estimates observed in African Americans.

To examine whether differences in the family structure of GENOA samples of whites and African Americans accounted for the lower heritabilities observed in African Americans, we created new samples that had the same distribution of sibship sizes. In the original samples used to calculate heritabilities in Chapter 4, the African American sample had a larger number of singletons (124 singletons in the white sample, 220 in the African American sample), and the white sample had three sibships larger than six individuals while the African American sample had none. For sibships of each size, the maximum possible number of sibships were selected from each group to create "reduced" samples

of whites and African Americans. Appendix 5.2 shows the pedigree structures of the original and reduced samples and the heritabilities of leukoaraiosis and the seven cognitive traits estimated using each of the samples. Heritabilities calculated using the reduced samples were very similar to those calculated using the original samples. For instance, the heritability of ln(leukoaraiosis) adjusted for age, sex, education, and total intracranial volume was 0.529 in the original sample of whites, 0.518 in the reduced sample of whites, 0.432 in the original sample of African Americans, and 0.417 in the reduced sample of African Americans. The negligible change in heritability estimates in the reduced samples indicates that differences in family structure of the white and African American GENOA samples is not responsible for the observed lower heritabilies of the traits observed in African Americans.

Since neither age nor family structure are responsible for lower heritability estimate of leukoaraiosis in GENOA African Americans, it is likely that the lower heritability reflects a true difference in the population parameters of the groups studied and is not an artefact of the analysis. GENOA African Americans exhibit a greater amount of variability in leukoaraiosis than GENOA whites, and additive genetic effects explain a smaller proportion of the variability in African Americans. The lower heritability of leukoaraiosis may be due to a more prominent role of non-genetic (environmental) factors affecting leukoaraiosis variation in African Americans. This may occur if the African Americans had greater variability in these important non-genetic factors than whites (for example, if factors related to socioeconomic status affects variation in leukoaraiosis and African Americans have a larger range of socioeconomic status than

whites). The lower heritability in African Americans may also be due to context-dependent effects having a greater effect on leukoaraiosis in African Americans. If gene-gene or gene-environment interactions play a larger role in the genetic architecture of leukoaraiosis in African Americans, heritabilities may be lower because variation due to additive genetic effects that do not account for variation due to interactions.

Lower heritabilities of leukoaraiosis and the seven cognitive traits examined in Chapter 4 as well as differences in the genetic and environmental correlations between GENOA whites and African Americans suggest that non-genetic factors have a greater effect in African Americans than in whites for all of the brain traits studied. Though similar patterns were observed in whites and African Americans, genetic correlations (evidence of shared genetic effects) among cognitive traits tended to be higher and more significant in whites. While very little evidence of environmental correlation (shared environmental effects) between cognitive traits was observed in whites, eight of the 21 pairs of cognitive traits (38%) had evidence of significant environmental correlation in African Americans. Therefore, it is likely that non-genetic factors are indeed playing a larger role in affecting variation in brain traits in African American GENOA sample.

Knowledge about the Genetic Architecture of Common Chronic Conditions

Genetic architecture

The genetic architecture that underlies the variability of a trait includes the number of genes that affect the trait, the number of alleles within each involved gene as well as their allele frequencies and effect sizes on trait mean levels, the type of locus (regulatory, exonic, intronic, intergenic), the relative contribution of each allele to both the genetic

and overall phenotypic variances, and the variability (e.g. plasticity) and covariability (e.g. pleiotropy) when these genes are considered alone or when interacting with other genes and measures of the environment (both internal and external). A large amount of information about the genetic variations that impact complex human traits comes from association studies, particularly GWAS. GWAS and the resources for conducting them, such as the International HapMap Project (The International HapMap Consortium, 2003), are designed to identify trait-associated genetic variation that is relatively common in the population (with minor allele frequency (MAF) > 0.1), and are underpowered to detect variants that are rare (Manolio, 2008). The existence of common variants that are associated with risk of common disease is exemplified by the *APOE* ε4 allele, which is present at relatively high frequency (approximately 10%-25%) in many ethnic groups and clearly confers risk of cardiovascular disease and dementia (Chu, 2009; Dickstein, 2010; Reich, 2001). GWAS studies have been conducted for a wide variety of human traits, ranging from common chronic diseases to behavioral phenotypes to anthropometrics, and have provided researchers with useful and often surprising insights into the genetic architecture of human traits.

Allele frequencies and effect sizes of common variants identified in GWAS

With very few exceptions, the common variants found to affect human traits have very small effect sizes. An exploration and summarization of 531 associations between single nucleotide polymorphisms (SNPs) and traits from 151 published GWAS studies found that reported odds ratios for dichotomous traits ranged from 1.04 to 29.4, but that the median odds ratio was 1.33 (Hindorff, 2009), indicating that for the majority of traits,

common SNPs exhibit very modest effects. The risk allele frequencies of the SNPs found to be associated with traits in this survey of GWAS studies were variable, having a median of 36% and an interquartile range from 21%-53%. Very little data has been compiled regarding the effect sizes and allele frequencies from GWAS of quantitative traits. In our GWAS of leukoaraiosis, the effect sizes of the most significant SNPs were in the range of 0.23 to 0.86 standard deviations and their MAFs ranged from 2.3% to 49% (the median MAF was 20.4% in whites and 13.2% in African Americans). Given the way in which the first generation genotyping chips (e.g., Affymetrix® Genome-Wide Human SNP Array 6.0 and the Illumina® Human1M-Duo BeadChip) emphasized common alleles with excellent linkage disequilibrium (LD) coverage, it is likely that these statistics do not necessarily represent the attributes of the causative mutations but rather represent only the state of our knowledge to date.

Types of SNPs that have shown effects in GWAS

A somewhat surprising finding of GWAS has been the genomic locations of identified SNPs. Although protein-coding regions make up only about 1.5% of the genome (Lander, 2001), it was predicted that the majority of SNPs associated with traits and diseases would be found in these regions and that nonsynonymous SNPs would be particularly influential as causal variants. While it is true that nonsynonymous sites are significantly overrepresented in GWAS findings when compared to randomly selected SNPs on genotyping arrays, these account for only a small proportion of the significant and replicated associated variants(~9%) (Hindorff, 2009). SNPs in promoter regions, which affect gene transcription and thus the amount of gene expression, were also significantly

overrepresented but account for only 2% of associated SNPs. Synonymous SNPs, found in gene coding regions but with no effect on the amino acid sequence of proteins, accounted for an additional 2% of associated SNPs but were not overrepresented. The remainder of associated SNPs in genic regions were intronic, accounting for 45% of associated SNPs. Almost half of the SNPs identified as associated with human traits and diseases were from intergenic regions (43%), which is surprising because until recently it was assumed these intergenic regions had very little effect on gene expression and were largely non-functional. In GWAS and meta-analysis of leukoaraiosis in whites and African Americans presented in Chapter 3, 45 SNPs had p-values < $1x10^{-5}$. Of these SNPs, 13 SNPs (71.1%) were intergenic (outside the boundaries of the coding region of a gene) and the remainder (28.9%) were intronic..

