GENOMIC STUDIES OF COGNITIVE FUNCTION AND ALZHEIMER'S DISEASE

by

ERIC L. GILLIS

(Under the Direction of Changwei Li and Toni Miles)

ABSTRACT

Currently, many longitudinal cognition studies measure cognitive function with various instruments and in different cognitive domains. This issue makes making inferences about causal factors problematic because misclassification of the outcome can bias the null hypothesis. To address this issue, our study validated the use of cognitive trajectories (CT) to accurately predict low cognitive function versus high cognitive function. CT analysis provides a uniform method to detect cases of impairment when instruments for measuring CI are not consistent across data.

Conclusively identifying risk factors for CI (AD) is an important public health priority as this information is required to effectively implement intervention strategies that have the potential to reduce morbidity and health care costs in an increasing senior population. Identifying risk factors and novel genes associated with AD will provide information that may illuminate the possible biological mechanisms that can influence the severity or rate of cognitive decline in AD patients.

INDEX WORDS: Alzheimer's disease, cognitive impairment, genome-wide gene-based analysis, genome-wide association study, trajectory analysis, Mendelian randomization

GENOMIC STUDIES OF COGNITIVE FUNCTION AND ALZHEIMER'S DISEASE

by

ERIC L. GILLIS
BS, Kennesaw State University, 1994
MPH, Georgia Southern University, 2005

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA
2018

GENOMIC STUDIES OF COGNITIVE FUNCTION AND ALZHEIMER'S DISEASE

by

ERIC L. GILLIS

Major Professors: Changwei Li

Committee: Toni Miles
Hanwen Huang
Jonathan Flacker

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
May 2018

DEDICATION

All efforts to achieve this goal were inspired by Msgt. Alfred L. Gillis, ANG, Retired; my personal Superman. And to my wife, Regina Gillis, who patiently listened to my complaints and encouraged me to “roll with the punches”.

ACKNOWLEDGEMENTS

I would like to acknowledge Dr. Toni Miles and Dr. Changwei Li for all their efforts in assisting me in achieving this goal. I would like to thank my Georgia Southern University Professors, Drs. Robert Vogel and Karl Peace for inspiring me to “reach beyond my grasp”.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER ONE

INTRODUCTION AND LITERATURE REVIEW: GENOMIC STUDIES OF COGNITION TRAJECTORIES

Introduction

Cognitive impairment (CI) is a common condition primarily affecting persons 60 years of age and older and the incidence and prevalence of CI increase with age [1]. CI is characterized by deficits in concentration, memory, reasoning, language and learning that interfere with independent living [1]. Dementia is the most common syndrome of CI and Alzheimer's Disease is the primary cause of most dementias, accounting for 60% of all cases [1]. CI cases are expected to increase from an estimated 46.8 million individuals (2015) to 74.7 million individuals worldwide by 2030 [1]. Persons living with CI are disproportionately affected with other chronic illnesses and experience probability of hospitalizations 3 times higher than persons without CI [2]. As a result, these individuals have poor long-term health outcomes [2]. CI places a substantial economic burden on the US economy and healthcare system. It is estimated that dementia and Alzheimer's disease (AD) cost the US economy an estimated $159 to $215 billion annually [2].

Age is the most prominent risk factor for CI; however, studies have implicated several other risk factors, including smoking, depression, hypertension, physical inactivity, diabetes, obesity, hyperlipidemia, cardiovascular disease, kidney disease, nutrition, cognitive inactivity and chronic inflammation [3]. However, findings regarding the effect of these behaviors and conditions on CI and dementia are still inconsistent [4]. Additional studies that are rigorously designed are required to clarify these risk factors for CI. Presently, only age and smoking have been conclusively identified as risk factors for CI [4, 5]. Studies have established that dementia

rates increase dramatically with advancing age, and a multi-ethnic longitudinal cohort study of 21,123 participants conducted by Rusanen et al. determined that the risk of dementia was highest for smokers of 2 packs of cigarettes per day than for nonsmokers (unadjusted Cox proportional hazard 2.01 (95%CI1.57-2.58). Additionally, a positive dose response effect of smoking on CI was observed [4, 5]. Cognitive decline and risk of CI can also be attributed to genetic factors. Twin and family studies have determined that cognitive ability has considerable heritability [6, 7]. A twin-based study by Kremen et al. determined the heritability estimate for CI to be between 40% and 48%, while a single-nucleotide polymorphic-based heritability estimate was determined to be between 31% and 51% [7, 8]. Twin-based estimates tend to be higher because these study designs account for variation in gene-gene interactions as well as specific single genes [9].

To date, several genomic loci have been identified for CI in large-scale genomic-wide association studies (GWAS). So far, 3 GWAS have been conducted for CI, and a total of 131 novel loci have been reported for CI [10-12].A GWAS meta-analysis of 53,949 participants, conducted by Davies et al., was able to identify 13 loci associated with cognitive function, and a subsequent study of 35,298 European participants, conducted by Trampush, was able to identify 2 additional novel loci associations [10, 12]. A GWAS by Davies et al. of 112,151 participants investigating cognitive functions and educational attainment identified 36 loci associated with general cognitive function [11]. However, current findings for CI only explain a small proportion of its heritability [6, 7]. Measurement errors regarding CI, the cross-sectional nature of study designs, and relatively small sample sizes have been proposed as explanations for the missing heritability problem [13]. But many GWAS involving CI are inconsistent

regarding how they measure CI. The meta-analyses conducted by Davies et al. and Trampush et al. measured CI across multiple domains, but the same domains were not consistently measured among all participants. which could have led to misclassification that biased the results and underpowered their study. Additionally, neither study used longitudinal measures of CI to appropriately capture the variability in cognition, reduce measurement errors, or increase statistical power [10, 11]. Plomin et al. argues that GWASs with large sample sizes could potentially find loci accounting for as much as 75% of cognitive heritability [13]. Another important factor contributing to the missing heritability problem is the lack of gene-environment interaction studies. It has been demonstrated that environmental factors can attenuate or augment the effects of genes on cognition [9]. For instance, physical activity can negate the negative effects of genes on memory [9].

As the proportion of individuals in the US aged 65 years and older increases, the incidence and prevalence of CI, dementia, and Alzheimer's disease are also expected to increase along with their associated morbidities. Understanding the underlying mechanism of CI and developing effective interventions to reduce the risk of these conditions is a substantial public health challenge and priority. Therefore, the goal of our study is to identify novel genomic loci through genome-wide gene-environment interaction studies of longitudinal cognition measures and to delineate the causal role of potential risk factors in CI development through robust Mendelian randomization studies. The proposed analyses will be conducted among participants of the Health and Retirement Study (HRS), a nationally representative survey among US adults aged 45 years and older. To achieve our overall objective, we will pursue the following specific aims:

Specific Aim 1

To evaluate the validity of using cognitive trajectories to accurately identify CI cases.

Specific Aim 2:

To delineate the causal relationship of coronary artery disease and Alzheimer's disease by Mendelian randomization.

Specific Aim 3:

To identify novel susceptibility genes for Alzheimer's Disease through gene-based genome-wide association studies.

Background

Dementia places a substantial economic burden on the United States (US) economy. The annual direct health care disbursement for dementia totals an estimated $159 to $215 billion, with an estimated $41,000 to $56,000 annually per case [2]. As the most common form of dementia, AD has caused an estimated annual total disbursement of $109 billion [2]. These disbursements exceed those for cancer and heart disease [2]. Seventy-five to eighty-four percent of per-case expense can be attributed to home care and nursing home care [2]. The annual mean direct medical expense for CI compared to those without CI is marginal ($6,784 vs. $6,042, respectively) [14]. However, the direct annual medical expenses associated with prevalent dementia compared to CI are nearly double ($11,678 vs. $6784) [14]. As the US population ages, the estimated annual health care total disbursement due to dementia is expected to rise to $511 billion by 2040 [2, 14, 15]. Because of the expected increases in the burden of dementia, along with the associated financial impact, it is imperative that we investigate the pathogenesis of cognitive decline to delineate the underlying mechanisms of dementia and to

develop novel intervention strategies to prevent and/or slow cognition decline and improve health outcomes.

The term “mild cognitive impairment” has been defined as the period of cognitive decline where cognition is no longer considered normal with respect to age but does not disrupt the ability to perform normal daily activities and does not meet the criteria for the diagnosis of dementia [16-19]. Amnestic MCI has been implicated as the intermediary between normal cognitive function and dementia, and it is believed to be the precursor to AD [5-8]. It is estimated that 10% to 20% of the US population 65 years and older have MCI, and 5% to 10% with MCI will progress to dementia, compared to 1% to 2% of the population without MCI [5]. The conversion rate from MCI to AD is estimated to be between 10% and 15% annually [8]. Studies have suggested that persons with MCI are at greater risk of developing serious illnesses and may have difficulty managing these illnesses if they progress to dementia [10].

MCI comprises two subtypes, amnestic and nonamnestic, and they are further subdivided into deficits of a single cognitive domain and of multiple cognitive domains [4, 16]. The incidence and prevalence of MCI vary in observational studies. The Mayo Clinic Study of Aging among 1,969 participants found the prevalence of amnestic MCI to be 11.1% among those between 70 and 80 years of age and nonamnestic prevalence to be 4.9% in the same age group [10]. This Mayo Clinic Study of Aging also revealed that 80% of persons with amnestic MCI converted to AD within 6 years of follow-up [20]. The Einstein Aging Study (EAS), a longitudinal cognitive aging and dementia study among 1,944 participants, confirmed findings by the Mayo Clinic’s Study of Aging [4]. The EAS also reported that there was only a marginal difference in prevalent MCI between men and women (22.2% and 21.0%, respectively) [4]. However, a significant

difference in MCI prevalence was identified between African Americans (AA) and European Americans (EA) (27.3% and 19.1%, respectively). The authors attributed this difference to the higher prevalence of nonamnestic MCI in AA (16.3% and 6.86%, respectively) [4]. Cognition is generally composed of 5 domains (Figure 1.1): Memory, Attention, Language, Visual-Spatial Perception, and Executive Functioning. The predominant feature of amnestic MCI is significant memory loss. Persons with amnestic MCI begin to forget important details that they once could recall (i.e., phone numbers, appointments, conversations or recent events), but other cognitive functions such as use of language, visuospatial skills and executive function are preserved [16]. As the condition progress, these cognitive domains can also be affected. Verbal and visual episodic memory are the major types of memory affected in amnestic MCI, and greatly impact the ability to function independently [21].

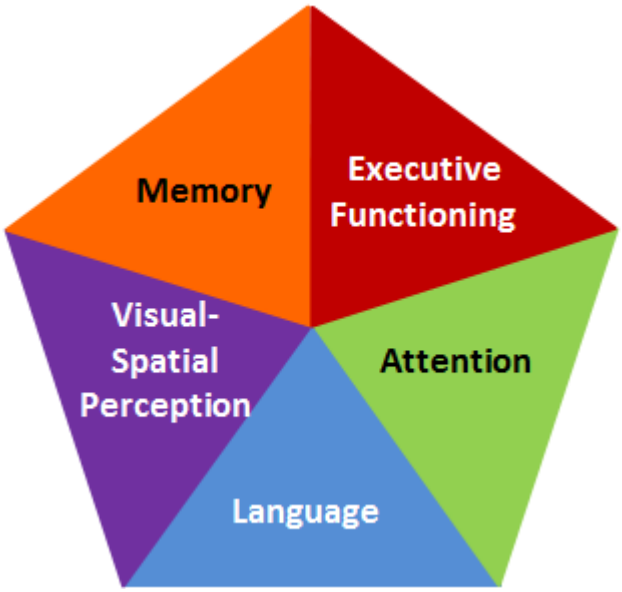

Figure 1.1. Cognitive Domains (Cognitive Therapeutics, http://cognitivetherapeutics.com/Services/interventions.php)

Executive Functioning includes cognitive abilities such as reasoning, problem solving, judgment, and cognitive flexibility;

Attention refers to the ability to focus on a specific piece of information for a sustained period of time while suppressing awareness of other competing distractions;

Language refers to the ability to execute verbal functions including spontaneous speech, speech repetition, speech comprehension, naming, reading and writing;

Visual-Spatial Perception involves the ability to accurately perceive and understand the visual relationships between objects and space;

Memory refers to the ability to retain information and utilize it later for adaptive purposes (Cognitive Therapeutics, http://cognitivetherapeutics.com/Services/interventions.php).

The operational criteria for the diagnosis of CI is as follows [19, 22];

1. Consistent lapses in memory confirmed by a close informant.
2. Deficits in cognitive domains and memory, evidenced by lower than normal performance for age on validated cognitive assessment tests.
3. Retention of abilities to perform activities of daily living.
4. Retention of global cognitive function.
5. Nondemented.

Screening for CI could potentially improve health outcomes and reduce morbidity through the implementation of early interventions. A review by Lin et al. identified 27 diagnostic accuracy studies examining succinct screening methods that were of moderate to good quality for the detection of MCI [23]. Only 6 of these methods had been evaluated in more than one

appropriately designed diagnostic accuracy study [13]. These methods are the Mini-Mental State Examination (MMSE), the Questionnaire on Cognitive Decline, the Clock-Drawing Test, the Telephone Interview for Cognitive Status, the Mini-Cog and the Montreal Cognitive Assessment (MoCA) [13]. The MMSE was the major method, used by 51% of healthcare professionals and endorsed by the American Academy of Neurology [24]. However, the MMSE has limitations that pose great concern for researchers. One, it overestimates impairment in persons over 60 years of age with less than 9 years of education [25]. Two, it is insensitive to mild forms of cognitive deficits as well as right hemisphere dysfunction [15]. And three, it inaccurately detects cognitive impairment in persons with average and below average verbal IQs [15]. MoCA is another widely used method for MCI screening. One study found that the MoCA performed better than the MMSE in screening for MCI; sensitivity and specificity of the MoCA were 80.48% and 81.19%, respectively, compared to those of 66.34% and 72.94%, respectively, for the MMSE [24]. An additional advantage of the MoCA over the MMSE is that it measures executive and visuospatial functions more comprehensively [26]. However, the MoCA does take longer to administer than the MMSE [26]. Meanwhile, efforts have been made to predict the progression of MCI to dementia. Lee et al. developed a multi-domain risk prediction index that integrated functional dependence, magnetic imaging and cognitive test scores to classify amnestic MCI patients into categories of high, moderate, and low risk for conversion to AD (N = 382 US Caucasians). The prediction index from MCI to AD has a sensitivity of 83.4% and specificity of 76.4% [27].

