Socio-Contextual and Multi-Omic Associations with Cognitive Function and Structural Brain Measures in Older African Americans

by

Dima L. Chaar

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Epidemiological Science)
in the University of Michigan
2023

Doctoral Committee:

Associate Professor Jennifer A. Smith, Chair
Professor Sharon L.R. Kardia
Research Assistant Professor Wei Zhao
Professor Xiang Zhou

Dima L. Chaar

dimac@umich.edu

ORCID iD: 0000-0001-6755-0179

© Dima L. Chaar 2023

Dedication

This dissertation is dedicated to my parents Ferial and Jarir, my brother Jad and my husband Rami. I love you all infinitely.

Acknowledgements

I would like to extend my gratitude to many individuals who have supported me during this dissertation process. First, I would like to thank my chair, Jennifer Smith, for her mentorship, guidance, patience and commitment to my success during the PhD program. From creating an environment from which I was able to learn genetic epidemiology to pushing my research and critical thinking skills to the next level. I have learnt lot from Jen in the past four years and this dissertation would not have been possible without her mentorship and support.

I would also like to my dissertation committee members for their support and expertise on this journey. Thanks to Wei Zhao for her helping me hone my coding and critical thinking skills, especially for helping me overcome obstacles associated with Aim 1 in the midst of the Covid-19 pandemic. I thank Xiang Zhou for his statistical insights and knowledge, and for always being available to answer my questions. Finally, I thank Sharon Kardia for all her support and mentorship which opened the doors to this journey and helped me throughout. I am also grateful to the participants of the Genetic Epidemiology Network of Arteriopathy for their participation and sharing of data that made this dissertation possible. I am also thankful to the Department of Epidemiology faculty and staff for their support throughout the years and for accepting me into the PhD program.

I am also grateful to be a part of the Genomics Research Collaborative which has been a helpful space to learn, collaborate and connect with others. I would like to thank members of our lab, past and present, including: Farah Ammous, Zheng Li, Yi Zhe Wang, Lisha Lin, Lauren

Opsasnick, Jiacong Du, MJ Kho, Miao Yu, Lulu Shang, Scott Ratliff, Chris Crowe, Hasan Abu Amara, Patricia Peyser, and Lawrence Bielak. I would also like to thank the members of the Center for Social Epidemiology and Population Health (CSEPH) as well as my PhD cohorts for their friendship and encouragement. I would also like to thank my friends/family for their love and support including Roslyn Glover, Heather Sommer, Dahlia Musharrafieh, Rana Andary, Sarah Abi-Chahine, Aida Charafeddine, Lama Aldeeb, Soula Saasouh, Nellie Said, and Ji Youn Shin.

Finally, I want to thank my family for their endless love and support throughout this dissertation process. I am grateful to my parents Ferial and Jarir who encouraged me and supported me throughout this entire journey. Thank you to my Mom and Dad for always believing in me and encouraging me to work hard and reach my goals. Thank you for your endless support, sacrifices and love, for inspiring me everyday, and for your words of advice/encouragement when I needed it the most. Thank you to my brother Jad for his love and support through this process. Thank you for always picking me up when I feel down and for always being there for me. And lastly, thank you to my husband and best friend Rami for his endless patience, love, dedication and support. I am grateful for his encouragement of my professional goals, for always believing in me and for being “my person” who brings harmony to my soul. I love you all infinitely.

Preface

Chapter II of this dissertation has been published as Chaar DL, Nguyen K, Wang Y-Z, Ratliff SM, Mosley TH, Kardia SLR, Smith JA, Zhao W. SNP-by-CpG Site Interactions in *ABCA7* Are Associated with Cognition in Older African Americans. *Genes*. 2022; 13(11):2150. https://doi-org.proxy.lib.umich.edu/10.3390/genes13112150.

Table of Contents

Dedication ii

Acknowledgements iii

Preface v

Table of Contents vi

List of Tables x

List of Figures xiii

List of Abbreviations xv

Abstract xvi

Chapter 1 . Introduction 1

1.1 Overview 1

1.2 Specific Aims 3

1.2.1 Aim 1 5

1.2.2 Aim 2 6

1.2.3 Aim 3 7

1.3 Background 9

1.3.1 Preclinical dementia and the dementia continuum 9

1.3.2 Cognitive function and brain structure 10

1.3.3 Individual-level and neighborhood risk factors for cognitive impairment, cognitive decline and dementia 13

1.3.4 Role of genetic factors in Alzheimer's disease and cognitive function 14

1.3.5 Role of epigenetic and transcriptomic factors in Alzheimer's disease and cognitive function 16

1.3.6 The importance of multi-omics and socio-contextual research in African Americans 19

1.4 Study Design and Measures in The Genetic Epidemiology Network of Arteriopathy 20

1.4.1 Study design and source population 20

1.4.2 Measures of cognitive function and brain structure 22

1.4.3 Genetic (SNP) data 25

1.4.4 DNA methylation data 25

1.4.5 Gene expression data 27

1.5 Summary 27

1.6 References 28

Chapter 2 . SNP-by-CpG Interactions in *ABCA7* are Associated with Cognition in Older African Americans 45

2.1 Abstract 45

2.2 Introduction 46

2.3 Materials and Methods 49

2.3.1 Sample 49

2.3.2 Measures 50

2.3.3 Statistical analysis 54

2.4 Results 58

2.4.1 Sample characteristics 58

2.4.2 Correlation among six cognitive outcomes 58

2.4.3 Correlation among ABCA7 SNPs 59

2.4.4 Genetic associations 59

2.4.5 Epigenetic associations 60

2.4.6 Genetic-epigenetic associations 60

2.4.7 Gene expression associations 62
2.5 Discussion 63
2.6 Conclusion 69
2.7 References 70
2.8 Tables 78
2.9 Figures 81
2.10 Supplementary Material 83
Chapter 3 . Neighborhood Environment Associations with Cognitive Function and Structural Brain Measures in Older African Americans 99
3.1 Abstract 99
3.2 Introduction 100
3.3 Materials and Methods 103
3.3.1 Sample 103
3.3.2 Measures 105
3.3.3 Statistical analysis 109
3.4 Results 112
3.4.1 Sample Characteristics 112
3.4.2 Correlation among cognitive and WMH outcomes 113
3.4.3 Correlation among the neighborhood exposures 113
3.4.4 Associations between neighborhood characteristics and cognitive/WMH outcomes 114
3.4.5 Mediation analysis 116
3.5 Discussion 117
3.6 Conclusion 124
3.7 References 125
3.8 Tables 134
3.9 Figures 139

3.10 Supplementary Methods 142

3.11 Supplementary Material 149

Chapter 4 . Multi-Ancestry Transcriptome-Wide Association Studies of Cognitive Function, White Matter Hyperintensity, and Alzheimer's Disease 163

4.1 Abstract 163

4.2 Introduction 164

4.3 Materials and Methods 166

4.3.1 Sample 166

4.3.2 Measures 167

4.3.3 Statistical Methods 172

4.4 Results 176

4.5 Discussion 180

4.6 Conclusion 186

4.7 References 187

4.8 Tables 197

4.9 Figures 201

4.10 Supplementary Material 217

Chapter 5 . Conclusion 230

5.1 Summary and Implications of Main Findings 230

5.2 Strengths and Limitations 236

5.3 Future Directions 237

5.4 Conclusion 239

5.5 References 240

List of Tables

Table 1-1. Descriptions of cognitive functions and neurocognitive domains associated with each cognitive test and measure. 23