Gene deserts harbor common variants that may affect regulation of gene expression at the level of transcription

Gene deserts, which are long stretches of DNA that contain no protein-coding genes, make up approximately 25% of the genome (Venter, 2001). The surprising discovery from GWAS that gene deserts often contain SNPs that have a replicated impact on human traits such as colorectal cancer (Tenesa, 2009) and Crohn's disease (Mathew, 2008) spurred interest in examining what mechanistic effect these SNPs may be having on gene expression. One of the main hypotheses is that gene deserts contain regulatory regions that play a role in determining expression of the genes in distant but flanking regions (Taylor, 2005). The theory posits that flanking genes are physically positioned next to these regulatory regions through the packaging of DNA in the nucleus, and that the

regions exert their effects through blocking or enhancing the transcriptional machinery. Evidence for this hypothesis includes the finding that some gene deserts contain evolutionarily conserved sequences. One of the earliest findings in this area of research was the presence of five enhancer sequences between 225kb and 780kb in a gene desert upstream of the *DACH1* gene, expressed in organ development, that are evolutionarily conserved in human, mouse, frog, and fish (Nobrega, 2003). The Encyclopedia of DNA Elements (ENCODE) project also showed that some genes have alternative promoter sequences up to several hundred kilobases away from their known transcription start sites, which may also constitute some of the associated SNPs in intergenic regions (ENCODE Project Consortium, 2007).

Gene deserts may harbor miRNAs, which affect trait variation at the level of translation

Until recently, it was assumed that the only regions of the genome that were transcribed, at least to a large degree, were protein-coding regions. Surprisingly, the ENCODE project revealed that up to 90% of the human genome is transcribed, including many non-coding regions that were thought to be silent (ENCODE Project Consortium, 2007). This led to the discovery of the importance of microRNAs (miRNAs) in regulating gene expression. Over 700 miRNAs, evolutionarily conserved noncoding RNAs of about 25 nucleotides, have been discovered in the human genome (Anglicheau, 2010). MiRNAs are now recognized to play a role in post-transcriptional gene expression by mechanisms such as repressing translation or degrading coding mRNA (Shyu, 2008). It is believed that each miRNA regulates hundreds of targets and that they also interact competitively and

synergistically with each other (Anglicheau, 2010). There is already preliminary evidence that miRNAs play a role in human diseases such as cardiovascular disease (Diez, 2009) and cancer (Croce, 2008).

Pleiotropy has been detected in studies of GWAS findings

Another interesting result of the current GWAS era is that it is providing insight into the frequency and mechanisms underlying pleiotropy. An assessment of 118 GWAS articles containing 56,411 significant associations conducted by Johnson et al. (2009) revealed that several SNPs and a large number of genes were associated with multiple traits. One SNP in the *APOC1* gene, for example, was associated with 11 traits ranging from Alzheimer's disease (AD) to coronary heart disease. Johnson and colleagues also found that 1% of genic regions, based on 100kb bins, exhibited associations with at least 13 traits, many of which appear to be unrelated based on current biological knowledge. For example, one bin on chromosome 2 was associated with 87 traits ranging from abdominal aortic calcification to bipolar disorder. Another interesting finding from this study of GWAS results was that the protein-coding genes shown to be associated with human diseases and traits were significantly overrepresented in Gene Ontology categories such as cell adhesion, signal transduction, ion transport, and protein phosphorylation. These proteins are key regulators of multiple downstream molecular and physiological pathways, making them key candidates for pleiotropic gene effects. Studies of pleiotropy using GWAS findings are just beginning, and their methodologies have yet to be fully developed. In this dissertation, we began to examine the utility of using the percentage of shared SNP associations at a nominal significance level (α=0.01) as a metric for

assessing pleiotropy among leukoaraiosis and cognitive traits. While we were unable to draw strong conclusions due to our relatively small sample sizes, we did find preliminary evidence that percentage of shared SNP associations from GWAS may have value as a metric of pleiotropy because it has a much stronger relationship with biometrically estimated genetic correlation than environmental correlation. This indicates that further investigation of this method for assessing pleiotropy is an endeavor that is worth pursuing.

Knowledge of Genetic Architecture from Studies of Model Organisms

Genetic studies in model organisms

In many ways, our understanding of the complex genetic architecture of common traits in humans is still years behind what has been demonstrated in model organisms. Model organisms represent a much more accessible avenue for studying genetic variation, since the researcher is able to experimentally homogenize the genetic and environmental backgrounds of the organisms in order to assess the effects of single genetic variations in isolation. Studies conducted in animal models, particularly of quantitative trait loci, have revealed several consistent findings regarding the genetic architecture of complex phenotypes, and many of the observations parallel or extend the conclusions drawn from human studies. The lessons learned from studying behavioral, anthropomorphic, and other quantitative traits in model organisms may provide insight into understanding how genetic variation affects human behavior at multiple levels: DNA sequence, transcription, protein abundance, metabolism, and in differing environmental contexts (Mackay, 2009).

Allele frequencies, effect sizes, and the polygenic model

In a recent review of the genetic architecture of quantitative traits in flies, mice, and humans, Flint and Mackay note the consistent observation that there tend to be large numbers of loci, each with small effects, that impact single traits (Flint, 2009). Allelic effects in animal models appear to follow an exponential distribution in which there are only a few loci with large to moderate effects, and increasing numbers of loci with smaller effects, and it appears that this is true for almost all traits studied. Flint and Mackay attest that this seems to be the pattern observed in humans also, as large effects in GWAS studies appear to be the exception rather than the rule and a number of genes with small effects have been observed for quantitative traits such as height (Flint, 2009; Gudbjartsson, 2008). In addition, many of the quantitative trait loci (QTLs) detected in animal studies tend to follow an additive polygenic model (Kenney-Hunt, 2006), and there is evidence that this also true for some human traits such as height (Visscher, 2008).