Treatment for CI is unclear. A meta-analysis conducted by Fitzpatrick-Lewis et al. identified that pharmacological interventions were ineffective and that behavioral and cognitive training

interventions only had modest effect [28]. There have been indications that interventions should target the stage of amnestic MCI, as this stage may be the most susceptible to treatment and carries the greatest risk of progression to AD [27]. Therefore, effective screening methods may improve the efficiency of intervention studies [27]. Meanwhile, studies are warranted to understand the underlying mechanisms, identify causal risk factors, and genes associated with CI and AD so that effective novel intervention strategies can be developed to prevent and treat CI and AD.

Risk Factors for CI and AD

The etiology of MCI is multifaceted. Biological, social, behavioral and environmental factors have been implicated in the pathogenesis of this condition as well as its progression to dementia and AD [3, 29-34]. Many factors have been associated with CI, including age, race, smoking, drinking, obesity, physical inactivity, depression, cognitive inactivity, hypertension, diabetes, cardiovascular disease (CVD), and chronic kidney disease. Of these factors, age and smoking are the conclusive factors that increase the risk of CI. However, there are limited studies that assess a broad range of risk factors [29]. Plassman and colleagues identified physical inactivity, diabetes, diet, smoking, depression, and cognitive inactivity as risk factors for cognitive decline [29]. Alternatively, Barnes and colleagues proposed that seven modifiable risk factors (obesity, hypertension, smoking, depression, cognitive and physical inactivity) are responsible for an estimated half of all dementia and AD cases [3]. More recent studies have corroborated some of these risk findings [35-43], and other studies have identified chronic kidney disease, chronic inflammation, alcohol consumption, vitamin A, education, homocysteine, metabolic syndrome, and high cholesterol as potential risk factors as well [31,

33, 34, 39, 44-51]. But many of these studies are observational and lack validation [17]. Nor do they examine a combination of risk factors [3, 29]. A systematic review by Deckers and colleagues found that depression, diabetes, cognitive inactivity, physical inactivity, hypertension, diet, obesity, smoking, alcohol consumption, chronic inflammation and cholesterol/hyperlipidemia were associated with an increased risk of cognitive decline [3]. Moreover, although an association was found for coronary heart disease, renal dysfunction, and cognitive inactivity, further studies are required to clarify the association as there were inconsistent results due to a small number of studies [3]. This study did not assess the association between education, vitamin A, homocysteine and cognitive decline because these are relativity new candidates as risk factors. Gender was not associated with MCI but race demonstrated an association with MCI [52]. African Americans (AA) were at greater risk of overall MCI; however, this was largely due to nonamnestic MCI being more prevalent among AA [52]. Amnestic MCI was more prevalent among Caucasians, and therefore, they are at higher risk than the AA population of developing dementia or AD [52].

The top risk factors for MCI are depression, cognitive inactivity, physical inactivity, hypertension, smoking, and diabetes. Studies have determined that depression substantially increases the risk of all dementias. A study by Diniz et al. estimated that persons experiencing depression had 85% higher risk of dementia than persons that do not and 90% higher risk for AD [3]. Cognitive activity has been shown to decrease the risk of AD by 62% [3]. Persons who do not engage in physical activity increased their risk of emerging dementia by 39%, and having hypertension increased risk by 61% [3]. The risk of dementia is increased by 47% for persons

affected with diabetes [3]. Conversely, studies have suggested that moderate alcohol consumption had a protective effect [3].

Studies have also identified cardiovascular disease (CVD) as a major risk factor for cognitive decline [53-55]. CVDs are the pathophysiological core of vascular cognitive impairment and causes 20% to 30% of all dementias [56]. The brain requires 25% to 50% of total blood glucose, 20% of total oxygen and 15% of total cardiac output to function properly [56]. A meta-analysis indicated that 10% of patients developed dementia following their first stroke, and more than 30% developed dementia after a second stroke [57]. The Framingham Heart Study implicated that elevated systolic blood pressure was associated with declining cognitive function. Poor cognitive performance risk increased for every 10 mmHg in systolic blood pressure above 120 mmHg [57]. However, currently, only age and smoking are conclusive risk factors for CI. As stated, smoking more than 2 packs per day increases the risk of dementia and Alzheimer's disease by more than 100% (HR 2.14 95%CI 1.65-4.03) [5].

Heritability and Genetic Studies of CI

Cognition is also determined by genetic factors and their interactions with environmental factors. The heritability of cognition ranges from 40% to 48% in population-based studies and from 31% to 51% in family and twin-based studies [6, 7]. Genomic studies, including candidate gene studies, genome-wide complex trait analysis (GCTA), and GWAS of CI, have made substantial progress. However, the current genomic findings explain only a small proportion of the CI heritability. Many factors have been postulated for the variation in heritability. This phenomenon is commonly referred to as the missing heritability problem (MHP) [13]. It is postulated that MHP is the result of multiple genes of small effect size being responsible for the

heritability of compound traits [13]. As a result of this complexity, associations conducted by GWAS fail to identify or replicate significant SNPs [13]. Plomin and colleagues suggested two reasons for the MHP: first, that commercial DNA arrays do not include rare DNA variants, and second, that twin and adoption studies may overestimate heritability [13].

Genome-wide complex trait analysis (GCTA)

Genome-wide complex trait analysis (GCTA) is a new population-based method for estimating the genetic variance explained by all SNPs genotyped in a sample, not solely SNPs that belong to family members as used in twin and adoption studies [13]. GCTA requires large sample sizes in which everyone has been genotyped for hundreds of thousands of SNPs. The analysis does not identify genes that are correlated with specific traits but uses probability similarities across thousands of SNPs to forecast phenotypic similarities, pair by pair, in large samples of unrelated persons. GCTA can miss genetic effects due to unrelated causal variants that are not associated with common SNPs on genotyping arrays [13]. An additional drawback to GCTA is that it only captures additive genetic effects [13], while twin and adoption studies account for both additive and nonadditive effects [13]. Despite these deficits, GCTA has been an invaluable method to estimate the variance in cognitive ability due to genetic factors. GCTA was initially used to investigate cognitive ability in a group of 3,500 unrelated adults, and the results determined heritability estimates of crystallized and fluid intelligence to be 0.40 and 0.51, respectively [13]. Plomin and colleagues conducted a study to compare heritability estimates generated by GCTA to those generated by a twin study design in a sample of 3,154 pairs of 12-year-old twins [13]. General, nonverbal, verbal cognitive ability and language abilities were the measures assessed, and a ratio measure of GCTA heritability estimates to twin-based estimates

was used to assess the accountability of GCTA estimates to twin-based estimates. This study determined that general cognitive ability had the largest estimates at .35, 95% CI(.12, .58) and .46, 95% CI (.36, .48) by GCTA and twin design, respectively, with a ratio estimate of .76 [13]. This finding suggests that twin-based estimates may be inflated, but this hypothesis is not clearly supported, as twin-based studies account for both additive and nonadditive effects, while GCTA accounts only for additive effects and thus has lower heritability estimates [13]. However, this study does provide evidence that GWASs with appropriately large sample sizes will have greater success in identifying SNPs/genes associated with cognitive ability, accounting for roughly 75% of cognitive heritability using current DNA technology [13].

Candidate Genes and Polymorphisms

Candidate gene studies focus on genes that are functionally relevant to cognitive functions based on a priori knowledge and test associations of SNPs within the candidate genes with cognitive impairment. Candidate gene studies offer a powerful approach for detecting genetic variants that influence cognition. Many candidate genes have been studied for AD and its related phenotypes, particularly proteins or their precursors involved in senile plaque formation [58-63].

Liu et al. found that genes TOMM40 and MAPT/STH, associated with amyloid-β and tau proteins, respectively, were associated with amnestic MCI among 209 Chinese participants [63]. This study also reported that the apolipoprotein gene (APOE), a gene responsible for transporting cholesterol, was associated with MCI [63]. This study determined that polymorphisms TOMM40 rs157581 and rs2075650, MAPT/STH rs242562, and APOE 4 rs8126696 were associated with amnestic MCI [63]. However, this study was relatively small

and only conducted in a Chinese population; therefore, the findings may not be generalizable across different populations.

Golanska and colleagues discovered that the *APBB2* coding β amyloid precursor protein-binding family member 2 was associated with CI among 150 Caucasian centenarians [59]. In these centenarians, *APBB2* rs13133980 and hCV1558625 were associated with an increased risk of severe CI [59]. Similarly to those mentioned above, the major limitation to this study was its substantially limited sample size.

A cohort study involving 5,994 65-year-old osteoporotic Caucasian men and 9,704 osteoporotic Caucasian women determined that genes *BIN1, CD33, CELF1, CR1, HLA CLUSTER*, and *MEF2C1* were associated with cognitive decline among women and *ABCA7, HLA cluster*, *PICALM, PTK2B, SLC24A4*, and *SORL1* were associated with cognitive decline among men. A major limitation of this study is that it did not explain the function of the genes identified and it also used two different methods to assess cognitive impairment.

Li and colleagues determined that polymorphism rs1699102 of the SORL1 (sortilin-related receptor) gene, which modulates amyloid-β production and amyloid precursor protein processing, was associated with rapid cognitive decline among 780 non-demented Chinese participants older than 50 years of age [62]. Other polymorphisms of SORL1 were not assessed in this study.

Genome-wide Association Studies

Genome-wide association studies (GWAS) evaluate associations of dense panels of SNPs covering the entire genome with disease outcomes and are a powerful hypothesis-free approach for the discovery of susceptibility loci for common complex traits such as cognition

and CI. Early studies used a case-control design and commercial genotyping chips with different genotyping resolutions, such as the Affymetrix 100k or 500k and Illumina 300k chip [64]. GWASs were conducted in three stages: discovery, replication and meta-analyses [64, 65]. In the discovery stage, genome-wide association analyses were performed for all SNPs with disease or disease-related measures, and SNPs with suggestive findings (P<1E-5 or 1E-6) were carried forward to the replication stage analyses [64, 65]. In the replication stage, promising findings identified in the discovery stage were evaluated for association with the same disease or disease-related trait in a second independent population [64, 65]. Then, results across the discovery and replication stages were combined using meta-analysis methods. SNPs with replication p-values <0.05, meta-analysis stage p-value <5E-8 and having consistent effect size directions across the discovery and replication stages were deemed significant [64, 65]. An important consideration for conducting GWASs is the appropriate sample size (power) to detect a causal genetic variant.

To date, it has been challenging for GWASs to identify associations between SNPs and cognitive abilities, largely due to sample size limitations that can influence the power required to detect an effect. Moreover, very few GWASs have investigated genetic variants associated with cognitive function [10-12]. Trampush and colleagues conducted a GWAS meta-analysis of 35,298 participants, firstly, to investigate associations between genetic variants and cognitive function and, secondly, to determine the degree to which general cognitive function correlates to previously published neurobehavioral phenotypes [10]. This study discovered a novel SNP for cognitive function: a variant of CENPO rs76114856, which encodes a centromere complex. The study also revealed that genetic variants determining cognitive performance were also strongly

correlated with years of education, absence of psychiatric disorders and smoking status [10]. The major limitation of this study was that the population's age range spanned from adolescents to seniors [10]. It has been determined that cognitive ability is reduced in early childhood in relation to adulthood, which can distort the effect in favor of detecting a significant finding.

Davies et al. conducted the largest GWAS to date for 3 cognitive domains (verbal-numerical reasoning, reaction time and memory) and educational attainment using the UK Biobank participants (N=112,151). The study identified 149 novel SNPs associated with verbal-numerical reasoning, 36 SNPs associated with reaction time, and 1,115 SNPs associated with educational attainment, while 327 SNPs were associated with general cognitive function [11]. The study did not identify any SNPs associated with memory. Because educational attainment shared an association with SNPs previously identified for general cognitive function, it was suggested that educational attainment could possibly serve as a proxy for cognitive function [11]. This study did not measure cognitive functions by standardized test, posing a limitation in that the study's findings had little overlap with those reported in previous studies using standard cognitive tests. In addition, all participants were British Caucasians, which limited generalizability [11].

A meta-analysis conducted by Davies et al. using a cohorts of middle-aged and senior participants (N=53,949) from the Cohort for Heart and Aging Research in Genetic Epidemiology (CHARGE) identified 13 SNPs within 4 genes that were associated with general fluid cognitive function: *TOMM40, APOE, MEF2C and ABCG1* [12]. These genes were previously identified as being associated with AD [12]. The results of this study demonstrated that cognitive function was heritable and likely polygenic [12]. Limitations included that the same cognitive domains

measured were not uniformly applied to participants, nor were the same tests administered to all participants, which increased the risk of misclassification.

Gene-Environment Interaction

Education and APOE

Cognition is also influenced by the interaction of genes with environmental factors such as lifestyle behaviors. Environmental factors can modify the association between genes/polymorphisms and cognitive ability. Gene-lifestyle interaction analyses have been proposed to explore the missing heritability of many complex traits, including CI [8, 66-72]. Such analyses account for the influence of environmental factors and help to identify genetic variants that function under certain environments. Using powerful 2-degree-of-freedom (df) joint interaction tests [8,66-72], gene-lifestyle interaction studies have successfully identified many novel loci underlying complex traits such as blood pressure [8,66-72]. Unfortunately, gene-lifestyle interaction analyses on cognitive function have only been explored in candidate gene studies [8,66-72]. These studies have primarily focused on the interactions between the APOE gene and lifestyles, due to extensive genetic research identifying APOE Є4's association with Alzheimer's disease and cognitive decline [73, 74]. Twenty-five to thirty percent of the US population are carriers of the APOE Є4 allele [73].