Table 2-1. Sample characteristics of Genetic Epidemiology Network of Arteriopathy (GENOA) African Americans (N=634). 78

Table 2-2. Interaction of *ABCA7* sentinel SNPs and CpG sites on general cognitive function (FDR q<0.1; N=494). 79

Table 2-3. Estimated effect of CpG site on general cognitive function for given ABCA7 SNP genotype group (N=494). 80

Table S2-4. Pearson's correlations among the six cognitive measures (n=634) 83

Table S2-5. Pearson's correlations among the five sentinel ABCA7 SNPs (n=634) 84

Table S2-6. Association between ABCA7 sentinel SNPs and general cognitive function (n=634) 85

Table S2-7. Association of CpGs in the ABCA7 region and general cognitive function (p<0.05; n=494) 86

Table S2-8. Interaction between ABCA7 sentinel SNPs and CpG sites on general cognitive function (p<0.05; n=494) 87

Table S2-9. Pearson's correlations among ABCA7 CpG sites[a] (n=494) 88

Table S2-10. Estimated effect of CpG site on general cognitive function for given ABCA7 SNP genotype group, after excluding outlying values for CpG sites[a] 89

Table S2-11. Estimated effect of CpG[a] site on general cognitive function for given ABCA7 SNP genotype group, after adjusting for SNP effect. 90

Table S2-12. Interaction between ABCA7 sentinel SNPs and CpG sites on neurocognitive measurements (n=494) 91

Table S2-13. Interaction between ABCA7 sentinel SNPs and CpG sites[a] on transcripts in the ABCA7 gene region (p<0.05; n = 429) 92

Table S2-14. Estimated effect of CpG site on ABCA7 transcripts for given ABCA7 SNP genotype group (n=429) 93

Table S2-15. Association of SNPs[a] on transcripts in the ABCA7 gene region (p<0.05; n=429). 94

Table S2-16. Association of CpG sites[a] on transcripts in the ABCA7 region (p<0.05; n=429)... 95

Table 3-1. Sample characteristics of Genetic Epidemiology Network of Arteriopathy (GENOA) African Americans (N = 542) 134

Table 3-2. Associations between neighborhood socioeconomic disadvantage and cognitive function/ White matter hyperintensity 135

Table 3-3. Associations between neighborhood socioeconomic disadvantage and cognitive measures (Model 2a; N=477).. 136

Table 3-4. Associations between simple density of neighborhood destinations per square mile for 1-mile buffer size and cognitive function/WMH.. 137

Table 3-5. Associations between simple density of neighborhood destinations per square mile for 1-mile buffer size and cognitive measures (Model 2a; N=477) 138

Table S3-6. Pearson's correlations among the cognitive/WMH outcomes (N=466) 149

Table S3-7. Pearson's correlations among neighborhood socioeconomic disadvantage and neighborhood simple density measures per square mile for 1-mile buffer size (N=542)........... 150

Table S3-8. Associations among neighborhood socioeconomic disadvantage and neighborhood simple density measures per square mile for 1-mile buffer size after adjusting for census tract population density (N=542) 151

Table S3-9. Pearson's correlations among neighborhood socioeconomic disadvantage and simple and kernel densities per square mile for 1-mile buffer size (N=542)[a] 152

Table S3-10. Associations between simple density of neighborhood destinations per square mile for ½-, 1- and 3- mile buffer sizes and cognitive function/WMH 153

Table S3-11. Associations between simple density of neighborhood destinations per square mile for ½-, 1- and 3- mile buffer sizes and cognitive measures (N=477) 155

Table S3-12. Associations between kernel density of neighborhood destinations per square mile for ½-, 1- and 3- mile buffer sizes and cognitive function/WMH 157

Table S3-13. Associations between kernel density of neighborhood destinations per square mile for ½-, 1- and 3- mile buffer sizes and cognitive measures (N=477) 159

Table 4-1 Sample characteristics of expression quantitative trait locus (eQTL) mapping study and genome-wide association studies (GWAS) participants. 197

Table 4-2. Genes for WMH identified both by METRO followed by fine-mapping with FOCUS and by TWAS-Fusion conducted by Sargurupremraj et al. (2020) 198

Table 4-3. Genes for AD identified both by METRO followed by fine-mapping with FOCUS and by TWAS-Fusion followed by fine-mapping with FOCUS conducted by Bellenguez et al. (2022) 199

Table S4-4. Genes associated with general cognitive function using METRO followed by fine-mapping with FOCUS (N=266 genes; $P<2.9x10^{-6}$)........ 217

Table S4-5. Genes associated with white matter hyperintensity using METRO followed by fine-mapping with FOCUS (N=23 genes; $P<2.9x10^{-6}$)........ 222

Table S4-6. Genes associated with Alzheimer's disease using METRO followed by fine-mapping with FOCUS (N=69 genes; $P<2.9x10^{-6}$)........ 223

List of Figures

Figure 1-1. Conceptual model of three aims in GENOA. Arrows refer to mechanistic pathways in Aim 1 (green), Aim 2 (blue), and Aim 3 (yellow). 5

Figure 2-1. Regional plot of the association between DNA methylation in the ABCA7 region and general cognitive function. 81

Figure 2-2. Linear prediction of CpG sites (% methylated) on general cognitive function for a given SNP genotype group in the ABCA7 region. 82

Figure S2-3. Flow diagram illustrating sample sizes for genetic (n = 634), epigenetic (n = 494), and transcriptomic (n = 429) analyses in GENOA AA. 96

Figure S2-4. Models used to assess genetic, epigenetic and genetic-epigenetic interaction associations with general cognitive function. 97

Figure S2-5. Transcript expression of ABCA7: ENSG00000064687 (12 ATP binding cassette subfamily A member 7 98

Figure 3-1. Directed acyclic graph (DAG) of the hypothesized associations for the epigenetic mediation between neighborhood characteristics (exposures) and cognitive/WMH outcomes. 139

Figure 3-2. Distributions of cognitive and structural brain measures 140

Figure 3-3. Quantile-quantile plots for the epigenetic mediation of the associations between neighborhood characteristics and cognitive function. 141

Figure 3-4. Flow diagram illustrating sample sizes for neighborhood density and neighborhood socioeconomic disadvantage analyses for cognitive measures in GENOA African Americans. 161

Figure 3-5. Flow diagram illustrating sample sizes for neighborhood density and neighborhood socioeconomic disadvantage analyses for white matter hyperintensity in GENOA African Americans. 162

Figure 4-1. Manhattan plots of $-\log_{10}$ p-values for gene-trait associations in METRO. 201

Figure 4-2. Quantile-quantile plots of $-\log_{10}$ p-values for gene-trait associations in METRO... 204

Figure 4-3. Venn diagrams comparing number of genes associated with general cognitive function, white matter hyperintensity and Alzheimer's disease using METRO, prior to and following FOCUS fine-mapping. 205

Figure 4-4. Contribution weights of expression prediction models across all significant genes identified by METRO. 207

Figure 4-5. Venn diagram comparing number of METRO-identified genes associated with general cognitive function following FOCUS fine-mapping and genes identified by Davies et al. (2018) gene-based and SNP-based analyses. 208

Figure 4-6. Functional enrichment analysis on the fine-mapped gene set identified for general cognitive function using METRO TWAS (N=266 genes). 209