Different alleles at the same locus may have differential effects

An interesting finding from model organisms is that there may be another layer of underlying complexity to the mechanisms in which allelic variants affect quantitative traits. Drosophila geneticists have begun to use association mapping in order to determine the molecular basis of allelic variation, and are now studying quantitative trait nucleotides (QTNs) at QTLs. It has been consistently shown that differing alleles at the same locus do not have consistent effects on particular traits (Mackay, 2009). For example, two independent P-element insertions into the exactly the same genetic location but at opposite orientations in Drosophila have very different phenotypic effects: one

increases trehalose sensitivity and decreases starvation resistance while the other has no effect on trehalose sensitivity but increases starvation resistance (Rollmann, 2006). An example of the relevance of these findings to human genetics is that we may find that it is not only the number of copies of an allele that matters at a CNV locus, but also the precise orientation and constitution of the genetic variants within the CNV.

Gene-environment and gene-gene interaction are common

Animal studies have also shown that context-dependence (gene-environment interaction and epistasis) plays a key role in almost all phenotypes studied (Flint, 2009). Interactions between QTLs and sex are a very common finding in model animal studies, with a large fraction of detected QTLs (approximately 19% to 50% depending on the trait studied) exhibiting sex-specific effects (Kenney-Hunt, 2006; Cheverud, 2001). For example, a recent study on Drosophila life span found that 41 of 58 detected QTLs (70.7%) showed differential trends in males and females, and furthermore that 16 of 21 statistically significant epistatic interactions (76.2%) also exhibited differential sex effects (Magwire, 2010). In a study of adiposity in mice, Cheverud et al. (2001) observed that the effect sizes for individual QTLs and epistatic effects were of similar magnitude (about one-third of a standard deviation unit, explaining an average of 5% of the phenotypic variance). In this study, accounting for epistatic effects between adiposity QTLs doubles the proportion of total variance of adiposity explained in females (from 14% to 28%) and increases it from 27% to 35% in males.

Flint and Mackay believe that similar patterns of context-dependence are likely to also exist in humans, but that the probitively large sample sizes needed to detect these findings in humans is the reason why findings have not been consistently replicated similar to main effects of SNPs (Flint, 2009). However, candidate gene studies on human traits that have examined context-dependence through statistical tests of interaction or biometrical modeling have demonstrated the presence of both sex specific and epistatic effects. For example, Reilly et al. (1994) showed that alleles at the *APOE* locus exhibited gender-specific influences on the correlations and covariances of nine plasma lipid protein traits (*APOE* alleles had a significant influence on 27.8% of correlations and covariances in females, but only 8.3% in males). In addition to gender-specific effects, *APOE* alleles have also been shown to have epistatic effects in studies of human traits such as leukoaraiosis through interaction with polymorphisms in the *ACE* and *MTHFR* genes (Szolnoki, 2004). In the candidate gene study we conducted as part of this dissertation, epistatic effects had the largest cross-validated influence on trait variation. Four pairs of SNP-SNP interactions explained 9.59% of the trait variation in independent test samples, while the four most highly associated SNPs explained only 3.72%.

Pleiotropy is common, and also displays gene-gene and gene-environment interaction

The large degree of epistasis and context-dependent effects extends beyond single allelic variants and also is found for pleiotropic effects (Mackay, 2009). As discussed in previous chapters, the degree of pleiotropy detected in QTL studies of Drosophila and mice studies is extensive for both anthropometric and behavioral phenotypes (Mackay, 2005). For example, a review of complex behavior in Drosophila indicates that large

numbers of pleiotropic genes interact in an epistatic fashion to regulate behavioral traits, and that these pleiotropic genes commonly have sex- and environment-specific effects (Mackay, 2009). There is evidence that the complexity of the genetic architecture of human traits also exhibits a high degree of pleiotropy and interactions. For example, most human behaviors are quantitative traits that are affected by complex networks of interacting genes that are context-dependent (Falconer, 1998), and similar rules for pleiotropic action are likely to shape these traits (Mackay, 2009). In our study of pleiotropy in leukarasaiosis and cognitive traits, we found a large degree of pleiotropy among the pairs of cognitive traits but very little pleiotropy between leukoaraiosis and individual cognitive traits. Fourteen of 21 cognitive trait pairs in whites (67%) and 4 of 21 cognitive trait pairs in African Americans (19%) had significant genetic correlations (ranging from 0.26 to 0.92), indicating that these share a large proportion of genetic effects. We also observed a strong relationship between biometrically estimated genetic correlations and the percentage of shared SNP associations at a nominal α level from GWAS (which ranged from 0.6% to 16.1% in trait pairs exhibiting significant genetic correlation), providing preliminary evidence that metrics can be developed to evaluate pleiotropy using GWAS findings.

Future Directions

Given the importance of understanding the genetic architecture of human traits, we have provided state-of-the-art reflection of the current approaches available to investigate genetic causes of variation in risk of common chronic conditions. As the technologies for measuring different types of genome-wide phenomena continually increase, there is a

concomitant increase in the complexity of next-generation of studies to explore this topic. In final last section of the dissertation, we review some of the most exciting arenas that are likely to inform the next set of research questions to elucidate the genetic architecture of leukoaraiosis. First, we discuss the potential role of epigenetic effects in trait variation. Next, we present potential new statistical methods for studying the heritability of disease traits and examining pleiotropy among traits. Finally, we discuss the necessity of moving toward a systems-level approach for integrating GWAS, epigenetic, and transcriptomic data in order to obtain a more complete picture of the genetic architecture of disease traits.

Epigenetic effects play a role in trait variation

Recently, the ENCODE project definitively demonstrated that transcriptional regulation of the DNA sequence involves a set of processes acting in concert, including modifications to histones and to the DNA itself (ENCODE Project Consortium, 2007). This brought the attention of researchers to another potential contributor to variation in human traits, termed epigenetic mechanisms, which collectively include histone modfication, DNA methylation, and miRNA. These are a set of heritable and non-heritable phenomena that play a key role in a variety of cellular processes and have been hypothesized as a link between environmental factors and chronic disease susceptibility (Waterland, 2009). Covalent modification of cytosine residues in DNA by methylation plays an active role in gene silencing (Egger, 2004), and there is evidence that changes in methylation at key sites affect transcription in a variety of human disease processes including cancer (Esteller, 2008), athersclerosis (Turunen, 2009), and functional

abnormalities in the brain (Wilson, 2008; Pogribny, 2009). Epigenomic technologies, developed to study epigenetic mechanisms on a genome-wide scale, include commercial array-based techniques and techniques based on restriction endonucleases, and second-generation technologies for massively parallel large-scale analysis are currently being developed along with analysis strategies for this exciting new type of data (Feinberg, 2010). Whole-genome methylation profiles (27,578 markers in 14,495 genes) have been obtained for GENOA participants using the Illumina® Infinium HumanMethylation27 BeadChip, and GENOA investigators have already begun to test for association between methylation profiles and target organ damage phenotypes. As a next step, we plan to explore pleiotropic effects by identifying methylation markers that are associated with multiple traits and by multivariately modeling the combined effects of SNPs and methylation markers on the genetic correlation of trait pairs.