Smoking and APOE

In the Rotterman study, when stratified by genotypes of the *APOE* gene, smoking was a strong risk factor for Alzheimer's disease, but only among individuals without the *APOE* Є4 alleles, not among those carrying the *APOE* Є4 alleles [75]. Similarly, in a study by Reitz et al., current smokers lacking the *APOE* Є4 alleles had faster memory decline.

Studies investigating the joint effect of education and APOE regarding cognitive ability have produced mixed results. A longitudinal study (Canadian Study of Health and Aging) comprising 1,185 seniors determined that educational attainment and physical activity were modifiers of the association between APOE and CI [76]. Higher education and regular physical activity reduced the risk of CI [76]. The educational joint effect was corroborated by Cook and Fletcher (2015), whose study determined that carriers of the *APOE* Є4 allele who had completed high school or less had a greater risk of cognitive decline than those who had completed college [77]. However, a study by Seeman et al. (2005) reported a three-way interaction of education, *APOE* Є4 and time with respect to cognitive ability (memory), showing that cognitive decline was greatest for persons with the Є4 allele and educational attainment above 8 years, which indicated that education may amplify the effect of the APOE Є4 allele. In contrast, Van Gerven et al. (2012) were not able to identify any significant interactions between *APOE* Є4 and educational attainment with respect to cognitive decline.

Cognitive activity and APOE

The β-amyloid (Aβ) protein forms the characteristic plaque, a precursor of Alzheimer's disease, and is the primary cause of neurodegeneration, which leads to cognitive decline in AD [78]. Wirth and colleagues (2014) hypothesized that extensive lifetime cognitive activities such as reading and cognitive games may reduce Aβ burden. They tested whether a joint effect existed between cognitive activity and APOE Є4 with respect to AD pathogenesis. Their study determined that lifetime cognitive activity in carriers of this allele regulated Aβ plaque deposition. *APOE* Є4 carriers who engaged in greater cognitive activity had lower Aβ plaque deposition [78]. However, nonsignificant results were determined for noncarriers [78].

BMI and APOE

BMI is a risk factor for CVD, diabetes and hypertension. Each of these conditions have been independently associated with cognitive decline [73]. Rajan et al. (2014) investigated the joint effect of body mass index (BMI) and *APOE* Є4 with respect to cognitive decline in 4,055 seniors, while controlling for age, education, gender, heart disease, stroke, diabetes and hypertension. They determined that BMI modified the association between *APOE* Є4 and cognitive decline [73]. The *APOE* Є4 allele was not associated with increased cognitive decline in obese persons; in general, cognitive decline was slower in obese persons with the *APOE* Є4 allele than normal BMI persons with the *APOE* Є4 allele [73].

Cortisol and APOE

Cortisol is a steroid produced by the adrenal gland in response to stress and is implicated as a risk factor for cognitive decline [79]. Lee et al. (2008) investigated the joint effect of cortisol and APOE genotypes with respect to cognitive function in 967 seniors 50-70 years of age. They determined that elevated levels of cortisol metrics (pretest, mean and AUC) were associated with poor cognitive performance [79]. Carriers of two APOE Є4 alleles who had elevated cortisol were at greater risk of cognitive deficits than those without or with only one copy of the Є4 allele. This finding provides evidence that APOE Є4 modifies the association between cortisol and cognitive function [79].

Overall, these gene-environment interaction studies using the candidate gene approach help elucidate the mechanism of CI development and AD.

Mendelian Randomization

Mendelian randomization (MR) is a robust technique that uses genes/SNP proxies for potential risk factors to delineate causal relationships between the factors and a specified disease outcome. According to Burgess et al. (2017), the following assumptions must be met for a genetic variant to be used as a proxy: "a genetic variant must be associated with the risk factor; the genetic variant is not associated with confounders of the risk factor–outcome relationship; and the genetic variant is not associated with the outcome conditional on the risk factor and confounders of the risk factor–outcome relations" [80]. This technique is resistant to confounding, reverse causation, and selection bias. However, this approach must consider population stratification and pleiotropy [81].

Some unique features of MR have made it widely adopted for causal inference. First, a genetic variant can impact exposure through an effect on constitutional tendency [81], for instance, variants that influence the tendency to smoke. Secondly, genetic variants may impact an intermediate phenotype, such as serum cholesterol [81]. Third, a biological response to an environmental exposure can be altered by a genetic variant [81]. Fourth, genetic variants can classify modifiable exposures that are not easily measured [81]. Fifth, maternal genotypes can be investigated as causal factors of intrauterine exposure [81]. And lastly, genetic variants can be used to characterize exposures that impact disease risk [81].

AD Susceptibility Genes

Genome-wide association studies (GWAS) have identified 21 risk loci that affect AD phenotype (*CR1, BIN1, INPP5D, MEF2C, TREM2, CD2AP, HLA-DRB1/HLA- DRB5, EPHA1, NME8, ZCWPW1, CLU, PTK2B, PICALM, SORL1, MS4A4/MS4A6E, SLC24A4/RIN3, FER- MT2, CD33, ABCA7, CASS4,*

APOE) [82, 85]. And It is estimated that these loci account for 28% of AD's heritability and 30% of the familial risk [86]. Having a first degree relative affected with AD increases the relative risk for AD (3.5 95% CI 2.6-4.6) [86]. Studies have found that 30-48% of AD patients had an affected first degree relative [86]. The most prominent genetic risk factor for AD was thought to be the *APOE* gene but comprehensive European and International GWAS have identified novel risk genes associated with the amyloid β pathway and many of these genes are linked to the immune system, *CLU, CR1, ABCA7, CD33, EPHA1* and *MS4A* gene cluster [83, 86, 87]. Several more AD risk genes are correlated with synaptic function (*PICALM, CD33, CD2AP, EPHA1, and BIN1*) and lipid metabolism were also identified (*CLU* and *ABCA7*) [87].

Statistical Analysis

To address Specific Aim 1, of whether cognitive trajectories accurately identify CI cases, trajectory modeling was conducted to determine group assignment and cognitive trends over time for normal vs. cognitive impairment or worse using the summary cognitive scores collected at specific time points.Descriptive baseline statistics were also generated for ADAMS and Rand HRS.

To assess the validity of cognitive trajectory modeling to predict cognitive status, we calculated the sensitivity,specificity, positive predictive value , and negative predictive value using the physicians' diagnostic evaluation of the ADAMS participants as the gold standard. Duke University specialists (neurologist, neuropsychologist, cognitive neuroscientist and geropsychiatrist) reviewed and designated a preliminary research diagnosis of cognitive status: normal, cognitive impairment but not demented, and demented (definition based on DSM-II-R and DSM-IV) [75]. The preliminary diagnosis was revised by the study geropsychiatrist when

medically justified [75]. All analyses were executed in SAS 9.4 (SAS Institute, Cary, North Carolina, USA).

To address Specific Aim 2, of using a two sample Mendelian randomization study to delineate causal relations of coronary artery disease and AD. An inverse regression analysis modeling was conducted to assess the effect of SNPs associated with CAD and SNPs associated with AD..

To address Specific Aim 3, of identifying novel genomic loci through gene-based analysisof AD, we used GATES and ECS (VEGAS) statistical methods to identify genes associated with AD. This procedure was implemented in KGG4 software.

Implications

Currently, many longitudinal cognition studies measure cognitive function with various instruments and in different cognitive domains. This issue can make investigating causal inferences problematic because misclassification of the outcome can bias the null hypothesis. To address this issue, our study validated the use of cognitive trajectories (CT) to accurately detect dementia and cognitive impairment. To date, there are no known studies investigating the validity of CT to detect CI. CT analysis provides a uniform method to detect cases of impairment when instruments for measuring CI are not consistent across data.

Understanding the underlying biological mechanism and conclusively identifying risk factors for CI and AD are an important public health priority as this information is required to effectively implement intervention strategies with the potential to reduce morbidity, mortality and health care costs in an increasing senior population. Identifying novel SNPs associated with CI and AD

will provide information that may illuminate the possible gene-gene interactions that can influence the severity or rate of decline of CI.

References

1. Liao, W., et al., *A profile of The Clinical Course of Cognition and Comorbidity in Mild Cognitive Impairment and Dementia Study (The 4C study): two complementary longitudinal, clinical cohorts in the Netherlands.* BMC Neurol, 2016. **16**(1): p. 242.
2. Hurd, M.D., et al., *Monetary costs of dementia in the United States.* N Engl J Med, 2013. **368**(14): p. 1326-34.
3. Deckers, K.E.A., *Target risk factors for dementia prevention: a systematic review and Delphi concensus study on the evidence from observational studies.* Int J Geriatr Psychiatry, 2015. **30**: p. 234-246.
4. Katz, M.J., et al., *Age-specific and sex-specific prevalence and incidence of mild cognitive impairment, dementia, and Alzheimer dementia in blacks and whites: a report from the Einstein Aging Study.* Alzheimer Dis Assoc Disord, 2012. **26**(4): p. 335-43.
5. Rusanen, M., et al., *Heavy smoking in midlife and long-term risk of Alzheimer disease and vascular dementia.* Arch Intern Med, 2011. **171**(4): p. 333-9.
6. Cox, A.J., et al., *Heritability and genetic association analysis of cognition in the Diabetes Heart Study.* Neurobiol Aging, 2014. **35**(8): p. 1958 e3-1958 e12.
7. Kremen, W.S., et al., *Early identification and heritability of mild cognitive impairment.* Int J Epidemiol, 2014. **43**(2): p. 600-10.
8. Davies, G., et al., *Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N=53949).* Mol Psychiatry, 2015. **20**(2): p. 183-92.

9. Papenberg, G., U. Lindenberger, and L. Backman, *Aging-related magnification of genetic effects on cognitive and brain integrity.* Trends Cogn Sci, 2015. **19**(9): p. 506-14.

10. Trampush, J.W., et al., *GWAS meta-analysis reveals novel loci and genetic correlates for general cognitive function: a report from the COGENT consortium.* Mol Psychiatry, 2017.

11. Davies, G., et al., *Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N=112151).* Molecular Psychiatry, 2016. **21**: p. 758-767.

12. Davies, G., et al., *Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (n=53 949).* Molecular Psychiatry, 2015. **20**: p. 183-192.

13. Plomin, R.e.a., *Common DNA Markers Can Account for More Than Half of the Genetic Influence on Cognitive Abilities.* Psychol. Sci, 2013. **24**(4): p. 562-568.

14. Leibson, C.L., et al., *Direct medical costs and source of cost differences across the spectrum of cognitive decline: a population-based study.* Alzheimers Dement, 2015. **11**(8): p. 917-32.

15. Chapman, D.e.a., *Dementia and Its Implications for Public Health.* Prev Chronic Dis., 2006(Apr).

16. Peterson, R., *Mild Cognitive Impairment.* N Engl J Med, 2011. **364**(23): p. 2227-34.

17. Janoutova, *Is Mild Cognitive Impairment A Precursor of Alzheimer's Disease? Short Review.* Cent Eur J Public Health 2015. **23**(4): p. 365-367.

18. Morris, *Mild Cognitive Impairment Represents Early-Stage Alzheimer's disease.* Arch Neurol, 2001. **58**: p. 397-405.

19. Forelenza, C., et al, *Mild cognitive impairment (part 1): clinical characteristics and predictors of dementia.* Revista Brasileria de Psiquitria, 2013. **35**: p. 178-185.

20. Yanhong, O., M. Chandra, and D. Venkatesh, *Mild cognitive impairment in adult: A neuropsychological review.* Ann Indian Acad Neurol, 2013. **16**(3): p. 310-8.

21. Mistridis, P., et al., *The 12 Years Preceding Mild Cognitive Impairment Due to Alzheimer's Disease: The Temporal Emergence of Cognitive Decline.* J Alzheimers Dis, 2015. **48**(4): p. 1095-107.

22. Petersen, R.C.e.a., *Mild Cognitive Impairment: Clinical Characterization and Outcome.* Arch Neurol, 1999. **56**(Mar): p. 303-308.

23. Lin, P.J. and P.J. Neumann, *The economics of mild cognitive impairment.* Alzheimers Dement, 2013. **9**(1): p. 58-62.

24. Ciesielska, N., et al., *Is the Montreal Cognitive Assessment (MoCA) test better suited than the Mini-Mental State Examination (MMSE) in mild cognitive impairment (MCI) detection among people aged over 60? Meta-analysis.* Psychiatr Pol, 2016. **50**(5): p. 1039-1052.

25. Naugle RI, K.K., *Limitations of the Mini-Mental State Examination.* Cleve Clin J Med. , 1989. **53**(3): p. 277-81.

26. Lam, B., et al., *Criterion and convergent validity of the Montreal cognitive assessment with screening and standardized neuropsychological testing.* J Am Geriatr Soc, 2013. **61**(12): p. 2181-5.

27. Korolev, I.O., et al., *Predicting Progression from Mild Cognitive Impairment to Alzheimer's Dementia Using Clinical, MRI, and Plasma Biomarkers via Probabilistic Pattern Classification.* PLoS One, 2016. **11**(2): p. e0138866.

28. Fitzpatrick-Lewis, D., et al., *Treatment for mild cognitive impairment: a systematic review and meta-analysis.* CMAJ Open, 2015. **3**(4): p. E419-27.

29. Plassman, B.L., et al., *Systematic review: factors associated with risk for and possible prevention of cognitive decline in later life.* Ann Intern Med, 2010. **153**(3): p. 182-93.

30. Stacey, D., L.G. Ciobanu, and B.T. Baune, *A systematic review on the association between inflammatory genes and cognitive decline in non-demented elderly individuals.* Eur Neuropsychopharmacol, 2015.

31. Setien-Suero, E., et al., *Homocysteine and cognition: A systematic review of 111 studies.* Neurosci Biobehav Rev, 2016. **69**: p. 280-98.