Figure 4-7. Venn diagram comparing number of METRO-identified genes associated with white matter hyperintensity following FOCUS fine-mapping and genes identified by Sargurupremraj et al. (2020) gene-based and SNP-based analyses. 210

Figure 4-8. Functional enrichment analysis on the fine-mapped gene set identified for white matter hyperintensity using METRO TWAS (N=23 genes). 211

Figure 4-9. Venn diagram comparing number of METRO-identified genes associated with Alzheimer's disease following FOCUS fine-mapping and genes identified by Bellenguez et al. (2020) gene prioritization and SNP-based analyses. 212

Figure 4-10. Functional enrichment analysis on the fine-mapped gene set identified for Alzheimer's disease using METRO TWAS (N=69 genes). 213

Figure 4-11. Venn diagram comparing METRO TWAS results prior to and following FOCUS fine-mapping with TWAS results from Sargurupremraj et al. (2020) and Bellenguez et al. (2022). 214

Figure S4-12. Functional enrichment analysis on the gene set identified by METRO for general cognitive function and AD (N=22 genes; $P<2.90 \times 10^{-6}$). 225

Figure S4-13. Functional enrichment analysis on the fine-mapped gene set not previously identified by Davies et al. (2018)[15] for general cognitive function using METRO TWAS (N=82 genes). 226

Figure S4-14. Functional enrichment analysis on the fine-mapped gene set not previously identified by Sargurupremraj et al. (2020)[305] for white matter hyperintensity using METRO TWAS (N=12 genes). 227

Figure S4-15. Functional enrichment analysis on the fine-mapped gene set not previously identified by Bellenguez et al. (2022)[112] for Alzheimer's disease using METRO TWAS (N=45 genes). 228

List of Abbreviations

AA	African Americans
AD	Alzheimer's disease
COWA-FAS	Controlled Oral Word Association Test
CpG	Cytosine-phosphate-guanine
DNAm	DNA methylation
DSST	Digit Symbol Substitution Test
EA	European Americans
EBV	Epstein-Barr virus
eQTL	expression quantitative trait loci
EWAS	Epigenome wide association study
FDR	False discovery rate
FLAIR	fluid-attenuated inversion recovery
FUPC	First unrotated principal component
GENOA	Genetic Epidemiology Network of Arteriopathy
GMBI	Genetics of Microangiopathic Brain Injury
GReX	genetically regulated gene expression
GWAS	Genome wide association study
LD	Linkage disequilibrium
LOAD	Late-onset Alzheimer's disease
MCI	Mild cognitive impairment
MMSE	Mini Mental State Exam
MRI	Magnetic resonance imaging
NHW	Non-Hispanic whites
PC	Principal component
PCA	Principal component analysis
RA	Risk allele
RAF	Risk allele frequency
RAVLT	Rey Auditory Verbal Learning Test
SCWT	Stroop Color-Word Test
SD	Standard deviation
SE	Standard error
SES	Socioeconomic status
SNP	Single nucleotide polymorphism
SVD	Small vessel disease
TMTA	Trail Making Test A
TWAS	Transcriptome wide association study
VaD	Vascular dementia
WMH	White matter hyperintensity

Abstract

Dementia affects approximately 1 in 10 persons aged 65 years and older in the U.S., and African Americans (AA) are more likely to develop dementia compared to European Americans (EA). However, the underlying molecular mechanisms and impact of their interactions with socio-contextual risk factors on cognitive function and brain structures in AA are not fully understood. This dissertation examines the molecular effects of genetic, epigenetic, and transcriptomic markers, as well as socio-contextual determinants of health, on cognitive function and white matter hyperintensity (WMH) prior to dementia onset in a well-curated cohort of older AA from the Genetic Epidemiology Network of Arteriopathy (GENOA) study. In Aim 1, we investigated whether single nucleotide polymorphisms (SNPs), epigenetic variants, and/or their interactions in the *ABCA7* gene region, which was previously associated with Alzheimer's Disease (AD) in AA, are associated with general cognitive function in cognitively normal older AA. Although *ABCA7* sentinel SNPs and CpG sites were not associated with general cognitive function, we did see evidence of SNP-by-CpG interactions. We found that rs3764647 and rs115550680 may regulate the effects of DNA methylation (DNAm) on cognitive function. As such, while AD risk SNPs in *ABCA7* were not associated with cognitive function in this sample, DNAm at local CpGs may influence cognitive function in people with specific *ABCA7* genotypes. In Aim 2, we assessed whether DNAm from peripheral blood leucocytes mediates the relationships between neighborhood characteristics and cognitive function/WMH in cognitively healthy AA, using high-dimensional mediation methods. For a 1-mile buffer around a participant's residence, each additional fast-food destination or unfavorable food store with

alcohol per square mile was associated with 0.05 (p=0.04) and 0.04 (p=0.04) second improvements in visual conceptual tracking score, respectively. Also, each additional alcohol drinking place per square mile was associated with a 0.62 word increase in delayed recall score (p=0.03), indicating better memory function. Although the presence of these destinations encourages unhealthy diet and behaviors, they may provide meeting places for community members that allow for greater interaction and stimulation of cognitive health. In this study, there was no evidence that DNAm mediated the observed associations between neighborhood characteristics and cognitive function. Further examination of the pathways between neighborhood characteristics and cognitive function/WMH may allow for development of behavioral, infrastructural, and pharmaceutical interventions to facilitate healthy brain aging in older AA. In Aim 3, we conducted a multi-ancestry transcriptome wide association study (TWAS) that leveraged gene expression data collected from EA and AA in GENOA, through a joint likelihood-based inference framework, to identify genes associated with general cognitive function, WMH, and AD. After fine-mapping within genomic regions, we identified 266, 23, and 69 genes associated with general cognitive function, WMH, and AD, respectively (Bonferroni-corrected alpha level =$P<2.9x10^{-6}$). These genes were enriched for innate immunity, vascular dysfunction, and neuroinflammation. The WMH and AD TWAS also indicated that downregulation of *ICA1L* may contribute to overlapping AD and vascular dementia (VaD) neuropathology. To our knowledge, this study is the first TWAS of cognitive function and neurocognitive disorders that used expression mapping studies in multiple ancestries. This work may expand TWAS studies beyond a single ancestry group to identify gene targets for pharmaceutical or preventative treatment for dementia. Together, these studies advance

knowledge of the relationships between multi-omic mechanisms and socio-contextual factors that contribute to neurocognitive outcomes and structural brain measures in older AA.

Chapter 1 . Introduction

1.1 Overview

Adult-onset dementia consists of a group of neurocognitive disorders caused by abnormal brain changes that result in a gradual and irreversible loss of neurons and brain functions. These brain changes may lead to the loss of memory, language, problem-solving and other cognitive functions, impacting an individual's daily life and independence.[1] Approximately one-third of adults aged 85 and older have some form of dementia. Although dementia is more common among older people, it is not a normal part of aging. Currently, there are approximately 55 million people diagnosed with dementia. As the proportion of older people worldwide increases, this prevalence is expected to rise to 78 million by 2030 and 139 million by 2050.[1,2]

Alzheimer's disease (AD) accounts for 60-80% of late-onset dementia cases. Other less-common forms of dementia include vascular dementia (VaD; 5-10%), Lewy body dementia (5-10%), frontotemporal dementia (5-10%), Huntington's and Parkinson's-related dementias and mixed dementia.[3] These forms of dementia are often difficult to distinguish from AD because they share many pathological features and cognitive symptoms. Notably, VaD often co-occurs with AD and is underdiagnosed.[4–6] Both AD and VaD are characterized by noticeable cognitive impairment in areas of episodic and semantic memory, as well as executive function. However, AD also shows aggregation of amyloid-beta protein and neurofibrillary tangles in brain tissue that may precede the illness by 10-20 years,[7,8] while VaD may be caused by reduced blood flow

to the brain as a result of small vessel disease (SVD) or following one or more strokes, and is commonly seen in hypertensive patients.[9] Since VaD and AD often coexist, it has been hypothesized that vascular changes and other brain abnormalities may interact in ways that increase the likelihood of cognitive impairment. A further challenge in the field is distinguishing between individuals who are aging normally from those aging pathologically with multiple forms of dementia.