New methods for estimating heritability and pleiotropy with genome-wide SNPs

In contrast to the classic biometrical approach in which the covariance between relatives is partitioned into genetic and environmental sharing based on knowledge of familial relationships, the methods used by Yang et al. (2010) begin to lay the groundwork for estimating heritability using genotyped markers in unrelated individuals. While Yang et al. do not estimate heritability per se, they estimate the variance of a trait explained by considering all measured SNPs simultaneously and then perform simulations to show that additional genetic variance may be captured by estimating the degree to which causal alleles are not in LD with measured SNPs. As the density of our genotypes improves vastly with the use of resources developed through the 1,000 Genomes Project to impute

~17 million SNPs, it is possible that we may be able to reliably estimate heritability based on SNP genotypes, even in unrelated individuals, using the methods of Yang et al.

Yang method can be extended to estimate genetic correlations among traits (pleiotropy)

The methods of Yang and colleagues could also be extended to estimate the genetic correlations among pairs of traits in order to assess pleiotropy. In the biometrical approach, the covariance matrix between related individuals is used to estimate the phenotypic, genetic, and environmental correlations for a pair of traits. Using a method analogous to determining the proportion of biometric heritability accounted for by simultaneously considering the effects of all measured SNPs, we can begin to examine the proportion of genetic covariance due to measured genes and the remaining covariance due to non-measured additive effects. If incomplete LD is taken into consideration and accounted for, this method can utilize measured genetic markers to estimate genetic correlation and will provide a means for examining pleiotropy in unrelated samples.

Systems-level genetics

Many loci with complex relationships are likely to contribute to the heritability of human traits

As Yang and colleagues demonstrated, the many loci contributing to additive genetic variance in traits are distributed across the genome (Yang, 2010). At least for height, there appear to be many loci with effects that are smaller than the threshold currently detectable by GWAS methods, and new methods for identifying these currently unidentifiable contributing loci are sure to emerge in the near future. If animal models

provide any indication of the vast number of loci that may be contributing to complex traits, the finding that over 20% of the genome was associated with Drosophila mating behavior (Mackay, 2005) suggests that the high heritabilities observed in human traits may be due to thousands or even hundreds of thousands of loci contributing small effects to trait variation. The further finding that many of the genes implicated in Drosophila mating behavior were also implicated in a variety of seemingly unrelated or tangentially-related biological pathways (Mackay, 2005) suggests that the GWAS paradigm of a single genetic variant associated with a single trait is not prepared to even begin to capture the complexity of the genetic architecture that underlies traits of interest. Clearly, more advanced methods of capturing genetic variation, such as those proposed by Yang et al. and more sophisticated ways of viewing the interconnectedness of traits and their underlying contributing factors are needed.

Integrating information from contributing loci will require a systems-level approach

Studies of quantitative trait loci and preliminary studies of pleiotropic effects in model organisms and humans have set the stage for the development of systems-level genetics approaches to examining the complex genetic architecture that impacts human health and trait variation. Systems-level genetic approaches aim to integrate information from a variety of sources, including genetic, transcriptomic, proteomic, and epigenetic data in order to highlight biological pathways that play a key role in trait variability and identify novel pathophysiological mechanisms contributing to disease (McKnight, 2010). In addition to providing detailed genetic maps of variants associated with specific traits, systems genetics approaches will also provide an opportunity to examine pleiotropy, to

extrapolate genetic networks from networks of correlated transcripts, and to provide functional annotation for genes with unknown function, computationally predicted genes, or genes located in gene deserts (Ayroles, 2009).

Systems-level genetics approaches have been used to study genetic architecture and pleiotropy in model organisms

Systems-level genetics has already infiltrated the field of genetics research in model organisms, and much has been learned about the genetic architecture of quantitative traits from these studies. For example, Mackay and colleagues (2009) have begun to develop a systems biology approach for studying the complex genetic architecture of quantitative traits in Drosophila that integrates information about DNA sequence, gene expression, biological pathways, and multiple phenotypes. Using this approach, whole genome transcript profiles were obtained for genetically divergent and control lines of Drosophila, and groups of co-regulated genes were identified that are affected by specific genetic variants. Through this line of research in multiple model organisms, it has become apparent that there is a modular organization of gene expression patterns that are associated with tissue-specific expression, transcription factor binding sites, and a variety of Gene Ontology categories, emphasizing the importance of studying the mechanisms contributing to pleiotropic effects (Magwire, 2010; Ayroles, 2009; Jumbo-Lucioni, 2010; Zou, 2008; Gunsalus, 2005; Ihmels, 2002; Han, 2004). Systems genetic work in model organisms has the potential to strongly guide and supplement our knowledge of the biological pathways underlying disease pathology in humans. For example, the work of Jumbo-Lucioni (2010) on the systems genetics of body weight and energy metabolism

phenotypes in Drosophila may help predict correlated changes in related traits with medical interventions in humans (Jumbo-Lucioni, 2010). Model organisms may also be used to help confirm findings in humans when biological pathways are found to be associated with similar traits. For example, systems genetics approaches have identified a gene expression module related to inflammation and immune response that is associated with obesity in both mouse and humans (Chen, 2008; Emilsson, 2008).

Applications of systems biology to human trait research

Pathways analysis using functional annotation for an integrated set of GWAS and gene expression data provides a starting point for initiating systems-level genetics approaches to studying complex traits in humans. For example, Zhong et al. (2010) used a novel approach of integrating publicly available GWAS and gene expression data to functionally characterize the pathways enriched for type 2 diabetes. Using this approach, they identified both known pathways (e.g., calcium signaling, *PPAR* signaling, and *TGF*-beta signaling) and novel pathways (e.g., tight junction, complement and coagulation, and antigen processing) associated with this trait that replicated in independent cohorts. Systems-level genetics approaches have also been used to prioritize candidate genes identified through GWAS for further functional validation based on their biological characterization. Hsu et al. (2010) used GWAS data combined with gene expression signature profiling from cellular and animal studies to prioritize candidate genes associated with skeletal metabolism. This method appears to have utility, as two of the three prioritized candidate genes replicated in subsequent studies, while none of the 16 non-prioritized candidate genes replicated. Indeed, it appears that the agnostic nature of

the genome-wide approach has provided clues to the discovery of unexpected mechanisms of the regulation of gene expression and pleiotropy, but that these findings will need to be explored with a careful consideration of the biology underlying their function in order to provide utility for understanding human trait variation.