32. Rannikko, I., et al., *Predictors of Long-Term Change in Adult Cognitive Performance: Systematic Review and Data from the Northern Finland Birth Cohort 1966.* Clin Neuropsychol, 2016. **30**(1): p. 17-50.

33. Zeng, J., et al., *Marginal vitamin A deficiency facilitates Alzheimer's pathogenesis.* Acta Neuropathol, 2017.

34. Zammit, A.R., et al., *Cognitive Impairment and Dementia in Older Adults With Chronic Kidney Disease: A Review.* Alzheimer Dis Assoc Disord, 2016. **30**(4): p. 357-366.

35. Bertram, S., K. Brixius, and C. Brinkmann, *Exercise for the diabetic brain: how physical training may help prevent dementia and Alzheimer's disease in T2DM patients.* Endocrine, 2016. **53**(2): p. 350-63.

36. Berk, L., M. van Boxtel, and J. van Os, *Can mindfulness-based interventions influence cognitive functioning in older adults? A review and considerations for future research.* Aging Ment Health, 2016: p. 1-8.

37. Chen, R., et al., *Association of passive smoking with cognitive impairment in nonsmoking older adults: a systematic literature review and a new study of Chinese cohort.* J Geriatr Psychiatry Neurol, 2013. **26**(4): p. 199-208.

38. Cheng, S.T., *Cognitive Reserve and the Prevention of Dementia: the Role of Physical and Cognitive Activities.* Curr Psychiatry Rep, 2016. **18**(9): p. 85.

39. Williams, J.W., et al., *Preventing Alzheimer's disease and cognitive decline.* Evid Rep Technol Assess (Full Rep), 2010(193): p. 1-727.

40. van de Rest, O., et al., *Dietary patterns, cognitive decline, and dementia: a systematic review.* Adv Nutr, 2015. **6**(2): p. 154-68.

41. Singh, B., et al., *Association of mediterranean diet with mild cognitive impairment and Alzheimer's disease: a systematic review and meta-analysis.* J Alzheimers Dis, 2014. **39**(2): p. 271-82.

42. Collins, N., et al., *Smoking increases risk for cognitive decline among community-dwelling older Mexican Americans.* Am J Geriatr Psychiatry, 2009. **17**(11): p. 934-42.

43. Knott, V.J., A. Harr, and C. Mahoney, *Smoking history and aging-associated cognitive decline: An event-related brain potential study.* Neuropsychobiology, 1999. **40**(2): p. 95-106.

44. Danna, S.M., et al., *Association between Depressive Symptoms and Cognitive Function in Persons with Diabetes Mellitus: A Systematic Review.* PLoS One, 2016. **11**(8): p. e0160809.

45. Yates, L.A., et al., *Cognitive leisure activities and future risk of cognitive impairment and dementia: systematic review and meta-analysis.* Int Psychogeriatr, 2016. **28**(11): p. 1791-1806.

46. Lewis, J.R., *Early kidney disease, an important factor in cognitive decline or merely a harbinger of early vascular brain injury?* Am J Nephrol, 2015. **41**(4-5): p. 303-4.

47. Bettcher, B.M. and J.H. Kramer, *Longitudinal inflammation, cognitive decline, and Alzheimer's disease: a mini-review.* Clin Pharmacol Ther, 2014. **96**(4): p. 464-9.

48. Anstey, K.J., D.M. Lipnicki, and L.F. Low, *Cholesterol as a risk factor for dementia and cognitive decline: a systematic review of prospective studies with meta-analysis.* Am J Geriatr Psychiatry, 2008. **16**(5): p. 343-54.

49. Sachdeva, A., et al., *Alcohol-Related Dementia and Neurocognitive Impairment: A Review Study.* Int J High Risk Behav Addict, 2016. **5**(3): p. e27976.

50. Sabia, S., et al., *Alcohol consumption and cognitive decline in early old age.* Neurology, 2014. **82**(4): p. 332-9.

51. van den Kommer, T.N., et al., *Homocysteine and inflammation: predictors of cognitive decline in older persons?* Neurobiol Aging, 2010. **31**(10): p. 1700-9.

52. Au, B., S. Dale-McGrath, and M.C. Tierney, *Sex differences in the prevalence and incidence of mild cognitive impairment: A meta-analysis.* Ageing Res Rev, 2016.

53. de la Torre, J.C., *Cardiovascular risk factors promote brain hypoperfusion leading to cognitive decline and dementia.* Cardiovasc Psychiatry Neurol, 2012. **2012**: p. 367516.

54. Harrison, S.L., et al., *Cardiovascular disease risk models and longitudinal changes in cognition: a systematic review.* PLoS One, 2014. **9**(12): p. e114431.

55. Nash, D.T. and H. Fillit, *Cardiovascular disease risk factors and cognitive impairment.* Am J Cardiol, 2006. **97**(8): p. 1262-5.

56. Picano, E., et al., *Cognitive impairment and cardiovascular disease: so near, so far.* Int J Cardiol, 2014. **175**(1): p. 21-9.

57. Tadic, M., C. Cuspidi, and D. Hering, *Hypertension and cognitive dysfunction in elderly: blood pressure management for this global burden.* BMC Cardiovasc Disord, 2016. **16**(1): p. 208.

58. Fu, Y.e.a., *NEDD9 Gene Ploymorphism Influences the Risk of Alzheimer Disease and Cognitive Function in Chinese Older Persons.* Alzheimer Dis Assoc Disord, 2012. **26**(1): p. Jan - Mar.

59. Golanska, E.e.a., *APBB2 genetic ploymorphisms are associated with severe cognitive impairment in centenarians.* Experimental Gerontology, 2013. **48**: p. 391-394.

60. Lythgoe, C.e.a., *Population-based analysis of cholesteryl ester transfer protein identifies association between 1405v and cognitive decline: the Cache County Study.* Neurobiol Aging, 2014. **36**: p. 547.e1 - 547.e3.

61. Nettiksimmons, J.e.a., *Gene-based aggregate SNP associations between candidate AD genes and cognitive decline.* Age 2016. **38**(41).

62. Li, h.e.a., *SORL1 rs1699102 polymorphism modulates age-related cognitive decline and gray matter volume reduction in non-demented individuals.* European Journal of Neurology, 2016. **24**: p. 187-194.

63. Liu, X.e.a., *Association Study of Candidate Gene Polymorphisms with Amnestic Mild Cognitive Impairment in a Chinese Population.* PLoS One, 2012. **7**(7): p. e41198.

64. Spencer, C.C., et al., *Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip.* PLoS Genet, 2009. **5**(5): p. e1000477.

65. Bush, W.S. and J.H. Moore, *Chapter 11: Genome-wide association studies.* PLoS Comput Biol, 2012. **8**(12): p. e1002822.

66. Li C, H.J., Chen J, Zhao J, Gu D, Hixson JE, et al., *Genome-Wide Gene-Sodium Interaction Analyses on Blood Pressure: The Genetic Epidemiology Network of Salt-Sensitivity Study.* Hypertension, 2016. **68**(2): p. 348-355.

67. Simino J, S.G., Bis JC, Chasman DI, Ehret GB, Gu X, et al., *Gene-age interactions in blood pressure regulation: a large-scale investigation with the CHARGE, Global BPgen, and ICBP Consortia.* Am J Hum Genet 2014. **95**(1): p. 24-38.

68. Ahmad S, R.G., Varga TV, Ali A, Kurbasic A, Shungin D, et al., *Gene × physical activity interactions in obesity: combined analysis of 111,421 individuals of European ancestry.* PLoS Genet 2013. **9**(7): p. e1003607.

69. Qi L, C.Y., *Gene-environment interaction and obesity.* Nutr Rev, 2008. **66**(12): p. 684-694.

70. Zhang R, C.M., Zhao Y, Wu C, Guo H, Shi Y, et al., *A genome-wide gene-environment interaction analysis for tobacco smoke and lung cancer susceptibility.* Carcinogenesis, 2014. **35**(7): p. 1528-35.

71. Manolio TA, C.F., Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al., *Finding the missing heritability of complex diseases.* Nature, 2009. **461**(7265): p. 747-753.

72. Murcray CE, L.J., Gauderman WJ, *Gene-environment interaction in genome-wide association studies.* Am J Epidemiol, 2009. **169**(2): p. 219-226.

73. Rajan, K.B., et al., *Gene-environment interaction of body mass index and apolipoprotein E epsilon4 allele on cognitive decline.* Alzheimer Dis Assoc Disord, 2014. **28**(2): p. 134-40.

74. Boardman, J.D., et al., *Social disorder, APOE-E4 genotype, and change in cognitive function among older adults living in Chicago.* Soc Sci Med, 2012. **74**(10): p. 1584-90.

75. Ott A, S.A., Hofman A, van Harskamp F, Witteman JC, Van Broeckhoven C, et al., *Smoking and risk of dementia and Alzheimer's disease in a population-based cohort study: the Rotterdam Study.* Lancet 1998. **351**(9119): p. 1840-3.

76. Meng, X. and C. D'Arcy, *Apolipoprotein E gene, environmental risk factors, and their interactions in dementia among seniors.* Int J Geriatr Psychiatry, 2013. **28**(10): p. 1005-14.

77. Cook, C.J. and J.M. Fletcher, *Can education rescue genetic liability for cognitive decline?* Soc Sci Med, 2015. **127**: p. 159-70.

78. Wirth, M., et al., *Gene-environment interactions: lifetime cognitive activity, APOE genotype, and beta-amyloid burden.* J Neurosci, 2014. **34**(25): p. 8612-7.

79. Lee, B.K., et al., *Apolipoprotein e genotype, cortisol, and cognitive function in community-dwelling older adults.* Am J Psychiatry, 2008. **165**(11): p. 1456-64.

80. Burgess, S., Bowden, J., Fall, T., Ingelsson, E., Thompson, S., *<Sensitivity_Analyses_for_Robust_Causal_Inference.6.pdf>.* Epidemiology, 2017. **28**(1).

81. Smith, G.D. and S. Ebrahim, *Mendelian randomization: prospects, potentials, and limitations.* Int J Epidemiol, 2004. **33**(1): p. 30-42.

82. Swaminathan, S., et al., *Amyloid pathway-based candidate gene analysis of [(11)C]PiB-PET in the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort.* Brain Imaging Behav, 2012. **6**(1): p. 1-15.

83. Cuyvers, E. and K. Sleegers, *Genetic variations underlying Alzheimer's disease: evidence from genome-wide association studies and beyond.* The Lancet Neurology, 2016. **15**(8): p. 857-868.

84. Chaudhry, M., et al., *Genetic variation in imprinted genes is associated with risk of late-onset Alzheimer's disease.* J Alzheimers Dis, 2015. **44**(3): p. 989-94.

85. Bettens, K., K. Sleegers, and C. Van Broeckhoven, *Genetic insights in Alzheimer's disease.* The Lancet Neurology, 2013. **12**(1): p. 92-104.

86. Ashare, R.L., et al., *APOE varepsilon4, an Alzheimer's disease susceptibility allele, and smoking cessation.* Pharmacogenomics J, 2013. **13**(6): p. 538-43.

87. Jones, L., et al., *Genetic evidence implicates the immune system and cholesterol metabolism in the aetiology of Alzheimer's disease.* PLoS One, 2010. **5**(11): p. e13950.

CHAPTER TWO

ASSESSMENT OF COGNITIVE TRAJECTORIES TO PREDICT COGNITIVE STATUS

Gillis, E., Li, C. Miles, T. To be submitted to *PLoS One*

ABSTRACT

INTRODUCTION

It is estimated that 10% to 20% of persons 65 years and older are affected by cognitive impairment (CI), and 32% to 53% of these individuals will progress to some form of dementia. The goal of our study was to assess the validity of using trajectory analysis to accurately determine cognitive status and determine the underlying association of heart disease and low cognition trajectory.

METHODS

Participants were selected from the Aging Demographics and Memory Study (ADAMS). Modeling was conducted using the SAS Proc Traj procedure to determine group membership and cognitive trends over time for normal vs. cognitive impairment or worse.

RESULTS

We have determined that trajectories of cognitive impairment had a sensitivity and specificity of 85% and 70%, respectively, and a positive predictive value of 80% and a negative predictive value of 77%. Low cognition trajectory was associated with self-reported heart disease.

CONCLUSION

Our findings indicated that trajectory analysis achieves greater accuracy than the Mini-Mental State Examination (MMSE) and comparable accuracy to the Montreal Cognitive Assessment (MoCA).

Introduction

Cognitive impairment (CI) is a common condition of aging and is defined by deficits in memory, concentration, reasoning, language and learning that interfere with performance of daily activities required for independent living [1]. Dementia is the most common form of CI and affects an estimated 46.8 to 74.7 million individuals worldwide. Individuals affected with CI disproportionately experience other chronic illnesses and experience 3 times higher hospitalizations than those without CI [1].

Currently, there are several methods available for evaluating cognitive impairment. Lin et al. identified 27 accuracy diagnostic studies assessing cognitive impairment, although these studies varied in quality [3]. This variability in screening quality increases the likelihood of false positive as well as false negatives (misclassification). The Mini-Mental State Examination (MMSE) was the most widely used screening method, with 51% of healthcare professionals employing its use [4]. However, the MMSE has limitations. One, it overestimates impairment in persons over 60 years of age with less than 9 years of education [4]. Two, it is insensitive to mild forms of cognitive deficits as well as right hemisphere dysfunction [4]. And three, it inaccurately detects cognitive impairment in persons with average and below average verbal IQs [4]. Alternatively, the Montreal Cognitive Assessment (MoCA) is a better method for screening CI, as its sensitivity and specificity were 80.48% and 81.19%, respectively, compared to 66.34% and 72.94% for MMSE [4].

Typically, individuals affected by dementia/Alzheimer's disease have a lengthy preclinical phase prior to the onset of cognitive and physical deterioration [5, 6]. Trajectories of cognitive impairment may provide discernment into the temporal patterns of deterioration and identify

those at the greatest risk of becoming demented, as well as identify those who may benefit from participating in clinical studies investigating interventions [5, 6]. Trajectory modeling presumes that a study population comprises distinct groups with different latent trajectories [7]. This method distinguishes groups of individuals with similar trends in progression of the outcome of interest over time and assumes that the repeated measures of the outcome are outcomes of the initial exposure by adjusting for group behavior [7].