A greater understanding of the pathological cascade of events that influence cognitive function and lead to cognitive decline in older adults is critical for early intervention during the long preclinical or prodromal phase prior to dementia onset.[10,11] Biological pathways related to lipid metabolism, inflammation and immune function have been linked to cognitive decline and preclinical AD.[12,13] Additionally, genetic factors have a strong influence on cognition and dementia. Cognitive ability is highly heritable (from 55% in adolescence to 66% in young adulthood in twin studies)[14] and hundreds of genetic loci are associated with individual differences in cognitive ability,[15,16] including a handful that have been previously associated with AD.[15,16] However, identifying biological pathways associated with cognition has proven challenging, in part because many identified genetic loci are located within intergenic non-coding regions[15] which do not directly code for proteins. Through advances in high-throughput technologies and the integration of multi-omic studies, we can better understand downstream pathways and how they interact with the environment to affect dementia[17] and cognitive pathologies.[18,19]

Although the primary risk factor for late-onset dementia is age, there are significant disparities in incidence and prevalence by race and ethnicity.[20–22] Several studies have found that African Americans (AA) have a greater burden of and risk for developing dementia compared to

Non-Hispanic Whites (NHW).[21,23–25] On average, AA perform lower than their white counterparts on cognitive tests, have higher prevalence of AD, and have higher incident risk of AD.[26] Cognitive function in AA is especially important to study during the preclinical period because unique combinations of socio-contextual or genetic exposures may influence the biological mechanisms that underlie racial/ethnic health disparities. For example, these differences in cognitive performance, cognitive reserve and AD risk in AA may in part be caused by racial disparities in education (amount and quality), access to material and social resources, exposure to discrimination, and exposure to neurotoxicants.[27,28] Potential biological mediators for these social influences on health disparities include plasma biomarkers,[29] genomic risk factors,[12,30–32] and the influences of epigenomic[33] and transcriptomic factors.[34] Further, dementia research has mainly focused on diagnosis, mechanisms, as well as management and treatment of disease among NHW. As such, the lack of biological and epidemiologic research among AA poses a barrier to understanding how cognitive aging and the development of dementia differ in racial and ethnic minorities, particularly in the AA population. Given the multifactorial and complex nature of cognitive decline preceding dementia, it is important to integrate multi-omic layers of data to better understand these disparities. This may allow the identification of targets for intervention and treatment, especially in populations that are most at risk.[35]

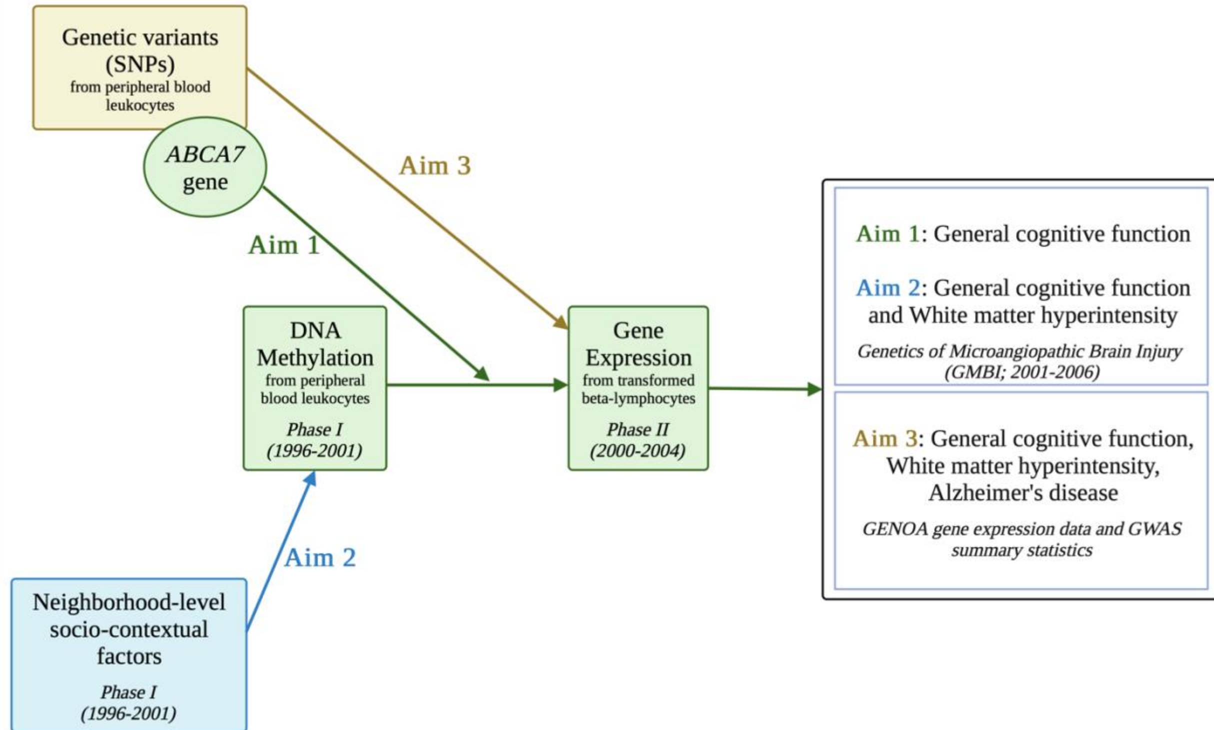
1.2 Specific Aims

In this dissertation, we will characterize the potential molecular effects of genetic, epigenetic, and transcriptomic markers, as well as socio-contextual determinants of health, on cognition and white matter hyperintensity (WMH) prior to dementia onset in a well-curated cohort of older AA adults from the Genetic Epidemiology Network of Arteriopathy (GENOA)

study. GENOA is one of the few studies to combine genetic, DNA methylation, and gene expression data with rich measures of socio-environmental context, cognitive function, and brain structure in a large cohort of AA. GENOA has both cross-sectional and longitudinal data (Phase I: 1995-2000, Phase II: 2000-2004, and several ancillary studies thereafter).[36]

Specifically, we will investigate whether single nucleotide polymorphisms (SNPs), epigenetic variants (CpGs) and/or their interactions in the *ABCA7* gene region, which was previously associated with AD in AAs, are associated with general cognitive function in cognitively normal older AAs (Aim 1). Next, we will investigate whether CpGs mediate the association between socio-contextual factors and cognitive/WMH outcomes in the same cohort of cognitively normal older AAs (Aim 2). Lastly, we will examine gene-trait associations for general cognitive function, WMH and AD to understand underlying transcriptomic mechanisms using multi-ancestry data from European Americans (EA) and AA (Aim 3). Our findings will assist in the ongoing efforts to better understand the etiological precursors of dementia and their impact on socioeconomic and racial/ethnic health disparities.