Conclusion

The results from this dissertation assist in characterizing genetic polymorphisms and their interactions with each other and with environmental factors that are associated with inter-individual variation in leukoaraiosis, and begin to examine the extent of pleiotropy among leukoaraiosis and measures of cognitive function. Although it is clear that there is not enough information to make strong conclusions about SNPs that definitively affect leukoaraiosis at this time, these studies pave the way for future research that may be able to more thoroughly assess the multiple genetic and environmental factors that impact leukoaraiosis and cognitive function. There is another great wave of research opportunities ahead arising from next-generation sequencing, epigenomic assays, and transcriptional profiles that can build upon new methods for estimating heritability and integrating information through systems-genetics to elucidate the many genetic and environmental factors that influence variation in human traits including leukoaraiosis. As these methods are developed and implemented, a more thorough understanding of the underlying biology that affects health and disease of the brain may ultimately provide greater accuracy in predicting disease risk and more effective design of prevention and treatment strategies for stroke and dementia.

Appendix 5.1. Summary of findings from candidate gene and GWAS studies in this dissertation

The three figures below show a graphical summary of the association results with leukoaraiosis from the candidate gene study conducted in GENOA whites (Chapter 2) and the GWAS studies conducted in GENOA whites and African Americans in Chapter 3 and Chapter 4. Differences between the samples and methods for the GWAS studies conducted for leukoaraiosis in Chapters 3 and 4 are described in the text of Chapter 5 above.

On each figure, the regions that harbor the four SNPs that passed all of the filters in the candidate gene study (Chapter 2) are indicated by the letters "a, b, c," and "d". The region that harbors the apolioprotein E (*APOE*) gene is indicated by the letter "e". The region that harbors the angiotensinogen-converting enzyme (*ACE*) insertion/deletion polymorphism, the genetic variant with the strongest evidence of association with leukoaraiosis in the literature, is indicated by the letter "f".

The vertical axis of each figure depicts chromosomes 1-22, and the horizontal axis depicts 10Mb segments (bins) of the chromosomes, numbered 0-24. For example, the cell in the upper left-hand corner designates base pairs 1 through 9,999,999 from the start of chromosome 1, the cell immediately adjacent to the right designates base pairs 10,000,000 through 19,999,999 from the start of chromosome 1. Since the chromosomes differ in length, the number of 10Mb bins varies by chromosome, ranging from 25 bins for chromosome 1 to 5 bins for chromosome 22. The remainder of the bins are colored in gray. Chromosomal locations are annotated according the National Center for Biotechnology Information (NCBI) Genome Build 36.3 of the human genome reference sequence.

Chromosomal regions with evidence of association with leukoaraiosis in GENOA whites

Chr.	10 Mb chromosomal region																								
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
1		xo	o	o			o			a												o		o	c
2			xo			x				xo	x			xo						x	x	o		o	
3			xo						xo				xo			o		xo							
4	o		xo				xo					xo			xo	xo	xo	o	xo						
5	x							x							xo	o		x							
6	xo	xo	o	xo			xo	o		o															
7					o							x													
8											xo			xo											
9	o																								
10							xo			x															
11	o				o							o	x												
12			o		o				b																
13			xo		xo		x		x	xo	xo														
14						o				x															
15			x			xo																			
16						d		o																	
17						f																			
18							x	xo																	
19					e		xo																		
20	xo	xo			xo	o																			
21		o	xo																						
22																									

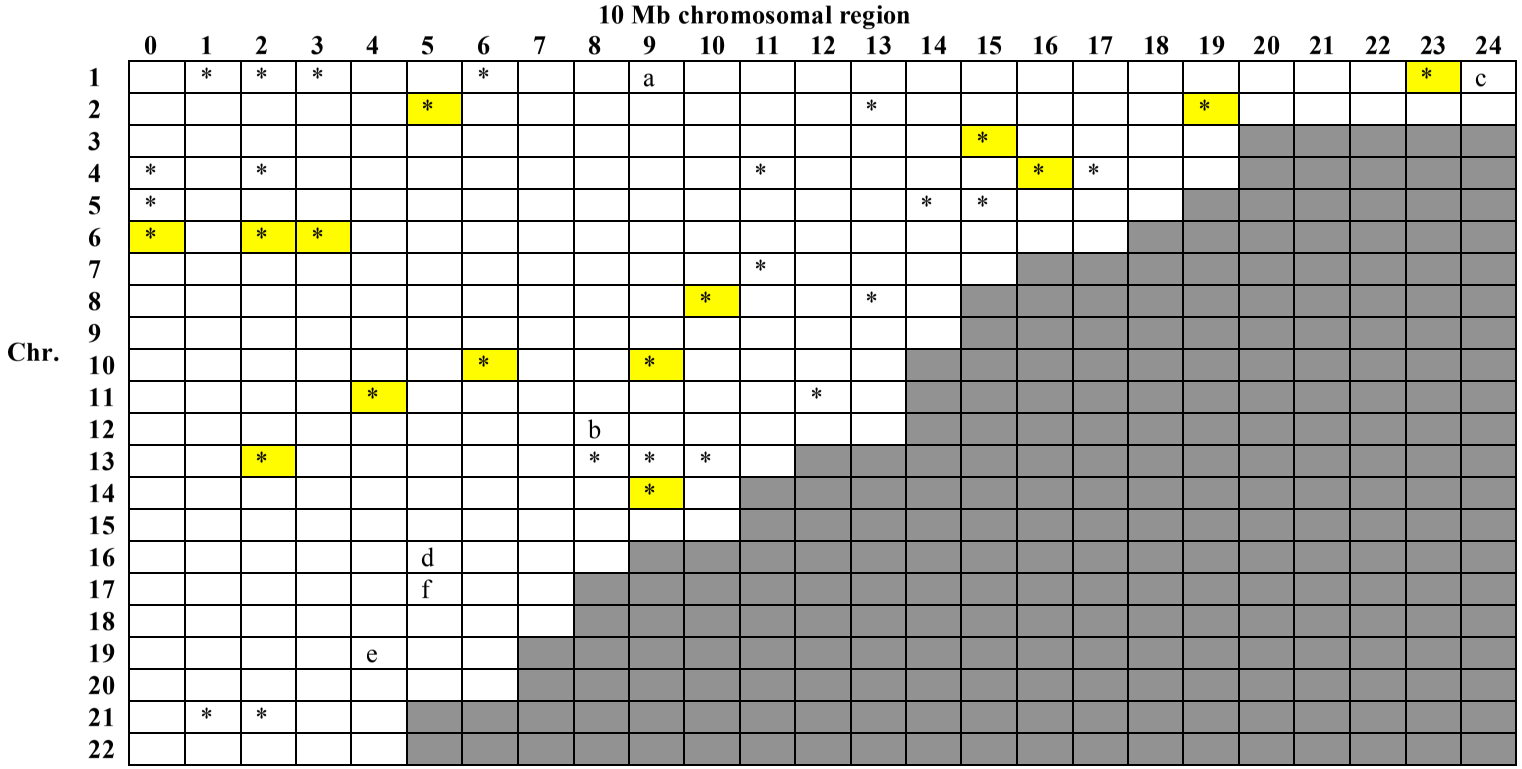
x = at least one SNP had p-value < $1x10^{-4}$ in the GWAS for leukoaraiosis presented in Chapter 3
o = at least one SNP had p-value < $1x10^{-4}$ in the GWAS for leukoaraiosis presented in Chapter 4
a = *F3*_rs3917643 (significant in candidate gene study, Chapter 2)
b = *KITLG*_rs995029 (significant in candidate gene study, Chapter 2)
c = *CAPN10*_rs7571442 (significant in candidate gene study, Chapter 2)
d = *MMP2*_rs9928731 (significant in candidate gene study, Chapter 2)
e = Apolipoprotein E (*APOE*)
f = Angiotensinogen (*ACE*)