Longitudinal data with time-based measures of cognition provide the underpinning for the analysis of the temporal patterns of decline. However, the statistical methods employed can affect the results obtained [8]. Generalized estimates of equations (GEE), random effects models, and growth mixture modeling have been popular methods for conducting trajectory analysis, and each method yields robust estimates. However, these models assume that all individuals in the population decline at a uniform rate and that all individuals are members of the same population specified by the individual's average cognitive trajectory [8]. Mixture models have the advantages of relaxing the single population assumption, thereby permitting the classification of individuals with similar trajectories, the evaluation of unobserved classes of individuals with homogenous trajectories and the evaluation of a class's distinct risk factors [8]. The SAS Proc Traj procedure is ideally suited to perform this type of analysis as it has several advantages. The procedure is easy to use; accommodates missing data, sample weights, and overlapping cohort designs; and can also accommodate irregular spacing of measurements [9]. The objective of this study was to determine if group-based trajectory modeling (mixture model) was a valid and comparable method to assign cognitive status (impaired vs. normal)

versus physician diagnosis in senior adults using the nationally representative cohort of Aging Demographics and Memory Study (ADAMS) participants.

Methods

Participants

To facilitate this study, participants were selected from the Aging Demographics and Memory Study (ADAMS) and the Rand Health and Retirement Study (Rand HRS, demographics information). Both studies are sub-studies of the University of Michigan Health and Retirement Study (HRS), which is a longitudinal representative survey study of health, housing, employment, disability, demographics, health care utilization and income within the US population of persons aged 50 years or older (and spouses) and is supported by the National Institute on Aging and the Social Security Administration [10]. The objective of the HRS is to enable research in the areas of financial fitness, health insurance policy, and retirement policy, and it integrates data from 1992, 1993, 1994, 1998, 2000, 2002, 2004, 2006, 2008, 2010, 2012 and 2014 [10].

The objective of the ADAMS was to acquire clinical assessments of dementia for participants who were 70 years of age or older [11]. Assessments were conducted between August 2001 and December 2003, with a final sample size of 856 participants [11]. The ADAMS was stratified regarding gender, age and cognitive test scores in order to ensure adequate sampling over the complete range of cognitive status [11]. Based on a battery of cognitive tests, respondents were classified into cognitive strata: demented, cognitive impairment but not demented (CIND) and normal [11]. Composite scores of the full battery of cognitive tests (ranging from 0 to 35 points) were used to classify self-respondents and used in trajectory modeling (scores against time).

The follow ups for these cognitive batteries were taken at 18-month intervals for the 2002, 2004, 2006, 2008, and 2010 waves.

The Rand HRS is a consolidated, user-friendly version of the HRS data containing derived variables that are consistently identified across waves. The data includes imputated information for medical expenditures, income and assets [12]. This data was merged with ADAMS data to obtain demographic information on ADAMS participants.

Statistical Analysis

Trajectory modeling (mixture model) was conducted using the SAS Proc Traj procedure to determine group membership and cognitive trends over time for normal vs. demented or worse using the summary cognitive scores collected at specific time points. A univariate model was chosen to investigate the validity of trajectory modeling to predict cognitive impairment. Two x two tables were generated to evaluate the sensitivity and specificity for detecting impairment, normal versus demented or worse. Descriptive baseline statistics were also generated for ADAMS participants.

To assess the validity of cognitive trajectory modeling to predict cognitive status, we calculated the sensitivity, specificity, positive predictive value and negative predictive value using the physicians' diagnostic evaluation of the ADAMS participants as the gold standard. Cognitive measures included vocabulary tests to assess baseline intelligence and an abridged version of the Telephone Interview for Cognitive Status (TICS). The abridged version included an object-naming test, a timed counting backwards test, the serial sevens subtraction test, an immediate and delayed 10-noun free recall test and the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE). Duke University specialists (a neurologist, neuropsychologist, cognitive

neuroscientist and geropsychiatrist) reviewed the test results as well as proxy interviews and designated a preliminary research diagnosis of cognitive status: normal, cognitive impairment but not demented, and demented (definition based on DSM-II-R and DSM-IV). The preliminary diagnosis was revised by the study geropsychiatrist when medically justified.. All analyses were executed in SAS 9.4 (SAS Institute, Cary, North Carolina, USA). We tested the association of self-reported heart problems and cognitive trajectories in 2000 Rand HRS participants using Chi-square.

Results

Table 2.1 displays the demographic characteristics of the ADAMS participants by cognitive status. The median age for cognitively normal individuals and CIND individuals were very similar but not statistically significant, 82 years and 83 years, respectively. There was a statistically significant difference in median education, 12 versus 10 for CIND individuals. There were no significant differences for BMI, number of chronic conditions, or income. For the remaining characteristics, the proportion of Caucasians and blacks was significantly different between the two groups, as was the proportion of those who ever drank alcohol.

Table 2.2 displays the trajectory group assignment and distribution of cognitive status among ADAMS participants, group 1 (cognitively impaired) and group 2 (cognitively normal), as well as the sensitivity, specificity, positive predictive value, and negative predictive value, 85%, 70%, 80%, and 77%, respectively. We also found that the relative risk of belonging to a low CI trajectory given heart problems was 1.4 (95% CI 1.06-1.77) (table 2.3). The results from the trajectory analysis (high and low trajectories) are displayed in Figure 2.1.

Discussion

We have demonstrated that trajectories of dementia and cognitive impairment had a sensitivity of 85% and 70% specificity, respectively, and a positive predictive value of 80% and a negative predictive value of 77%. To our knowledge, our study was the first to investigate the validity of cognitive trajectory analysis within an aging HRS population to assess cognitive status. Our findings indicated that trajectory analysis achieves greater accuracy than the MMSE and comparable accuracy to the MoCA. The validity of our findings was enhanced by the status of the HRS data. The response rate for HRS was 80%, with a re-interview rate for succeeding waves being 92-95% [13]. Additionally, the Survey Research Center at the University of Michigan has fruitfully tracked 98-99% of HRS participants [13]. ADAMS data provided information on dementia from participants from all regions of the United States utilizing a uniformed diagnostic protocol [10]. All testing was independently scored by 2 clinical neuropsychology technicians before a final review by a PhD-level neuropsychologist [10]. However, there are a few important limitations involving the use of ADAMS data. Firstly, the generalizability may be less than optimal due to the lower response rate of 56% [10]. And secondly, the ADAMS furnishes data on the prevalence of cognitive impairment, which hinders determining the causality of potential risk factors [10]. Nevertheless, the latter limitation does not affect the goal of this study, which was to determine the validity of trajectory analysis in predicting cognitive impairment. We demonstrated that trajectory analysis performs better than the widely used MMSE at predicting cognitive impairment [4]. Furthermore, we demonstrated that trajectory analysis may be a viable alternative when cognitive assessments

are not congruent in subsequent waves or secondary datasets, provided that the assessments are collected consistently over time.

Because our focus was determining the comparability of cognitive trajectories versus conventional physician diagnosis, we are very confident that our findings are well supported by the SAS Proc Traj procedure. The foundation of this procedure can be structured on a semiparametric group-based mixture model, which has the advantage of not being constrained by the single population assumption nor by the assumption that a single trajectory suitably approximates the entire population [14]. This procedure facilitates a graphical presentation of data that is easily understood and displays the repercussions of distinct trajectories of decline. This method can also serve to harmonize cognitive status across studies to increase power and validity of future studies.

References

1. Liao, W., et al., *A profile of The Clinical Course of Cognition and Comorbidity in Mild Cognitive Impairment and Dementia Study (The 4C study): two complementary longitudinal, clinical cohorts in the Netherlands.* BMC Neurol, 2016. **16**(1): p. 242.
2. Teeters, D.A., et al., *Mild Cognitive Impairment and Risk of Critical Illness.* Crit Care Med, 2016. **44**(11): p. 2045-2051.
3. Lin, J.S., et al., *Screening for cognitive impairment in older adults: A systematic review for the U.S. Preventive Services Task Force.* Ann Intern Med, 2013. **159**(9): p. 601-12.
4. Ciesielska, N., et al., *Is the Montreal Cognitive Assessment (MoCA) test better suited than the Mini-Mental State Examination (MMSE) in mild cognitive impairment (MCI) detection among people aged over 60? Meta-analysis.* Psychiatr Pol, 2016. **50**(5): p. 1039-1052.
5. Howieson, D.B., et al., *Trajectory of mild cognitive impairment onset.* J Int Neuropsychol Soc, 2008. **14**(2): p. 192-8.
6. Wilkosz, P.A., et al., *Trajectories of cognitive decline in Alzheimer's disease.* Int Psychogeriatr, 2010. **22**(2): p. 281-90.
7. Modeling, T., https://www.mailman.columbia.edu/research/population-health-methods/trajectory-analysis.11/28/17
8. Terrera, G.M., et al., *One size fits all? Why we need more sophisticated analytical methods in the explanation of trajectories of cognition in older age and their potential risk factors.* Int Psychogeriatr, 2010. **22**(2): p. 291-9.

9. Data, P.T.A.S.P.f.G.B.M.o.L., https://www.researchgate.net/profile/Bobby_Jones2/publication/266822262_Proc_TRAJ_A_SAS_Procedure_for_Group Based_Modeling_of_Longitudinal_Data/links/551815fc0cf2f7d80a3d27 79/Proc-TRAJ-A-SAS-Procedure-for-Group-Based-Modeling-of-Longitudinal-Data.pdf. 11/30/17.

10. Health and Retirement Study, I.f.S.R., University of Michigan, http://hrsonline.isr.umich.edu/.

11. The Aging, Demograhic and Memory Study (ADAMS),, http://hrsonline.isr.umich.edu/index.php?p=shoavail&iyear=XB. 11/30/17.

12. HRS, R., https://www.rand.org/labor/aging/dataprod/hrs-data.html.

13. Health and Retirement Study, I.f.S.R., University of Michigan, *Health and Retirement Study, Aging, Demographics, and Memory Study (ADAMS) Supplement.* 2013.

14. Nagin, D.S., *Group-based trajectory modeling: an overview.* Ann Nutr Metab, 2014. **65**(2-3): p. 205-10.

Table 2.1. ADAMS participant characteristics by cognitive status

Characteristics	NORMAL (N=139)	CIND or WORSE (N=533)
Age*	82 (95%CI 81.7-83.4)	83 (95%CI 83.0-85.2)
Education (yrs)*	12 (95%CI 12.2-13.2)	10 (95%CI 9.1-9.8)
BMI*	25.1 (95%CI 24.7-26.7)	26.5 (95%CI 26.4-28.1)
Chronic conditions*	2.0 (95%CI 1.75-2.16)	2.0 (95%CI 2.43-2.68)
Gender		
Male	66 (47.5%)	201 (37.7%)
Female	73 (52.5%)	332 (62.3%)
Race		
Caucasian	125 (89.9%)	388 (72.8%)
Black	11 (7.9%)	123 (23.1%)
Other	3 (2.2%)	22 (4.1%)
Smoke (Ever)		
yes	54 (62.1%)	104 (61.2%)
no	33 (37.9%)	66 (38.8%)
Drink (Ever)		
yes	46 (52.3%)	55 (32.2%)
no	42 (47.7%)	116 (67.8%)
Hypertension (Ever)		
yes	46 (52.3%)	102 (60%)
no	42 (47.7%)	68 (40%)

* median values

Table 2.2. Cognitive trajectories versus ADAMS status (N=325)

Trajectory group assignment	**ADAMS Status**			
	Cognitive impairment or worse	*Normal*		
Low	162	40	**Positive predictive value**	80%
High	28	95	**Negative predictive value**	77%
			Sensitivity	85%
			Specificity	70%

Table 2.3. Cognitive trajectories versus self-report Heart Disease (N=2000)

Trajectory group assignment	**Self-reported Heart Disease Status**		**Statistic**		**Prob**
	Yes	*No*	Chi-Square	Df =1	0.0192
Low	49	131			**Value**
High	361	1459	Relative Risk	MH	1.37(1.06-1.77)

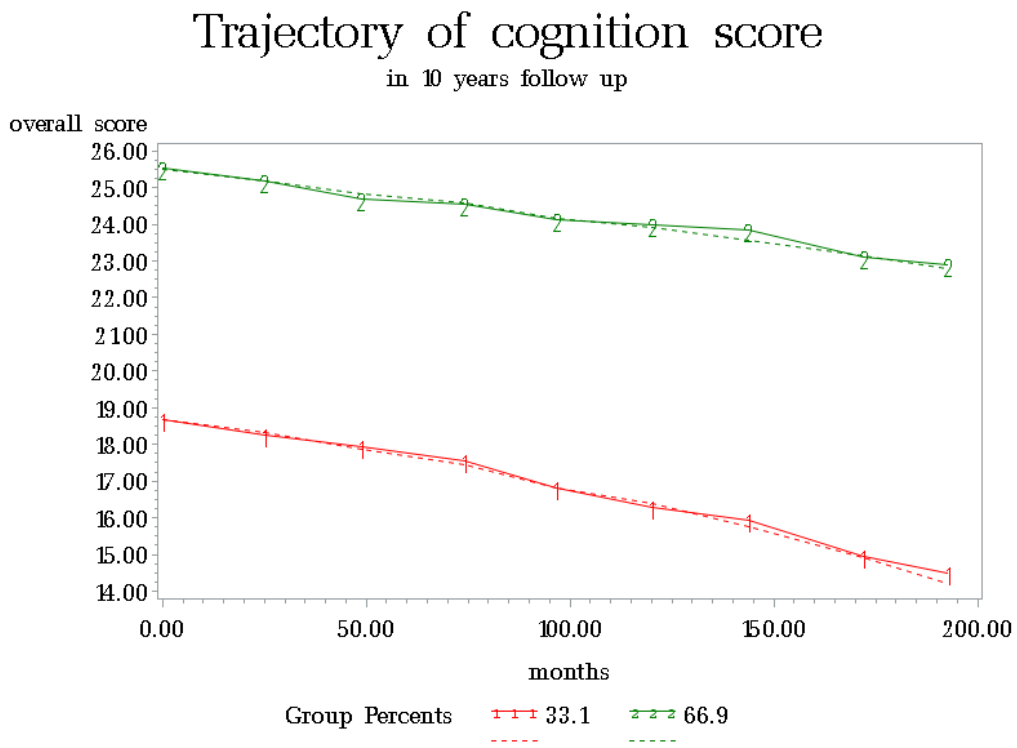

Figure 2.1. Trajectory of cognition scores in 10-year follow up

CHAPTER THREE

INVESTIGATING CAUSALITY IN THE ASSOCIATION OF CORONARY ARTERY DISEASE AND ALZHEIMER'S DISEASE

Gillis, E., Li, C. Miles, T. To be submitted to the *Journal of Alzheimer's Disease*

Abstract

AD is the most common cause of age-related dementia, accounting for 60% of dementia cases in the United States, and is the fifth-leading cause of death among persons 65 years and older. It has been postulated that the risk for cardiovascular disease (CVD) and AD may share the same biological pathway. The objective of our study was to evaluate the causal effect of CAD on AD risk by Mendelian randomization using summary genome-wide association study (GWAS) data from the International Genomics of Alzheimer's Project (IGAP) and CAD summary data from 4 large-scale GWASs. We determined that the odds of AD were 1.89 (95% CI 1.18-3.03) per doubling of the odds of CAD.