Figure 1-1. Conceptual model of three aims in GENOA. Arrows refer to mechanistic pathways in Aim 1 (green), Aim 2 (blue), and Aim 3 (yellow).



1.2.1 Aim 1

The *ABCA7* gene confers the largest genetic risk for AD in AA after the apolipoprotein E (*APOE)* ε4 allele.[37,38] However, the relationship between *ABCA7* and cognitive function has not been thoroughly examined. This aim will investigate whether previously identified AD risk SNPs in *ABCA7*, DNA methylation at CpG sites in *ABCA7* measured in peripheral blood leukocytes, and their interactions are associated with general cognitive function in 634 GENOA AA without dementia. To understand the potential functional consequences of our findings at the molecular level, we will also evaluate whether identified SNPs or CpGs are also associated with *ABCA7* gene expression from transformed beta lymphocytes in the same cohort. Studying the relationship between SNPs and CpGs in *ABCA7* and cognition may illuminate the role of *ABCA7*

in cognitive aging preceding AD. To our knowledge, this study will be the first assessment of the associations and interactions between DNA methylation and genetic risk factors in *ABCA7* on cognition in AA without dementia. Investigating the interplay of multi-omic markers and later-life cognition may help us characterize the underlying genetic architecture of cognition in older adults preceding dementia. It may also allow us to identify targets for intervention and treatment in AA, a population at high risk for AD and dementia.

Aim 1: To examine whether genetic and epigenetic variations in the *ABCA7* gene region, and/or their interactions, are associated with general cognitive function in older African Americans.
Hypothesis 1: *We hypothesize that a number of SNPs in ABCA7, DNA methylation sites in ABCA7, and their interactions are associated with general cognitive function in older AA from the GENOA study.*

1.2.2 Aim 2

To date, there are no treatments to prevent cognitive impairment or slow cognitive decline prior to onset of dementia. However, treating vascular risk factors, improving diet, and engaging in cognitively stimulating activities and environments may delay cognitive impairment.[39–41] In addition to individual health behaviors, socio-contextual factors such as low neighborhood socioeconomic status (SES), the presence of racial segregation, and low availability of healthy food, recreation, and social engagement are significant predictors of worsening cognitive health and increased susceptibility to dementia.[41] DNA methylation is associated with both cognitive function and WMH, as well as neighborhood-level disadvantage indicators; however, little is known about the role of DNA methylation in mediating the

associations between neighborhood-level factors and cognitive function or WMH. The few studies that have been conducted in this area focus primarily on EA and/or those with dementia, so additional research is needed to examine these pathways in other racial/ethnic groups and those without dementia.

In Aim 2, we will examine whether neighborhood-level factors (e.g., summary neighborhood SES as assessed by Census data and the densities of available healthy food, recreation, and social engagement) are associated with cognitive function and WMH in older AA without dementia. For significant associations, we will conduct epigenome-wide mediation analysis to identify CpG sites mediating the relationship between neighborhood factors and cognitive function/WMH using the Sobel-Comp method that assesses sparse mediation effects under the composite null hypothesis. Investigating DNA methylation as a mediator between neighborhood factors and cognitive function/WMH may help us understand potential underlying epigenetic pathways influencing cognitive function in older adults prior to the onset of dementia.

Aim 2: To examine whether DNA methylation in peripheral blood leukocytes mediates the relationship between neighborhood-level factors and cognitive function or white matter hyperintensity in older African Americans.

Hypothesis 2: *We hypothesize that associations of neighborhood-level socio-contextual factors with cognitive function and/or WMH are partially mediated by DNA methylation levels in older AA from the GENOA study.*

1.2.3 Aim 3

Genome-wide association studies (GWAS) have identified thousands of genetic variants associated with complex traits and diseases. However, GWAS results are difficult to interpret functionally because many potential causal variants may be located in non-coding regions, and their associations with health-related traits may be obscured by other variants in linkage disequilibrium (LD).[42,43] Transcriptome-wide association studies (TWAS) can be utilized to elucidate transcriptomic mechanisms underlying disease etiology by integrating GWAS with expression mapping studies. However, to date, TWAS methods have predominantly been performed in a single ancestry, typically EA, and few TWAS have focused on cognitive function or structural brain measures. Due to differences in allele frequencies LD patterns across different ancestries, genetically regulated gene expression (GReX) patterns may vary across populations of EA and AA. As a result, expression could thus impede TWAS effectiveness. Further, previous TWAS methods have not able to take advantage of recent expression quantitative trait loci (eQTL) studies conducted in different ancestries. As such, a powerful TWAS method that leverages data from different ancestries is important for identifying gene-trait associations.

In this aim, we will conduct a TWAS to identify genes associated with general cognitive function, WMH and AD, using gene expression data from both AA and EA adults. We will utilize a newly developed computational TWAS method, the Multi-ancEstry TRanscriptOme-wide analysis (METRO),[44] to leverage recent eQTL studies performed in multiple genetic ancestries (N=801 EA and N=1,032 AA individuals from GENOA) and summary statistics from large GWAS studies in EA. We will construct expression prediction models in these ancestries to capture the distinct genetic architectures underlying gene expression in each ancestry, which will provide complementary information to improve TWAS effectiveness in AA. Using METRO, we will apply a joint likelihood-based inference framework to leverage association evidence across

the EA and AA ancestries to increase TWAS power to better understand gene-trait associations in AA. This will allow us to both harness the power of using multiple ancestries as well as interrogate ancestry-dependent transcriptomic mechanisms underlying genetic associations with general cognitive function, WMH and AD.[44]

Aim 3: To conduct a transcriptome-wide association study (TWAS) using the Multi-ancEstry TRanscriptOme-wide analysis (METRO) to identify genes associated with cognitive function, white matter hyperintensity and Alzheimer's disease in older African Americans.
Hypothesis 3: *We hypothesize that a number of genes will be significantly associated with general cognitive function, WMH and/or AD, and that there will be overlapping genes and biological pathways between the three traits/diseases.*

1.3 Background

1.3.1 Preclinical dementia and the dementia continuum

The progression from normal cognitive function to dementia can last many years and is affected by multiple risk factors including age, sex, education, cardiovascular disease, socio-contextual factors (e.g., neighborhood conditions), and genetics. The pathophysiological process is thought to begin decades[7,8] prior to dementia diagnosis and is characterized by noticeable cognitive impairment and decline.[45,46] The distinction between preclinical (asymptomatic) and early clinical (symptomatic) disease is subtle, with symptoms emerging gradually over time. Individuals with preclinical dementia exhibit longitudinal decline on cognitive tests, even in the absence of clinically significant symptoms.[11,47–49] Clinical diagnosis is also difficult due to the

spectrum of symptom presentation in those with dementia. Currently, dementia is screened for using a brief assessment such as the Mini-Mental State Exam (MMSE),[50] and diagnosis requires impairment in at least two cognitive domains measured using a neuropsychological test battery. Since dementia is generally diagnosed by cognitive test performance below a specific threshold, investigating general cognitive function and age-related cognitive impairment prior to meeting the diagnostic threshold is important in understanding etiology and disparities in dementia risk which may inform interventions and therapeutics that could prevent disease progression can be developed.[51]

1.3.2 Cognitive function and brain structure

A. General cognitive function

Cognitive function refers to the action of knowing and processing information. It affects every individual throughout their life course and has the potential to influence the development of different important life outcomes.[52,53] Cognitive function has been shown to positively predict socioeconomic status,[54] educational achievement,[55] occupational status, job performance,[56] mate-choice,[57] life-expectancy[58–60] and dementia.[61] Conversely, studies have found lower cognitive performance to be strongly associated with both subsequent dementia and mortality.[62] Considering that individuals with higher measured general cognitive function tend to live longer and healthier lives, retaining high cognitive function in late adulthood is an important aspect of healthy aging.