Chromosomal regions with evidence of association with leukoaraiosis in GENOA African Americans

Chr.	10 Mb chromosomal region																								
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
1		o	x	o			o			a		xo					x	o	o		o		xo	xo	co
2		x		xo	o	xo					o		x	o	o		x			x			xo		
3		x		o			o				x	xo			o	x									
4	xo		x						x	o		o	o				o	x							
5	o											o	x		xo	o	x								
6	xo		o	o		o									o										
7	o	x				o	o		o	o	o	o			x	x									
8	xo	xo				x					x			o	o										
9									x	o	x														
10	o	o		x			xo	xo		x		xo	o	x											
11					x			o			o		x												
12	x			x			xo		bx	x	xo	o	x												
13			o					o	xo	x	o														
14			xo	xo						xo															
15				o				o		x															
16	o				x	dx																			
17		x		o	o	fxo	o																		
18																									
19					eo	o																			
20																									
21		x	x		o																				
22			o	x																					

x = at least one SNP had p-value < $1x10^{-4}$ in the GWAS for leukoaraiosis presented in Chapter 3
o = at least one SNP had p-value < $1x10^{-4}$ in the GWAS for leukoaraiosis presented in Chapter 4
a = *F3*_rs3917643 (significant in candidate gene study, Chapter 2)
b = *KITLG*_rs995029 (significant in candidate gene study, Chapter 2)
c = *CAPN10*_rs7571442 (significant in candidate gene study, Chapter 2)
d = *MMP2*_rs9928731 (significant in candidate gene study, Chapter 2)
e = Apolipoprotein E (*APOE*)
f = Angiotensinogen (*ACE)*

Chromosomal regions with evidence of association with leukoaraiosis in GENOA whites and African Americans



* = at least one SNP had p-value < $1x10^{-4}$ in the GWAS for leukoaraiosis presented in Chapter 3 or Chapter 4 in both whites and African Americans. Items in yellow indicate that the SNPs with p-value < 0.0001 in whites and African Americans were within 1Mb of each other.

a = *F3*_rs3917643 (significant in candidate gene study, Chapter 2)

b = *KITLG*_rs995029 (significant in candidate gene study, Chapter 2)

c = *CAPN10*_rs7571442 (significant in candidate gene study, Chapter 2)

d = *MMP2*_rs9928731 (significant in candidate gene study, Chapter 2)

e = Apolipoprotein E (*APOE*)

f = Angiotensinogen (*ACE*)

Appendix 5.2. Heritability of ln(leukoaraiosis) and cognitive traits in samples of equal family size in whites and African Americans

In order to determine whether the differences in sibship structures of the white and African American GENOA samples were responsible for the lower heritability estimates in African Americans, "reduced" samples of whites and African Americans were created so that each group had the same distribution of sibship sizes. For sibships of each size, the maximum possible number of sibships were selected from each group to create the reduced samples. The tables below show the sibship sizes of the original and reduced samples, as well as heritability estimates of adjusted traits calculated from each sample.

Pedigree structures of original and reduced samples of whites and African Americans

Number of Siblings in Sibship	Number of Sibships in Original Sample of Whites (N=762)	Number of Sibships in Original Sample of African Americans (N=720)	Number of Sibships in Reduced Samples (N=619)
8	1	0	0
7	2	0	0
6	2	2	2
5	7	8	7
4	19	19	19
3	47	44	44
2	176	120	120
1	124	220	124
Totals	**378 Sibships**	**413 Sibships**	**316 Sibships**

Comparison of trait heritabilities in original and reduced samples of whites and African Americans

Trait	h^2 (SE) Adjusted in Original Sample of Whites	h^2 (SE) Adjusted in Reduced Sample of Whites	h^2 (SE) Adjusted in Original Sample of African Americans	h^2 (SE) Adjusted in Reduced Sample of African Americans
Ln(leukoaraiosis+1)	0.529 (0.09)***	0.518 (0.10)***	0.432 (0.13)***	0.417 (0.14)***
RAVLT delayed recall	0.526 (0.10)***	0.574 (0.11)***	0.390 (0.11)***	0.379 (0.11)***
RAVLT total learning	0.516 (0.10)***	0.553 (0.11)***	0.440 (0.11)***	0.452 (0.12)***
DSST	0.843 (0.09)***	0.804 (0.10)***	0.556 (0.10)***	0.561 (0.10)***
COWA FAS	0.366 (0.10)***	0.394 (0.12)***	0.536 (0.11)***	0.572 (0.11)***
COWA animals	0.349 (0.10)***	0.379 (0.11)***	0.329 (0.10)***	0.329 (0.11)***
Stroop color word	0.429 (0.09)***	0.471 (0.11)***	0.440 (0.10)***	0.433 (0.10)***
Stroop difference	0.275 (0.09)***	0.139 (0.10)	0.154 (0.10)*	0.138 (0.10)

For all adjusted traits, biometric models included age, sex, and education. The biometric model for ln(leukoaraiosis+1) also included TIV.