Introduction

Alzheimer's disease (AD) is a neurodegenerative disease that causes cognitive degeneration and cannot be cured or prevented, nor can its progression be slowed [1, 2]. It is caused by the proliferation of the protein plaques beta amyloid and tau on the exterior and interior of neurons, causing cell death by interfering with neuron-to-neuron communication and the transport of essential nutrients into neurons [1]. AD is characterized by a progressive decline in memory, dysfunction in sensory motor function, impairments in language and spatial orientation [1]. These symptoms result in the inability to perform tasks required for independent living [1]. AD is the most common cause of age-related dementia, accounting for 60% of dementia cases in the United States, and is the fifth-leading cause of death among persons 65 years and older [2]. It is estimated that 10% of Americans 65 years and older are living with AD, and 2 in 1000 are affected annually in this age range [1, 2]. Age, family history of

AD and carrying the APOE 4 allele are the most prominent risk factors; however, atherosclerotic disorders have also been implicated as potential risk factors [1].

Studies have demonstrated that vascular pathology has a causal role in the development of AD and dementias [2-7]. Both AD and coronary artery disease (CAD) share genes associated with low-density lipoprotein cholesterol and total cholesterol [3]. It has been postulated that the risk for cardiovascular disease (CVD) and AD may share the same biological pathway [3]. Cholesterol has been robustly established as a risk factor for cardiovascular disease (CVD), and studies have determined that lowering circulating cholesterol may lower the risk of AD [3]. Coronary artery disease has been associated with reduced cognitive ability and the neuropathological lesion of Alzheimer's disease [4, 8]. Moreover, coronary heart disease was found to be associated with an increased risk of cognitive impairment and dementia (OR = 1.45, 95% CI 1.21-1.47) [9].

Mendelian randomization (MR) is a technique that uses genetic variants as proxies for an exposure to evaluate the causal relationship between an exposure and outcome in observational epidemiological studies that are limited by confounding and reverse causation [10]. Because this technique is resistant to confounding and reverse causation, MR provides supporting evidence for causality [10]. However, there are 3 assumptions that must be satisfied: The genetic variant is associated with the exposure, the genetic variant is not associated with confounders of the exposure-outcome association, and the genetic variant is independent of the outcome [10-12]. Currently, there are minimal studies evaluating the association between cardiovascular diseases and AD, and, as a result, findings are inconclusive. The objective of our study was to evaluate the causal effect of CAD on AD risk by Mendelian

randomization using summary genome-wide association study (GWAS) data from the International Genomics of Alzheimer's Project (IGAP) and CAD summary data from 4 large-scale GWASs.

Methods

The International Genomics of Alzheimer's Project (IGAP) is a two-stage meta-analysis of 4 previously published genome-wide association studies (The European Alzheimer's Disease Initiative, EADI; the Alzheimer's Disease Genetics Consortium, ADGC; the Cohorts for Heart and Aging Research in Genomic Epidemiology consortium, CHARGE; and the Genetic and Environmental Risk in AD consortium, GERAD) [13]. We obtained summary data on over 7 million genotyped or imputed single nucleotide polymorphisms (SNPs) from the stage 1 meta-analysis of European participants, 17,008 cases of Alzheimer's disease and 37,154 controls. The IGAP dataset included information on chromosome location, SNP position, SNP rsID, coded allele, non-coded allele, overall effect size for the coded allele, overall standard error for the coded allele and meta-analysis p-value. We also obtained summary data on 95 SNPs associated with CAD collected from 4 large-scale GWAS meta-analyses, UK Biobank, 1000 Genome Project, and the Coronary Artery Disease Genome-wide Replication and Meta-analysis (CARDIOGRAM). All participants were of European ancestry. The UK Biobank consisted of 4,831 CAD cases and 115,455 controls; the CARDIOGRAM, 63,746 CAD cases and 130,681 controls; the 1000 Genome Project, 60,801 CAD cases and 123,524 controls; and 1 independent GWAS conducted by Howson et al. included 88,192 CAD cases and 162,544 controls. Eighty-six significant SNPs were identified for CAD exposure and AD outcome with a p-value $<10^{-5}$.

Statistical Analysis

A two-sample Mendelian randomization for multiple genetic variants was conducted using R 3.4.3 software. To test for a violation of the MR assumption, we performed an Egger regression analysis. To examine the causal effect of CAD on AD, the inverse-variance weighted (IVW) regression method was employed to homogenize the SNP-CAD exposure and SNP-AD outcome coefficients to calculate the overall estimate of the CAD causal effect. IVW is essentially a weighted regression employed when the assumption of constant variance of the least square residuals is violated [14]. Considering that AD is binary, the resulting estimate of effect represents the log odds for AD per unit increase in the log odds for CAD. For this reason, we multiplied the log odds of CAD by 0.693, then exponentiated to represent the OR for AD per doubling in the OR for CAD [14]. We formally tested violations of the MR assumptions using Egger regression. Egger regression incorporates an intercept parameter to the weighted regression model that represents the pleiotropic effect of CAD SNPs on risk of AD [11, 14, 15]. The Egger and IVW regression models are:

$\beta Y = \theta o + \theta E\, \beta x + \epsilon$

$\beta Y = \theta E\, \beta x + \epsilon$

where β_Y is the SNP association with AD (outcome), θo is the intercept, θE is the casual estimate, and βx is the SNP association with CAD (exposure). The model is the same for the IVW minus the intercept term. Graphs were also produced using R.

Results

The results of the two-sample MR-IVW of 86 SNPs associated with CAD indicate a strong causal association with AD. The odds of AD were 1.89 (95% CI 1.18-3.03) per doubling of the odds of CAD (Table 3.1). The results of the Egger regression provide evidence that there is no pleiotropic effect involved in the association between CAD and AD (Intercept: OR 0.98, 95% CI 1.03-1.63, p=0.49). Results from the Egger regression also confirm the causal association found in the IVW regression (OR AD 4.6, 95% CI 0.76-27.7, p=0.96).

Discussion

Our findings provide supporting evidence that CAD is a causal factor in the risk of developing AD. This study also provides evidence to support the hypothesis that vascular diseases are associated with cognitive impairment and AD [16-18]. We have confirmed the findings of other studies investigating the causal role of vascular diseases in the risk of AD. An MR study conducted by Proitsi et al. found that the risk of AD increased by 23% per 1-unit increase in low-density lipoprotein (LDL), and a study of 1,138 Medicare participants found that heart disease was associated with an increased risk of AD (OR 1.40, 95% CI 1.10-1.80) [9, 19]. The Rotterdam Study, a cohort study of factors that determine cardiovascular disease occurrence, found that both clinically diagnosed AD and vascular dementia were associated with atherosclerosis [20]. Further, an early study conducted by Martin et al. found that amyloid plaque was abundant in the brains of non-demented CAD subjects [21].

A major strength of our study was that we were able to obtain summary data from 4 large-scale studies of CAD (The European Alzheimer's disease Initiative, EADI; the Alzheimer’s Disease Genetics Consortium, ADGC; the Cohorts for Heart and Aging Research in Genomic

Epidemiology Consortium, CHARGE; the Genetic and Environmental Risk in AD Consortium, GERAD) and AD summary data from the IGAP. These studies provide the substantially large sample sizes necessary to detect the small effect sizes indicative of genomic studies. Another strength was that all SNPs used in these analyses reached genome-wide significance. A major limitation of our study was that we used two independent population samples, which could have biased our results in favor of the null hypothesis. However, we were able to detect a causal effect between CAD exposure and the risk of AD due to the robustness of the IVW regression method.

In conclusion, we were able to identify CAD as a causal risk factor in the development of AD using a two-sample Mendelian randomization that incorporates the IVW regression method. We then tested that the MR assumptions were not violated using the Egger regression method. MR is a robust and flexible method for investigating causality using individual or summary level data.

References

1. Association, A.S., *2017 Alzheimer's Disease Facts and Figures.* 2017. **13**: p. 325-373.
2. Bleckwenn, M., et al., *Impact of coronary heart disease on cognitive decline in Alzheimer's disease: a prospective longitudinal cohort study in primary care.* Br J Gen Pract, 2017. **67**(655): p. e111-e117.
3. Karlsson, I.K., et al., *Genetic susceptibility to cardiovascular disease and risk of dementia.* Transl Psychiatry, 2017. **7**(5): p. e1142.
4. Beeri, M.S., et al., *Coronary artery disease is associated with Alzheimer disease neuropathology in APOE4 carriers.* Neurology, 2006. **66**(9): p. 1399-404.
5. Deckers, K., et al., *Coronary heart disease and risk for cognitive impairment or dementia: Systematic review and meta-analysis.* PLoS One, 2017. **12**(9): p. e0184244.
6. Murdock, D.G., et al., *KIAA1462, a coronary artery disease associated gene, is a candidate gene for late onset Alzheimer disease in APOE carriers.* PLoS One, 2013. **8**(12): p. e82194.
7. Wolozin, B. and M.M. Bednar, *Interventions for heart disease and their effects on Alzheimer's disease.* Neurol Res, 2006. **28**(6): p. 630-6.
8. Hagenaars, S.P., et al., *Polygenic risk for coronary artery disease is associated with cognitive ability in older adults.* Int J Epidemiol, 2016.
9. Luchsinger, J., et al., *Aggregation of Vascular Risk Factors and Risk of Incident Alzheimer's Disease.* Neurology, 2005. **65**(4): p. 545-551.

10. Zheng, J., et al., *Recent Developments in Mendelian Randomization Studies.* Curr Epidemiol Rep, 2017. **4**(4): p. 330-345.

11. Burgess, S., A. Butterworth, and S.G. Thompson, *Mendelian randomization analysis with multiple genetic variants using summarized data.* Genet Epidemiol, 2013. **37**(7): p. 658-65.

12. Bowden, J., et al., *Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator.* Genet Epidemiol, 2016. **40**(4): p. 304-14.

13. Lambert, J.C., et al., *Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease.* Nat Genet, 2013. **45**(12): p. 1452-8.

14. Gage, S.H., et al., *Assessing causality in associations between cannabis use and schizophrenia risk: a two-sample Mendelian randomization study.* Psychol Med, 2017. **47**(5): p. 971-980.

15. Burgess, S., A.S. Butterworth, and J.R. Thompson, *Beyond Mendelian randomization: how to interpret evidence of shared genetic predictors.* J Clin Epidemiol, 2016. **69**: p. 208-16.

16. Hofman, T.E.A., *Atherosclerosis, Apolipoprotein E, and prevalence of dementia and Alzheimer's disease in Rotterdam Study.* The Lancet, 1997. **349**(9046): p. 151-154.

17. Strickland, S., *Blood will out: vascular contributions to Alzheimer's disease.* J Clin Invest, 2018. **128**(2): p. 556-563.

18. Iturria-Medina, Y., V. Hachinski, and A.C. Evans, *The vascular facet of late-onset Alzheimer's disease: an essential factor in a complex multifactorial disorder.* Curr Opin Neurol, 2017. **30**(6): p. 623-629.

19. Proitsi, P., et al., *Genetic predisposition to increased blood cholesterol and triglyceride lipid levels and risk of Alzheimer disease: a Mendelian randomization analysis.* PLoS Med, 2014. **11**(9): p. e1001713.

20. Torre, *Alzheimer Disease as a Vascular Disorder.* Stroke, 2002. **33**: p. 1152-1162.

21. Martins, I.J., et al., *Apolipoprotein E, cholesterol metabolism, diabetes, and the convergence of risk factors for Alzheimer's disease and cardiovascular disease.* Mol Psychiatry, 2006. **11**(8): p. 721-36.

Table 3.1. Association between CAD and AD using two-sample Mendelian randomization

Method	OR	95% CI	P-value
IVW	1.89	1.18-3.03	0.006
Egger (Intercept)	0.98	1.03-1.63	0.49
Egger	4.6	0.76-27.7	0.096

CHAPTER FOUR

GENE-BASED GENOME-WIDE ASSOCIATION STUDY OF ALZHEIMER'S DISEASE

Gillis, E., Li, C. Miles, T. To be submitted to *PLoS One*

Abstract

AD is the sixth-leading cause of death in the US and is the fifth-leading cause death among persons 65 years and older; furthermore, approximately two-thirds of the 5.3 million seniors living with AD are women (3.3 million). Age, family history of AD, and genetics are the most prominent risk factors for AD, however, cardiovascular disease (CVD), low educational attainment, social and cognitive engagement inactivity, smoking and traumatic brain injury have been implicated as modifiable risk factors for AD. The objective of this study was to identify novel genes associated with Alzheimer's disease using gene-based genome-wide association.