There are socioeconomic and racial/ethnic disparities in cognitive function prior to dementia onset.[63,64] Several studies have shown that AA are at increased risk for mild cognitive impairment (MCI)[65,66] and conversion from MCI to AD, compared to NHW.[66] In cross-sectional studies of cognitive function, AA had lower cognitive test scores than NHW on various cognitive

measures across multiple cognitive domains.[67] Differences in cognitive test performance between AA and NHW may be due to methodological and sampling challenges in study design, but also due to differences in the burden of risk factors (e.g., socioeconomic status, stress, etc.) over the life course associated with increased incidence and progression of cognitive impairment. Considering that many risk factors may culminate and interact over the life course to contribute to cognitive impairment and decline, it is critical to understand the impact and interplay of such risk factors within AA populations to develop strategies to modify and mitigate dementia risk and burden.

Dementia typically results from decreased cognitive function over time. Thus, longitudinal studies that show within-individual cognitive decline over time, where participants serve as their own controls, are key in characterizing cognitive aging and its disparities. While there are consistent cross-sectional differences in dementia risk and cognitive test performance among AA and NHW, there are mixed results for cognitive decline. Some studies have shown that the rate of decline among blacks on tests of executive function is slower than in NHW.[63,67–70] Also NHW performed higher on cognitive tests but had faster rates of cognitive decline.[26] However, others found no difference in rates of cognitive decline at all.[68,71,72] Such findings in the literature may be explained by differences in cognitive reserve caused by racial disparities over the lifespan. Reserve-building opportunities, such as high educational attainment,[73,74] increased occupational complexity[75] and engagement in mentally stimulating leisure activities,[76] may slow cognitive decline through learned skills and behavioral patterns that are protective from age-related damage in the brain. These markers of cognitive reserve are also indicative of socioeconomic status, which is strongly associated with race and ethnicity.[27] In addition, other socio-contextual factors, such as the presence of racial discrimination, low healthcare utilization

and exposure to environmental neurotoxicants, have also been shown to be associated with cognitive decline.[77–79] Potential racial disparities in access to reserve-building opportunities may underlie observed racial disparities in rates of cognitive decline. However, inconsistencies across studies in associations of race/ethnicity with cognitive decline could also be explained by methodological factors such as differing sampling strategies across studies, regional variability among subgroups enrolled in specific studies, and use of different cognitive tests that vary in their sensitive to cognitive decline.

B. White matter hyperintensity

Cerebral SVD is the most common, chronic and progressive vascular disease in older adults.[80] Its changes affect arterioles, capillaries and small veins that supply white matter and deep structures of the brain with oxygen and nutrients.[80] Cerebral SVD causes one quarter of all ischemic strokes and is the most common cause of vascular cognitive impairment and VaD.[81–83] It manifests as lacunar infarction (ischemia from a perforated artery) and leukoaraiosis (diffuse ischemic changes). Leukoaraiosis is a subclinical marker of cerebrovascular disease and can be detected and measured as WMH[83] using magnetic resonance imaging (MRI) in the periventricular and deep white matter regions of the brain. Leukoaraiosis has been shown to predict ischemic stroke, cognitive decline and VaD.[81,83]

Predictors of leukoaraiosis progression include age, blood pressure, current smoking and presence of lacunar infarcts.[83] Uncontrolled hypertension is associated with ischemic damage of the brain and is thought to underlie cerebrovascular disease.[84] Leukoaraiosis is thought to be a mechanistic marker on the pathway from hypertension to clinical endpoints such as ischemic stroke and VaD. Hypertension also increases risk of developing impairments in mobility,

cognitive function and mood – pathways that are most likely mediated by the presence of WMH.[85] Progression of WMHs is associated with decline in information processing speed, general cognitive function, and MMSE scores. Studies have shown that the presence and severity of WMHs are important predictors of cognitive and functional impairment.[85]

1.3.3 Individual-level and neighborhood risk factors for cognitive impairment, cognitive decline and dementia

Risk factors related to structural and socio-contextual determinants of health may help us to better understand the disproportionate burden of cognitive impairment and dementia among AA. Approximately a third of AD cases worldwide might be attributable to modifiable risk factors—AD incidence may be reduced through improved access to education and healthcare, interventions on vascular diseases (e.g., via physical activity, smoking cessation, improved diet, etc.) and depression.[86] Educational attainment is associated with AD in NHW and AA, but the lower educational attainment among AA may be an important contributor to racial disparities in AD, according to one meta-analysis.[87] Other factors such as psychosocial stress, physical activity, and obesity have been indicated as individual-level risk factors related to cognitive impairment in AA. Altogether, AA are more likely to live in neighborhoods with social conditions (e.g., discrimination, education, SES, etc.) that may affect their stress levels, and in turn, affect their physiological regulation.[88] This may lead to higher levels of cognitive impairment or dementia.

Neighborhoods are defined as living and work environments that possess both physical and social attributes that may affect the health of their residents. Specifically, characteristics of the neighborhood environment are associated with cognitive function in older adults.[41] The

underlying mechanisms may relate to the contextual influences on personal mobility, sense of security and safety, the potential for social interactions and physical activity, access to healthy foods and green space, and exposure to pollution, crime and social deprivation. Since older adults are more likely to spend less time in motorized transportation, have less mobility, and have more time at home and/or in the neighborhood, the neighborhood may play an important role in their health and cognitive function.[89] For example, the neighborhood may play a strong role in providing social ties and stimulating recreation and social participation among older adults, which in turn may affect their psychological and cognitive health and overall well-being.[90,91] Neighborhood environments may provide stimulating activities that may delay the onset of cognitive impairment and reduce dementia pathology. Understanding how neighborhood environments impact dementia pathology may allow us to develop better interventions to prevent disease onset.

1.3.4 Role of genetic factors in Alzheimer's disease and cognitive function

Genetics have been shown to be a strong influence on cognitive function and dementia. AD has high heritability, ranging from 58-79%,[92] while episodic memory has 30-60% heritability.[93,94] Twin studies have found general cognitive function to have a heritability of more than 50%, starting from adolescence to young adulthood (ranging from 55-66%);[95–97] while SNP-based estimates are lower (20-30%). As such, there is a gap between SNP heritability estimates and twin- or family-based heritability estimates. This gap may be explained by the idea that GWAS does not capture other structural variants beyond SNPs, rare variants, poorly tagged or multiple independent variants, dominant and epistatic effects, epigenetics, and gene-environment interaction.[98] Differences in the measures of cognitive function used across studies,

as well as differences in the heritability of cognitive measures across age groups, may also contribute to differences in heritability estimates across studies.[98] There are also socio-contextual influences that change over time, which contribute to heritability estimates.[95,98] However, considering the relatively strong relationship between genetics and cognitive function, cognitive decline and dementia, understanding the body of genetic research pertaining to these outcomes will help us to better understand future research to prevent or treat preclinical dementia.