Null hypothesis of tests: $h^2 = 0$
* $0.01 < \text{p-value} < 0.05$
** $0.001 < \text{p-value} < 0.01$
*** $\text{p-value} < 0.001$

References

Anglicheau, D., Muthukumar, T., and Suthanthiran, M. (2010). MicroRNAs: Small RNAs with Big Effects. *Transplantation* **90(2):** 105-112.

Ayroles, J.F., Carbone, M.A., Stone, E.A., Jordan, K.W., Lyman, R.F., Magwire, M.M., Rollmann, S.M., Duncan, L.H., Lawrence, F., Anholt, R.R. et al. (2009). Systems Genetics of Complex Traits in Drosophila Melanogaster. *Nat Genet* **41(3):** 299-307.

Brown, M.B., and Forsythe, A.B. (1974). Robust Tests for the Equality of Variances. *JASA* **69:** 364-367.

Chen, Y., Zhu, J., Lum, P.Y., Yang, X., Pinto, S., MacNeil, D.J., Zhang, C., Lamb, J., Edwards, S., Sieberts, S.K. et al. (2008). Variations in DNA Elucidate Molecular Networks that Cause Disease. *Nature* **452(7186):** 429-435.

Cheverud, J.M., Vaughn, T.T., Pletscher, L.S., Peripato, A.C., Adams, E.S., Erikson, C.F., and King-Ellison, K.J. (2001). Genetic Architecture of Adiposity in the Cross of LG/J and SM/J Inbred Mice. *Mamm Genome* **12(1):** 3-12.

Chu, A.Y., Parekh, R.S., Astor, B.C., Coresh, J., Berthier-Schaad, Y., Smith, M.W., Shuldiner, A.R., and Kao, W.H. (2009). Association of APOE Polymorphism with Chronic Kidney Disease in a Nationally Representative Sample: A Third National Health and Nutrition Examination Survey (NHANES III) Genetic Study. *BMC Med Genet* **10:** 108.

Croce, C.M. (2008). Oncogenes and Cancer. *N Engl J Med* **358(5):** 502-511.

Dickstein, D.L., Walsh, J., Brautigam, H., Stockton, S.D.,Jr, Gandy, S., and Hof, P.R. (2010). Role of Vascular Risk Factors and Vascular Dysfunction in Alzheimer's Disease. *Mt Sinai J Med* **77(1):** 82-102.

Diez, J. (2009). Do microRNAs Regulate Myocardial Fibrosis? *Nat Clin Pract Cardiovasc Med* **6(2):** 88-89.

Egger, G., Liang, G., Aparicio, A., and Jones, P.A. (2004). Epigenetics in Human Disease and Prospects for Epigenetic Therapy. *Nature* **429(6990):** 457-463.

Emilsson, V., Thorleifsson, G., Zhang, B., Leonardson, A.S., Zink, F., Zhu, J., Carlson, S., Helgason, A., Walters, G.B., Gunnarsdottir, S. et al. (2008). Genetics of Gene Expression and its Effect on Disease. *Nature* **452(7186):** 423-428.

ENCODE Project Consortium, Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigo, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T. et al. (2007). Identification and Analysis of Functional Elements in 1% of the Human Genome by the ENCODE Pilot Project. *Nature* **447(7146):** 799-816.

Esteller, M. (2008). Epigenetics in Cancer. *N Engl J Med* **358(11):** 1148-1159.

Falconer, D.S., and Mackay, T.F. (1998). *Introduction to Quantitative Genetics*. Essex, England: Addison Wesley Longman Ltd.

Feinberg, A.P. (2010). Genome-Scale Approaches to the Epigenetics of Common Human Disease. *Virchows Arch* **456(1):** 13-21.

Flint, J., and Mackay, T.F. (2009). Genetic Architecture of Quantitative Traits in Mice, Flies, and Humans. *Genome Res* **19(5):** 723-733.

Gudbjartsson, D.F., Walters, G.B., Thorleifsson, G., Stefansson, H., Halldorsson, B.V., Zusmanovich, P., Sulem, P., Thorlacius, S., Gylfason, A., Steinberg, S. et al. (2008). Many Sequence Variants Affecting Diversity of Adult Human Height. *Nat Genet* **40(5):** 609-615.

Gunsalus, K.C., Ge, H., Schetter, A.J., Goldberg, D.S., Han, J.D., Hao, T., Berriz, G.F., Bertin, N., Huang, J., Chuang, L.S. et al. (2005). Predictive Models of Molecular Machines Involved in Caenorhabditis Elegans Early Embryogenesis. *Nature* **436(7052):** 861-865.

Han, J.D., Bertin, N., Hao, T., Goldberg, D.S., Berriz, G.F., Zhang, L.V., Dupuy, D., Walhout, A.J., Cusick, M.E., Roth, F.P. et al. (2004). Evidence for Dynamically Organized Modularity in the Yeast Protein-Protein Interaction Network. *Nature* **430(6995):** 88-93.

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(23):** 9362-9367.

Hsu, Y.H., Zillikens, M.C., Wilson, S.G., Farber, C.R., Demissie, S., Soranzo, N., Bianchi, E.N., Grundberg, E., Liang, L., Richards, J.B. et al. (2010). An Integration of Genome-Wide Association Study and Gene Expression Profiling to Prioritize the Discovery of Novel Susceptibility Loci for Osteoporosis-Related Traits. *PLoS Genet* **6(6):** e1000977.

Ihmels, J., Friedlander, G., Bergmann, S., Sarig, O., Ziv, Y., and Barkai, N. (2002). Revealing Modular Organization in the Yeast Transcriptional Network. *Nat Genet* **31(4):** 370-377.

Johnson, A.D., and O'Donnell, C.J. (2009). An Open Access Database of Genome-Wide Association Results. *BMC Med Genet* **10:** 6.

Jumbo-Lucioni, P., Ayroles, J.F., Chambers, M.M., Jordan, K.W., Leips, J., Mackay, T.F., and De Luca, M. (2010). Systems Genetics Analysis of Body Weight and Energy Metabolism Traits in Drosophila Melanogaster. *BMC Genomics* **11:** 297.