Introduction

It is projected by 2060 that persons 65 and older will comprise 24% of the US population, resulting in an increase in age-related diseases [1]. An estimated 5.3 million American seniors 65 years and older are currently living with Alzheimer's disease (AD), and 2 in 1000 are affected annually [1, 2]. AD is caused by the proliferation of protein plaques beta amyloid and tau within and outside neurons, causing cell death by interfering with the communication between neurons and transport of essential nutrients into neurons [1]. The disease is distinguished by an escalating decline in memory, dysfunction in sensory motor function, and impairments in language and spatial orientation [1, 3]. Ultimately, these symptoms result in the inability to perform tasks required for independent living [1]. AD is the sixth-leading cause of death in the US and is the fifth-leading cause death among persons 65 years and older; furthermore, approximately two-thirds of the 5.3 million seniors living with AD are women (3.3 million) [1].

Typically, persons 65 years and older survive, on average, 4 to 8 years after diagnosis; however, individuals may survive as long as 20 years after diagnosis [1]. Age, family history of AD, and genetics are the most prominent risk factors for AD, however, cardiovascular disease (CVD), low educational attainment, social and cognitive engagement inactivity, smoking and traumatic brain injury have been implicated as modifiable risk factors for AD [1, 3-5]. Studies have indicated that factors that increase the risk of CVD, such as obesity, diabetes, hypertension, high cholesterol, and physical inactivity, also increase the risk for AD [1, 3, 5]. Individuals with more years of formal education have a lower risk of AD and dementia [1, 3]. The rationale for this phenomenon is that more years of education is associated with increased neural connections, allowing alternate routes of neuron-to-neuron communication to facilitate cognitive tasks, commonly known as the cognitive reserve hypothesis [1]. The cognitive reserve hypothesis is also used to explain why individuals who engage in social and cognitive activity lower their risk of AD and other dementias [1].

AD has a very potent genetic determinant. Genome-wide association studies (GWAS) have identified 21 risk loci that affect AD phenotype (*CR1, BIN1, INPP5D, MEF2C, TREM2, CD2AP, HLA-DRB1/HLA- DRB5, EPHA1, NME8, ZCWPW1, CLU, PTK2B, PICALM, SORL1, MS4A4/MS4A6E, SLC24A4/RIN3, FER- MT2, CD33, ABCA7, CASS4, APOE*) [2, 6-8]. It is estimated that these loci account for 28% of AD's heritability and 30% of the familial risk [6]. Having a first-degree relative affected by AD increases the relative risk for AD (3.5 95% CI 2.6-4.6) [6]. Studies have found that 30-48% of AD patients had an affected first-degree relative [6]. The most prominent genetic risk factor for AD was once thought to be the *APOE* gene, but comprehensive European and international GWASs have identified novel risk genes associated with the amyloid β

pathway, and many of these genes are linked to the immune system, such as the *CLU, CR1, ABCA7, CD33, EPHA1* and *MS4A* gene clusters [4, 6, 9]. Several more AD risk genes correlated with synaptic function (*PICALM, CD33, CD2AP, EPHA1, and BIN1*) and lipid metabolism were also identified (*CLU* and *ABCA7*) [6].

Conclusively identifying which genes are responsible for AD risk has been elusive. Several strategies have been employed to identify these AD risk genes, including transcriptome, proteome, and methylome analysis. A study by Humphries et al. found that RNA sequences taken from AD patients vs. cognitive healthy controls disclosed differences in the expression patterns of two genes in AD patients, *TSC22D4* and *C7orf61* [10, 11]. This study also revealed methylation differences in two genes specific to late-onset AD, *PILRA* and *PILRB* [10]. However, single nucleotide polymorphism (SNP)-based GWASs are the most elementary method for gene-based analysis and have successfully identified several genes linked to AD. However, obtaining sufficient statistical power can be problematic due to multiple SNPs being correlated or their cumulative effect being correlated with the disease phenotype [11]. Furthermore, SNP-based GWASs require a rigid significance threshold of $5x10^{-8}$ to control false positives, and significant SNPs may be surrogate variants because they are in linkage disequilibrium with other SNPs that are the true disease risk variant, which contributes to the issue of obtaining adequate power [12].

Gene-based GWASs have several advantages. First, gene-based analysis improves statistical power by collectively analyzing all variants within a gene to obtain a single representative p-value for the significance of association [12]. Secondly, because genes are a basic unit of the human genome, and genes are distinctly consistent over different populations, the results from

gene-based analysis are more consistent than the results from SNP-based GWASs [12]. Thirdly, the multiple-testing correction problem associated with analyzing millions of SNPs is reduced with gene-based analysis, which requires correction for an estimated 200,000 to 300,000 genes [12]. And finally, because genes are used as the analysis unit, making inferences about biological pathways and protein-protein interactions is uncomplicated [12].

The objective of this study was to identify novel genes associated with Alzheimer's disease using gene-based genome-wide association. We used summary genome-wide association study (GWAS) data from the International Genomics of Alzheimer's Project (IGAP) and genomic data from the 1000 Genomes Project to conduct our investigation.

Methods

IGAP is a two stage meta-analysis of 4 published genome-wide association studies (The European Alzheimer's disease Initiative, EADI; the Alzheimer Disease Genetics Consortium, ADGC; the Cohorts for Heart and Aging Research in Genomic Epidemiology consortium, CHARGE; the Genetic and Environmental Risk in AD consortium, GERAD) [13]. Summary data consisted of over 7 million genotyped or imputed single nucleotide polymorphisms (SNPs) from the stage 1 meta-analysis of European participants, 17,008 Alzheimer cases and 37,154 controls. The stage 2 meta-analysis of 8,572 AD cases and 11,312 controls yielded 11,632 SNPs genotyped and tested for association and combined stage 1 and 2 p-values [13]. SNPs with a call rate of less than 95% were excluded, and only SNPs genotyped or imputed at a minimum of 40% for both AD cases and controls were analyzed. The IGAP dataset included variables such as chromosome location, SNP position, SNP rsID, coded allele, non-coded allele, overall effect size

for the coded allele, overall standard error for the coded allele and meta-analysis p-value. The genome data was obtained from the 1000 Genomes Project, a collaboration of researchers from the US, UK, China, and Germany that produced nearly all genetic variants in regions of the genome determined to be associated with disease [12, 13].

Statistical Analysis

To assess the gene-level association, we used GATES, a modified Simes test and an effective chi square/versatile gene-based test for genome-wide association studies (ECS/VEGAS) test. Both Gates and ECS combine the p-values from SNP-based GWASs to obtain an overall gene-level p-value. The Gates test produces the overall gene p-value as follows:

$$PG = \text{Min}\left(\frac{meP(j)}{me(j)}\right)$$

Where m_e is the number of p-values amid the m SNPs and $m_{e(j)}$ is the number of p-values amid the most significant SNPs [12, 14-17]. P-values are estimated by the following equation:

$M-\sum[I(\lambda i>1)(\lambda i-1)]Mi=1\lambda i>0$

where λi is the ith eigenvalue for the SNP-based p-value correlation coefficient matrix [12, 14-17]. The null hypothesis for the GATES test is that no SNP within the gene is associated with the disease outcome [12, 16]. All analyses were conducted using the Knowledge-based mining system for Genome-wide Genetic studies (KGG4.0). The ECS/VEGAS test sums SNP-based Chi square test within a gene into a gene-based test statistic

Results

The Manhattan plot (Figure 4.1) displays the results of the GATES association test for the stage 1 IGAP GWAS. Five genes were determined to be significant and were located on chromosomes 6, 7, and 11: *HLA-DRB6, EPHA1, EPHA1-AS1, PICALM, SORLI*, respectively. However, none of these genes were novel. Results from the effective chi square test yielded 7 highly significant genes on chromosomes 1, 5, 78, 11, and 16 (Figure 4.2) not captured by the GATES test: genes *CR1, VAV3, LOC101929719, ZYX, EPHX2, NUP160*, and *MTSS1L,* respectively.

Table 4.1 displays the top 20 significant genes resulting from the KGG4 gene scan for associated AD genes from IGAP stage 1 using GATES and ECS tests. Protein coding genes *CBLC, MIR6503, MTSS1L* were found to be novel genes, p-values 2.73E-47, 1.46E-09, 2.1E-4, respectively. The combined stage 1 and 2 IGAP KGG4 top 20 GATES and ECS scan results, as shown I Table 4.2, confirmed the GATES stage 1 *CBLC* gene finding and identifies *LOC101929719* as a novel gene.

Discussion

We were successful in identifying 4 novel AD susceptibility genes using publicly available GWASs of AD data from IGAP and genomic data from the 1000 Genomes Project. Our gene-based GWAS was implemented using GATES, an extended Simes test that requires only the SNP-based GWAS p-values. This test is optimal when there is only one or a few risk SNPs within the gene [12, 14]. The ECS test option is more potent for genes with many dense independent risk SNPs [12]. We implemented the GATES and ECS using KGG4.0, a software instrument specifically designed for the secondary analysis of SNP-based GWAS as well as gene-pair-based and gene-set based association analyses [12]. KGG4.0 is executed in Java with an easy to use graphic interface to expedite analysis and can process approximately 10 million SNPs within hours using a 15GB RAM computing system [12].

We were able to identify two novel loci through ECS*, LOC101929719 and MTSS1L*. *LOC101929719* is a RNA gene that regulates expression, and *MTSS1L* is a protein coding gene involved in actin binding and cytoskeletal adaptor activity with no known disease associations [19].

A major strength of our study was that we were able to obtain a large sample of 11,632 SNPs associated with AD and robustly test genes associated with AD. We were able to successfully identify 8 novel genes associated with AD using GATES and ECS test procedures. Additionally, GATES has been proven capable of attaining statistically valid gene-level p-values as well as accurate type I error rates in both simulated and permutated datasets [15, 18]. This test performs very well when the reference population and study population match closely [12]. In our study, we used IGAP participants who were of European ancestry and European participants from the 1000 Genome Project.

In conclusion, gene-based analysis can facilitate uncomplicated inferences about biological pathways and protein-protein interactions [12, 15]. However, gene-based analysis can only capture SNPS near or within genes; therefore, gene-based analysis should be complemented by SNP-based tests of SNPs outside of genes [12].

References

1. Association, A.S., *2017 Alzheimer's Disease Facts and Figures.* 2017. **13**: p. 325-373.
2. Swaminathan, S., et al., *Amyloid pathway-based candidate gene analysis of [(11)C]PiB-PET in the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort.* Brain Imaging Behav, 2012. **6**(1): p. 1-15.
3. Beydoun, M.A., et al., *Epidemiologic studies of modifiable factors associated with cognition and dementia: systematic review and meta-analysis.* BMC Public Health, 2014. **14**: p. 643.
4. Ashare, R.L., et al., *APOE varepsilon4, an Alzheimer's disease susceptibility allele, and smoking cessation.* Pharmacogenomics J, 2013. **13**(6): p. 538-43.
5. Barnes, D.E. and K. Yaffe, *The projected effect of risk factor reduction on Alzheimer's disease prevalence.* Lancet Neurol, 2011. **10**(9): p. 819-28.
6. Cuyvers, E. and K. Sleegers, *Genetic variations underlying Alzheimer's disease: evidence from genome-wide association studies and beyond.* The Lancet Neurology, 2016. **15**(8): p. 857-868.
7. Chaudhry, M., et al., *Genetic variation in imprinted genes is associated with risk of late-onset Alzheimer's disease.* J Alzheimers Dis, 2015. **44**(3): p. 989-94.
8. Bettens, K., K. Sleegers, and C. Van Broeckhoven, *Genetic insights in Alzheimer's disease.* The Lancet Neurology, 2013. **12**(1): p. 92-104.
9. Jones, L., et al., *Genetic evidence implicates the immune system and cholesterol metabolism in the aetiology of Alzheimer's disease.* PLoS One, 2010. **5**(11): p. e13950.
10. Humphries, C., et al., *Alzheimer disease (AD) specific transcription, DNA methylation and splicing in twenty AD associated loci.* Mol Cell Neurosci, 2015. **67**: p. 37-45.
11. Kang, G., B. Jiang, and Y. Cui, *Gene-based Genomewide Association Analysis: A Comparison Study.* Curr Genomics, 2013. **14**(4): p. 250-5.

12. Li, M.X., et al., *GATES: a rapid and powerful gene-based association test using extended Simes procedure.* Am J Hum Genet, 2011. **88**(3): p. 283-93.

13. Lambert, J.C., et al., *Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease.* Nat Genet, 2013. **45**(12): p. 1452-8.

14. Ma, L., A.G. Clark, and A. Keinan, *Gene-based testing of interactions in association studies of quantitative traits.* PLoS Genet, 2013. **9**(2): p. e1003321.

15. Gui, H., et al., *Comparisons of seven algorithms for pathway analysis using the WTCCC Crohn's Disease dataset.* BMC Res Notes, 2011. **4**: p. 386.

16. Bacanu, S.A., *On optimal gene-based analysis of genome scans.* Genet Epidemiol, 2012. **36**(4): p. 333-9.

17. Petersen, A., et al., *Assessing methods for assigning SNPs to genes in gene-based tests of association using common variants.* PLoS One, 2013. **8**(5): p. e62161.

18. Li, M.X., J.S. Kwan, and P.C. Sham, *HYST: a hybrid set-based test for genome-wide association studies, with application to protein-protein interaction-based association analysis.* Am J Hum Genet, 2012. **91**(3): p. 478-88.

19. www.genecards.org. accessed 3/16/18.