In addition to age, genetic variants in the *APOE* gene are the largest risk factor for AD in AA,[37] with one copy of the *APOE ε*4 allele increasing AD risk by 3-5 fold. [99–101] *ABCA7* is the second largest genetic risk factor for AD in AA, with genetic variants increasing AD risk by 70-80%.[37] There have been at least 75 loci associated with late-onset AD (LOAD) at genome-wide significance, in at least two EA GWAS.[31,102–112] With respect to general cognitive function, 148 genetic loci have been identified (among older EA adults), with biological pathways related to neural and cell development.[15,16] Some of the genes identified in the general cognitive function GWAS have also been associated with AD, including *APOE, TOMM40, ABCA7, ABCG1, MEF2C,* and *SLC39A1.*[15,16] Overlapping biological pathways include lipid metabolism, inflammation and immune function.[12,13]

While previous GWAS for general cognitive function and AD have shown some overlapping loci,[15,16] further studies of cognitively "resilient" individuals who live to an older age with intact cognitive function, despite the presence of AD neuropathology, have found the genetic architecture of cognitive resilience to be distinct from that of AD.[113] At this point, relatively little is known about the pathways involving genetic variants and cognitive aging in those without dementia. Thus, studying variants affect general cognitive function in those

without dementia, as well as their interactions with socio-contextual factors, may identify novel pathways for therapeutic targets.

1.3.5 Role of epigenetic and transcriptomic factors in Alzheimer's disease and cognitive function

A. Epigenetics

Epigenetics are a potentially reversible molecular link between an individual's environment and gene expression.[114] Epigenetic changes may mediate or be an effect modifier on the pathway from risk factor to disease outcome, or they may be early biomarkers of disease and thus may be used to improve early detection (and reduce misclassification).[115] DNA methylation is one of the most studied epigenetic modifications and involves the transfer of a methyl group to a C-5 position of a cytosine, most prominently in a cytosine:guanine (CpG) sequence of DNA. Depending on the genomic context, methylation may up- or down-regulate gene expression.[116] Epigenetic markers may explain individual variation in disease phenotypes and identify environmentally driven disease mechanisms, including gene-by-environment interactions.[117]

Epigenome-wide association studies (EWAS) interrogate CpG sites across the genome to evaluate the association between methylation levels and a trait of interest. Recently, an EWAS meta-analysis was performed on seven measures of cognitive function in circulating leucocytes among 6,809 healthy, older-aged adults in 11 independent cohorts, including GENOA.[118] At an epigenome-wide significance level, there was an association between cg21450381 (located in an intergenic region on chromosome 12) and global cognitive function (MMSE score), as well as between cg12507869 (located in *INPP5A*) and phonemic verbal fluency. *INPP5A* is a member of the INPP5 family of gene family that has been implicated in cerebellar degeneration in mice[119] and is associated with AD and cognitive decline in humans.[102,120] CpGs identified in the

cognitive function EWAS as suggestive, but not epigenome-wide significant, were in or near genes related to inflammation (*CCR9, PRRC2A, SOCS3*) and neurodegeneration (through the beta-amyloid precursor protein interactor, *GAPDH*), among others. To that end, there is increasing evidence that there are strong and specific changes in DNA methylation in both peripheral blood and the brain that may indicate, and/or lead to, cognitive decline and impairment prior to dementia onset.[118,121–123]

Epigenetic variation in the brain is also associated with AD.[121,122,124–126] In two studies, investigation of postmortem AD brain tissue showed epigenetic dysregulation in genes with pathways related to neuroinflammation, neurogenesis, and cognitive function.[127,128] Brain DNA methylation in five of 28 AD loci identified from GWAS (*ABCA7, SORL1, HLA-DRB5, SLC24A4,* and *BIN1*) were associated with hallmark AD pathologies, including Aβ load and tau tangle density.[128] There is also increasing evidence for AD-related alterations in DNA methylation, with specific brain regions being either hyper- or hypomethylated.[121–123] While there are still many unanswered questions in this research area, studies point to a strong but specific manner in which the epigenome is associated with AD pathogenesis in the brain.

B. Transcriptomics

Transcriptomics is the study of RNA transcripts in a cell (i.e., mRNAs, non-coding RNAs, and small RNAs) and their quantity at a specific developmental stage or with respect to a specific physiological condition.[129] Studying RNA is essential for interpreting the functional elements of the genome, such as the transcriptional structure of genes, alternative splicing patterns, post-transcriptional modifications, and gene expression levels during developmental processes and/or conditions.

TWAS characterize underlying genetically regulated mechanisms between genetic variants and health-related outcomes by aggregating genomic information into functionally relevant units that map genes to their expression.[130] To date, only a few TWAS for cognitive function have been conducted, and they have all been in relatively small samples of EA (N<700).[131,132] These studies have shown that *CCR2*[129] and *POU6F1*[130] are associated with cognitive function. Gene set enrichment conducted in the latter study[132] pointed to protein and RNA metabolism, the immune system, and infectious disease pathways.

A TWAS[133] was conducted to study transcriptomics underlying AD and detected 13 genes for AD dementia diagnosis (based on cognitive status) and pathology, including a previously identified TWAS gene *TRAPPC6A*.[133] Of the 13 genes identified, 6 were previously identified in AD GWAS, including *TOMM40*.[30,134] Pleiotropic effects suggested biological mechanisms linking AD risk genes, via *β*-amyloid and tangles, with AD dementia. This mechanism is further supported by the association between RNA expression of transcripts in *SORL1* and *ABCA7* genes with paired helical filament tau tangle density, and *BIN1* with *β*-amyloid load.[128]

Considering that these processes are involved in both normal and pathologic brain aging, and that some studies have shown gene expression in brain regions affected by AD (e.g., hippocampus) and peripheral blood among genes related to neuronal function and repair to be upregulated in cognitively impaired individuals[135,136] and then transcriptionally downregulated[137,138] in AD cases, it is hypothesized that there may be complex compensatory mechanisms preceding dementia onset.[136,138,139] TWAS may further clarify previous GWAS results and elucidate biological mechanisms underlying the gene-trait associations.

1.3.6 The importance of multi-omics and socio-contextual research in African Americans

The central dogma informs us that there is a cascade of information from the genetic code being transcribed into mRNA, which is then translated into proteins. The epigenome acts as a regulatory mechanism that mediates environmental influences on the expression of genes in a dynamic and adaptive fashion. In addition, the transcriptome consists of RNA molecules that are translated into proteins, which may undergo post-translational modifications. All of these levels interact in a complex and nonlinear way to contribute to phenotypic variations. Integrating data from different types of "omic" data (i.e., genomic, epigenomic, transcriptomic, and proteomic) would allow a more comprehensive prediction of how complex traits or phenotypes are expressed, and potentially shed light on the evolutionary mechanisms (i.e., natural selection) that shape new phenotypes.[140]