Kenney-Hunt, J.P., Vaughn, T.T., Pletscher, L.S., Peripato, A., Routman, E., Cothran, K., Durand, D., Norgard, E., Perel, C., and Cheverud, J.M. (2006). Quantitative Trait Loci for Body Size Components in Mice. *Mamm Genome* **17(6):** 526-537.

Kooperberg, C., and LeBlanc, M. (2008). Increasing the Power of Identifying Gene × Gene Interactions in Genome-wide Association Studies. *Genet Epidemiol* **32:** 255-263.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W. et al. (2001). Initial Sequencing and Analysis of the Human Genome. *Nature* **409(6822):** 860-921.

Levene, H. (1960). *Robust Tests for the Equality of Variance*. In *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, I. Olkin, et al. eds., Stanford University Press: Palo Alto, CA, pp. 278-292.

Mackay, T.F. (2009). The Genetic Architecture of Complex Behaviors: Lessons from Drosophila. *Genetica* **136(2):** 295-302.

Mackay, T.F., Heinsohn, S.L., Lyman, R.F., Moehring, A.J., Morgan, T.J., and Rollmann, S.M. (2005). Genetics and Genomics of Drosophila Mating Behavior. *Proc Natl Acad Sci U S A* **102(Suppl 1):** 6622-6629.

Magwire, M.M., Yamamoto, A., Carbone, M.A., Roshina, N.V., Symonenko, A.V., Pasyukova, E.G., Morozova, T.V., and Mackay, T.F. (2010). Quantitative and Molecular Genetic Analyses of Mutations Increasing Drosophila Life Span. *PLoS Genet* **6(7):** e1001037.

Manolio, T.A., Brooks, L.D., and Collins, F.S. (2008). A HapMap Harvest of Insights into the Genetics of Common Disease. *J Clin Invest* **118(5):** 1590-1605.

Mathew, C.G. (2008). New Links to the Pathogenesis of Crohn Disease Provided by Genome-Wide Association Scans. *Nat Rev Genet* **9(1):** 9-14.

McKnight, A.J., Currie, D., and Maxwell, A.P. (2010). Unravelling the Genetic Basis of Renal Diseases; from Single Gene to Multifactorial Disorders. *J Pathol* **220(2):** 198-216.

Nobrega, M.A., Ovcharenko, I., Afzal, V., and Rubin, E.M. (2003). Scanning Human Gene Deserts for Long-Range Enhancers. *Science* **302(5644):** 413.

Paternoster, L., Chen, W., and Sudlow, C.L. (2009). Genetic Determinants of White Matter Hyperintensities on Brain Scans: A Systematic Assessment of 19 Candidate Gene Polymorphisms in 46 Studies in 19,000 Subjects. *Stroke* **40(6):** 2020-2026.

Pogribny, I.P., and Beland, F.A. (2009). DNA Hypomethylation in the Origin and Pathogenesis of Human Diseases. *Cell Mol Life Sci* **66(14):** 2249-2261.

Pritchard, J.K., Stephens, M., and Donnelly, P. (2000). Inference of Population Structure using Multilocus Genotype Data. *Genetics* **155:** 945-959.

Reich, D.E., and Lander, E.S. (2001). On the Allelic Spectrum of Human Disease. *Trends Genet* **17(9):** 502-510.

Reilly, S.L., Ferrell, R.E., and Sing, C.F. (1994). The Gender-Specific Apolipoprotein E Genotype Influence on the Distribution of Plasma Lipids and Apolipoproteins in the Population of Rochester, MN. III. Correlations and Covariances. *Am J Hum Genet* **55(5):** 1001-1018.

Sandhu, M., Wood, A., and Young, E. (2010). Genomic Risk Prediction. *Lancet* **376:** 1366-1367.

Rollmann, S.M., Magwire, M.M., Morgan, T.J., Ozsoy, E.D., Yamamoto, A., Mackay, T.F., and Anholt, R.R. (2006). Pleiotropic Fitness Effects of the Tre1-Gr5a Region in Drosophila Melanogaster. *Nat Genet* **38(7):** 824-829.

Sandhu, M., Wood, A., and Young, E. (2010). Genomic Risk Prediction. *Lancet* **376:** 1366-1367.

Shyu, A.B., Wilkinson, M.F., and van Hoof, A. (2008). Messenger RNA Regulation: To Translate Or to Degrade. *EMBO J* **27(3):** 471-481.

Szolnoki, Z., Somogyvari, F., Kondacs, A., Szabo, M., Fodor, L., Bene, J., and Melegh, B. (2004). Specific APO E Genotypes in Combination with the ACE D/D Or MTHFR 677TT Mutation Yield an Independent Genetic Risk of Leukoaraiosis. *Acta Neurol Scand* **109(3):** 222-227.

Taylor, J. (2005). Clues to Function in Gene Deserts. *Trends Biotechnol* **23(6):** 269-271.

Tenesa, A., and Dunlop, M.G. (2009). New Insights into the Aetiology of Colorectal Cancer from Genome-Wide Association Studies. *Nat Rev Genet* **10(6):** 353-358.

The International HapMap Consortium. (2003). The International HapMap Project. *Nature* **426(6968):** 789-796.

Turunen, M.P., Aavik, E., and Yla-Herttuala, S. (2009). Epigenetics and Atherosclerosis. *Biochim Biophys Acta* **1790(9):** 886-891.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A. et al. (2001). The Sequence of the Human Genome. *Science* **291(5507):** 1304-1351.

Visscher, P.M. (2008). Sizing Up Human Height Variation. *Nat Genet* **40(5):** 489-490.

Waterland, R.A. (2009). Is Epigenetics an Important Link between Early Life Events and Adult Disease? *Horm Res* **71 (Suppl 1):** 13-16.

Wilson, A.G. (2008). Epigenetic Regulation of Gene Expression in the Inflammatory Response and Relevance to Common Diseases. *J Periodontol* **79(Suppl 8):** 1514-1519.

Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W. et al. (2010). Common SNPs Explain a Large Proportion of the Heritability for Human Height. *Nat Genet* **42(7):** 565-569.

Zhong, H., Yang, X., Kaplan, L.M., Molony, C., and Schadt, E.E. (2010). Integrating Pathway Analysis and Genetics of Gene Expression for Genome-Wide Association Studies. *Am J Hum Genet* **86(4):** 581-591.

Zou, L., Sriswasdi, S., Ross, B., Missiuro, P.V., Liu, J., and Ge, H. (2008). Systematic Analysis of Pleiotropy in C. Elegans Early Embryogenesis. *PLoS Comput Biol* **4(2):** e1000003.