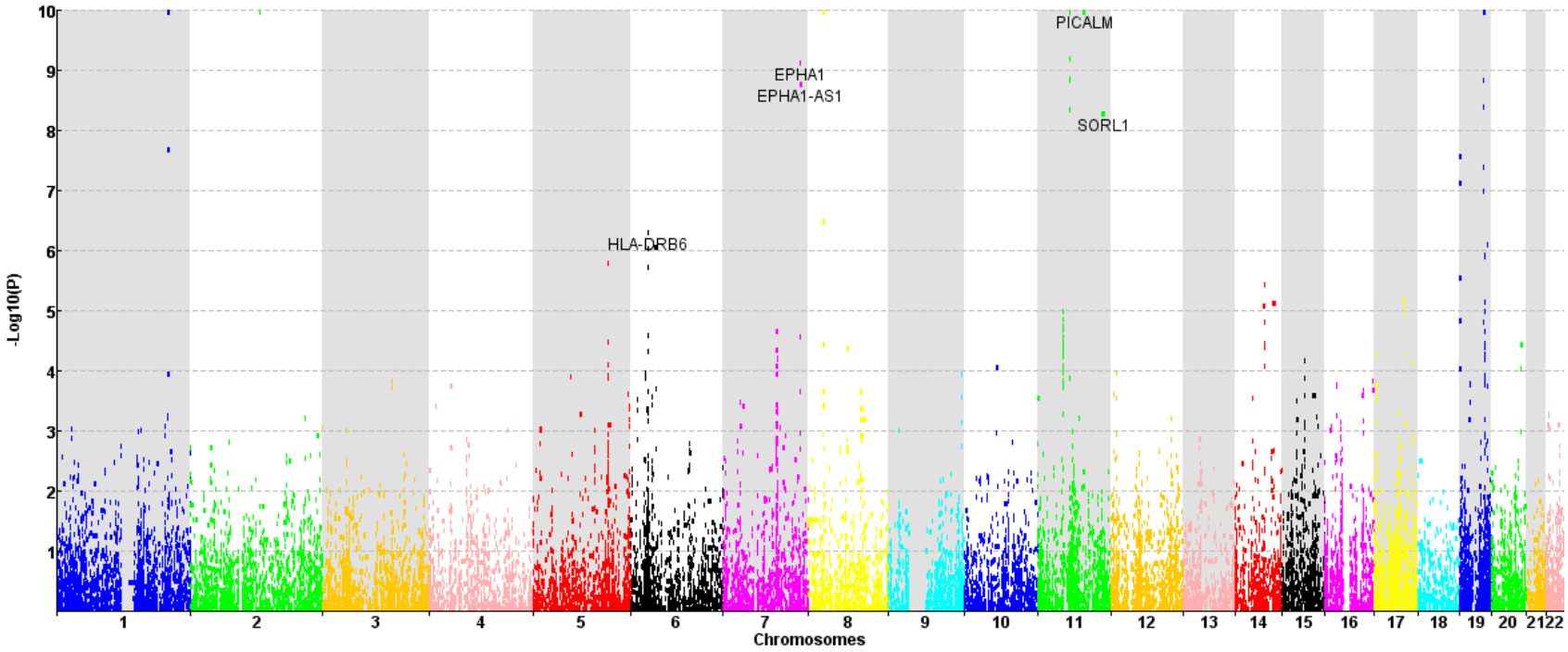

Figure 4.1. Gates IGap stage 1 Manhattan plot

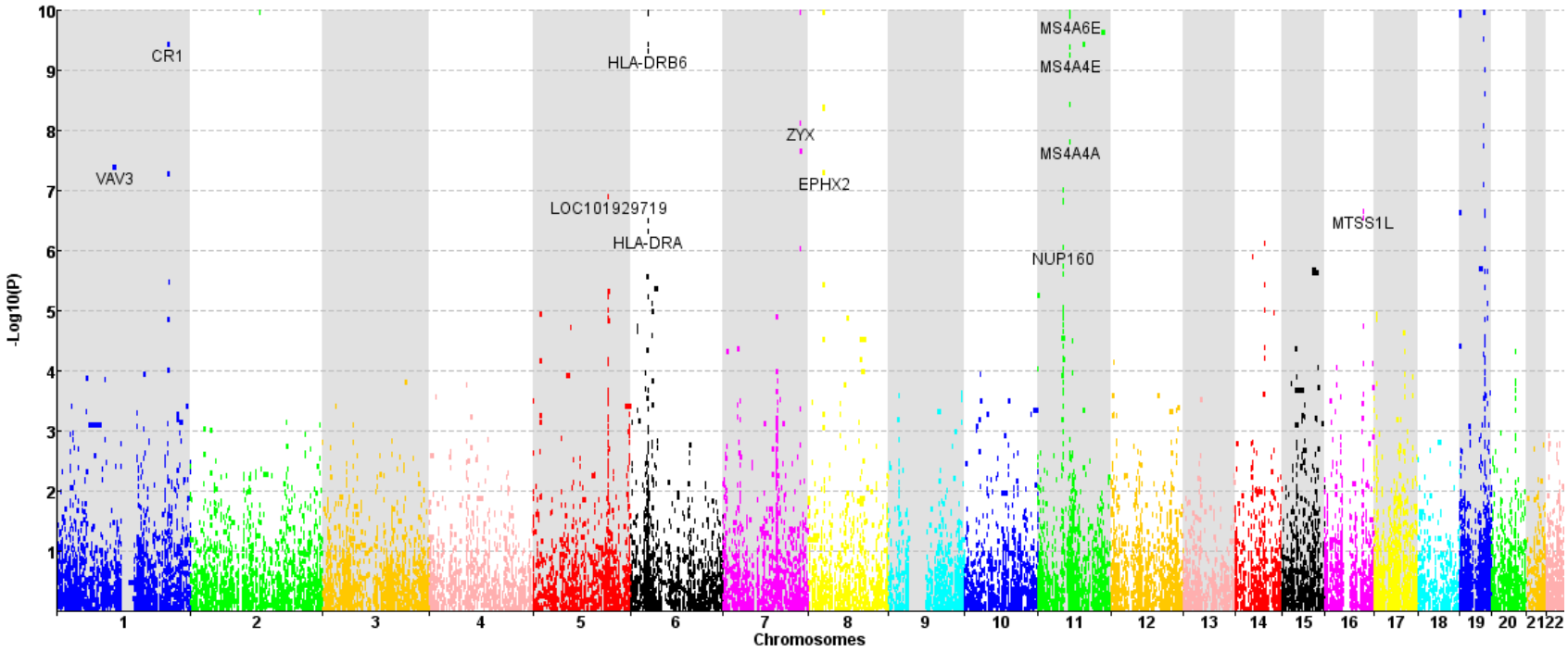

Figure 4.2. ECS IGAP combined 1 and 2 Manhattan plot

Table 4.1. Top 20 significant genes associated with AD stage 1(KGG4 gene scan results)

Stage 1 Gates							Stage 1 ECS						
Gene Names	Chr	Start Position	Group	*Loci	p-value	Report by Other GWAS	Gene Names	Chr	Start Position	Group	Loci	p-value	Report by Other GWAS
CR1	1	207669472	protein-coding gene		1.85E-13	Y	BIN1	2	127805598	protein-coding gene		1.68E-16	Y
BIN1	2	127805598	protein-coding gene		2.89E-14	Y	HLA-DRB5	6	32485153	protein-coding gene		4.26E-22	Y
EPHA1	7	143088204	protein-coding gene		7.63E-10	Y	MIR6843	8	27468117	non-coding RNA		1.69E-14	Y
EPHA1-AS1	7	143104905	non-coding RNA		1.75E-09	Y	CLU	8	27454433	protein-coding gene		1.7E-13	Y
CLU	8	27454433	protein-coding gene		5.79E-16	Y	BCAM	19	45312315	protein-coding gene		2.04E-107	Y
MIR6843	8	27468117	non-coding RNA		6.15E-16	Y	BCL3	19	45251977	protein-coding gene		1.61E-70	Y
MS4A6E	11	60102354	protein-coding gene		9.72E-12	Y	RELB	19	45504706	protein-coding gene		3.25E-53	Y
MS4A4A	11	60048013	protein-coding gene		9.86E-11	Y	CBLC	19	45281125	protein-coding gene		2.73E-47	N
PICALM	11	85668213	protein-coding gene		1.06E-10	Y	CLPTM1	19	45458480	protein-coding gene		6.94E-34	Y
MS4A6A	11	59939970	protein-coding gene		6.5E-10	Y	EXOC3L2	19	45715634	protein-coding gene		1.59E-24	Y
MIR6503	11	59976543	non-coding RNA		1.46E-09	N	TRAPPC6A	19	45666185	protein-coding gene		5.46E-22	Y
MS4A4E	11	59968725	protein-coding gene		1.5E-09	Y	NKPD1	19	45653007	protein-coding gene		1.57E-19	Y
BCAM	19	45312315	protein-coding gene		7.63E-68	Y	ABCA7	19	1040101	protein-coding gene		2.32E-19	Y
CBLC	19	45281125	protein-coding gene		3.27E-42	N	PPP1R37	19	45596430	protein-coding gene		1.25E-18	Y
BCL3	19	45251977	protein-coding gene		8.31E-42	Y	BLOC1S3	19	45682002	protein-coding gene		3.04E-17	Y
NKPD1	19	45653007	protein-coding gene		4.95E-20	Y	PVR	19	45147097	protein-coding gene		2.17E-16	Y
PPP1R37	19	45596430	protein-coding gene		9.25E-20	Y	CEACAM16	19	45202420	protein-coding gene		9.01E-16	Y
CLPTM1	19	45458480	protein-coding gene		1.2E-19	Y	CLASRP	19	45542297	protein-coding gene		1.08E-15	Y
TRAPPC6A	19	45666185	protein-coding gene		4.57E-18	Y	APOC4	19	45445494	protein-coding gene		5.04E-14	Y

Table 4.2. Top 20 significant genes associated with AD – stage1 and 2(KGG4 gene scan results)

Stage 1&2 Combined Gates							Stage 1&2 Combined ECS						
Gene Names	Chr	Start Position	Group	*Loci	p-value	Report by Other GWAS	Gene Names	Chr	Start Position	Group	Loci	p-value	Report by Other GWAS
CR1	1	207669472	protein-coding gene		1.85E-13		BIN1	2	127805598	protein-coding gene		1.68E-16	Y
BIN1	2	127805598	protein-coding gene		2.89E-14		HLA-DRB5	6	32485153	protein-coding gene		4.26E-22	Y
CLU	8	27454433	protein-coding gene		5.79E-16		MIR6843	8	27468117	non-coding RNA		1.69E-14	Y
MIR6843	8	27468117	non-coding RNA		6.15E-16		CLU	8	27454433	protein-coding gene		1.7E-13	Y
MS4A6E	11	60102354	protein-coding gene		9.72E-12		BCAM	19	45312315	protein-coding gene		2E-107	Y
BCAM	19	45312315	protein-coding gene		7.63E-68		BCL3	19	45251977	protein-coding gene		1.61E-70	Y
CBLC	19	45281125	protein-coding gene		3.27E-42		RELB	19	45504706	protein-coding gene		3.25E-53	Y
BCL3	19	45251977	protein-coding gene		8.31E-42		CBLC	19	45281125	protein-coding gene		2.73E-47	Y
NKPD1	19	45653007	protein-coding gene		4.95E-20		CLPTM1	19	45458480	protein-coding gene		6.94E-34	Y
PPP1R37	19	45596430	protein-coding gene		9.25E-20		EXOC3L2	19	45715634	protein-coding gene		1.59E-24	Y
CLPTM1	19	45458480	protein-coding gene		1.2E-19		TRAPPC6A	19	45666185	protein-coding gene		5.46E-22	Y
TRAPPC6A	19	45666185	protein-coding gene		4.57E-18		NKPD1	19	45653007	protein-coding gene		1.57E-19	Y
APOC4	19	45445494	protein-coding gene		2.2E-17		ABCA7	19	1040101	protein-coding gene		2.32E-19	Y
APOC4-APOC2	19	45445494	other		2.66E-17		PPP1R37	19	45596430	protein-coding gene		1.25E-18	Y
BLOC1S3	19	45682002	protein-coding gene		3.1E-17		BLOC1S3	19	45682002	protein-coding gene		3.04E-17	Y
RELB	19	45504706	protein-coding gene		5.19E-17		PVR	19	45147097	protein-coding gene		2.17E-16	Y
APOC2	19	45449238	protein-coding gene		6.97E-16		CEACAM16	19	45202420	protein-coding gene		9.01E-16	Y
CLASRP	19	45542297	protein-coding gene		1.17E-15		CLASRP	19	45542297	protein-coding gene		1.08E-15	Y
EXOC3L2	19	45715634	protein-coding gene		9.03E-14		APOC4	19	45445494	protein-coding gene		5.04E-14	Y
PVR	19	45147097	protein-coding gene		2.64E-12		LOC105372419	19	45588586	unknown		6.2E-13	Y

CHAPTER FIVE

CONCLUSION

We demonstrated that trajectory analysis can perform better than the widely used MMSE at predicting cognitive impairment and that trajectory analysis may be a viable alternative when cognitive assessments are not congruent in subsequent waves or secondary datasets provided that the assessments are collected consistently over time. We also demonstrated that heart disease is associated with low trajectory which confirms the hypothesis that cardiovascular pathways may contribute to cognitive impairment. This evidence provides the rationale for our next to two objectives, determining that coronary artery disease is likely a casual risk factor for cognitive impairment and AD by Mendelian randomization, and that multiple genes may contribute to the development cognitive impairment and AD.

This study was able to identify CAD as a casual factor in the development of AD using a two-sample Mendelian randomization that incorporates the IVW regression method. We then tested that the MR assumptions were not violated using the Egger regression method. MR is a robust and flexible method for investigating causality using individual or summary level data.

This study also successfully identified 4 novel AD susceptibility genes using publicly available GWAS of AD data from IGAP and genomic data from 1000 Genomes Project. Our gene-based GWAS was implemented using GATES, an extended Simes test that requires only the SNP-based

GWAS p-values. Gene-based analysis can also facilitate uncomplicated inferences about biological pathways and protein-protein interactions.