It is especially important to study risk of multifactorial disease in different populations and ethnic groups using these multi-omic layers. By understanding how genetic risk factors and molecular variation interplay with important contextual variation in a group of individuals, we may better understand the biological mechanisms underlying disease risk and onset,[140,141] as well as modifiable socio-contextual factors that contribute to the health disparities between EA and AA. Recent analysis of the GWAS catalog has revealed a lack of diversity and under-representation of non-European ancestral populations: only 19% of GWAS populations are non-European, even though over 75% of the world population live in Africa and Asia.[142] Individuals of African and Hispanic or Latin American ethnicity, specifically, contribute less to GWAS and may have a greater impact on discovery due to their higher level of genetic variation, compared to European or Asian populations. Genetic variants that affect a phenotype may vary across ethnicities, even if the underlying genetic mechanisms are the same. These differences are due to

allelic heterogeneity across different ancestral groups from mutation and linkage disequilibrium (LD), or SNP correlation, patterns that differ across ethnic groups. A thorough investigation of the relationship between these multi-omic layers and later-life cognition and brain structures (WMH) can help characterize the underlying genetic architecture of cognition in older adulthood, prior to dementia onset, in understudied AA populations. This may allow the identification of targets for intervention and treatment, especially in populations like AA that are most at risk.[35]

1.4 Study Design and Measures in The Genetic Epidemiology Network of Arteriopathy

1.4.1 Study design and source population

A. The Genetic Epidemiology Network of Arteriopathy (GENOA)

The GENOA study is a community-based longitudinal study aimed at examining the genetic effects of hypertension and related target organ damage.[143] EA and AA hypertensive sibships were recruited if at least 2 siblings were clinically diagnosed with hypertension before age 60. All other siblings were invited to participate, regardless of their hypertension status. Exclusion criteria included secondary hypertension, alcoholism or drug abuse, pregnancy, insulin-dependent diabetes mellitus, active malignancy, or serum creatinine levels >2.5mg/dL. In Phase I (1996-2001), 1,854 AA participants (Jackson, MS) and 1,583 EA participants (Rochester, MN) were recruited.[143] In Phase II (2000-2004), 1,482 AA and 1,239 EA participants were successfully followed up, and their potential target organ damage from hypertension was measured. Demographics, medical history, clinical characteristics, information on medication use, and blood samples were collected in each phase. Methylation levels were measured only in

AA participants using blood samples collected in Phases I and II. Written informed consent was obtained from all participants, and approval was granted by participating institutional review boards (University of Michigan, University of Mississippi Medical Center, and Mayo Clinic).

B. Genetics of Microangiopathic Brain Injury (GMBI) ancillary study

In an ancillary study, the Genetics of Microangiopathic Brain Injury (GMBI; 2001-2006), 1,010 AA and 967 EA Phase II GENOA participants that had a sibling willing and eligible to participate underwent a battery of established neurocognitive tests to assess several domains of cognitive function, including learning, memory, attention, concentration, and language. The goal of GMBI was to investigate susceptibility genes for ischemic brain injury. Ischemic brain damage to the subcortical and periventricular white matter (leukoaraiosis) was quantified by MRI as WMH in participants with no history of stroke or neurological disease and no implanted metal devices.

C. Exclusion criteria

Within GENOA, participants were excluded if they were diagnosed with the following: secondary hypertension, alcoholism or drug abuse, pregnancy, insulin-dependent diabetes mellitus, active malignancy, or serum creatinine levels >2.5mg/dL. For our study, to create a sample of "cognitively normal" AA adults, we excluded participants who were less than 45 years of age, had evidence of stroke, and/or preliminary evidence of dementia as indicated by a score of <24 on the MMSE.[144,145] The MMSE is a 30-question assessment of cognitive function that can be rapidly administered as a diagnostic instrument by healthcare professionals.[144] MMSE has been used to pre-screen for cognitive decline using its total score. Several studies have reported

lower performance on cognitive tests like MMSE to indicate lower cognitive functioning among individuals who go on to develop dementia.[62,146]

1.4.2 Measures of cognitive function and brain structure

A. Five neurocognitive domain measures

The following five neurocognitive domains were evaluated a year after Phase II, on average, as part of GMBI:[147,148]

1. The Weschler Adult Intelligence Scale-Revised: Digit Symbol Substitution Test (DSST) measured complex visual attention, sustained and focused concentration, response speed and visuomotor coordination. The DSST relates to the executive function of working memory.[149] In this test, participants matched symbols to numbers according to a key located at the top of the page. The DSST score comprised the number of symbols correctly matched within 90 seconds.
2. The Controlled Oral Word Association Test (COWA-FAS) tested for verbal fluency (phonetic association) and language. This required participants to generate as many words as possible that start with F, A, and S in 1 minute. The score consisted of the total number of admissible words generated.
3. The Rey Auditory Verbal Learning Test (RAVLT) measured delayed recall, relating to the cognitive functions of new learning, immediate memory span and vulnerability to interference in learning and recognition memory. Its score was determined by the number of words recalled after a 30-minute delay. Scores ranged from 0 to 15.
4. The Stroop Color-Word Test (SCWT) assessed concentration effectiveness by requiring participants to state the color of a word, rather than the word written. The score sums the

number of color words that were correctly stated in 45 seconds. Specifically, the ability to shift perceptual sets in response to novel stimuli, was tested.

5. The Trail Making Test A (TMTA) evaluated visual conceptual tracking as participants need to connect a set of 25 circles quickly and accurately. TMTA provides information on the cognitive functions of visual search, scanning, processing speed and executive functions. The natural logarithm of seconds to completion for the task was used and recoded so that higher scores indicate better cognitive function. The maximum was 240 seconds to complete.

Table 1-1. Descriptions of cognitive functions and neurocognitive domains associated with each cognitive test and measure.

Cognitive outcome	Description of cognitive measure	Cognitive domain[150]
General cognitive function	Summary measure of overall cognitive performance.	Complex measure encompassing multiple domains.
Weschler Adult Intelligence Scale-Revised: Digit Symbol Substitution Test (DSST)	Complex visual attention, sustained and focused concentration, response speed and visuomotor coordination.	Executive function, working memory.[147,149]
The Controlled Oral Word Association Test (COWA-FAS)	Verbal fluency (phonetic association) and language.	Fluency (language) and executive function.[151]
Rey Auditory Verbal Learning Test (RAVLT)	Delayed recall, relating to the cognitive functions of new learning, immediate memory span and vulnerability to interference in learning and recognition memory.	Episodic memory and fluency (verbal learning).[152]
Stroop Color-Word Test (SCWT)	Concentration effectiveness, or the ability to shift perceptual sets in response to novel stimuli (also called the Stroop Effect).[153]	Attention, processing speed, cognitive flexibility[154] and working memory.[155]
Trail Making Test A (TMTA)	Visual searching and scanning, processing speed, motor function.	Complex attention, executive functions.[156]

B. General cognitive function

General cognitive function, a measure of overall cognitive performance, can be captured by a summary measure of tests in multiple cognitive domains.[16] General cognitive function is calculated as the first unrotated principal component (FUPC) from a principal component analysis (PCA) of the five positively correlated cognitive tests taken by all participants in the full sample. This data reduction procedure loads all tests on the first unrotated principal component, and scores on this component can be calculated for each person. In GENOA, the FUPC accounts for 53% of the total variance in the neurocognitive measures and loading values of the five measures ranged from 0.52 to 0.87.

Cognitive decline is calculated as the slope of an individual's general cognitive function change (change in cognitive function over time) between the initial cognition measurement (GMBI, approximately a year after Phase II) and Phase III. Studying cognitive decline allows examination of intra-individual differences in the rate of decline in cognitive functioning.

C. White matter hyperintensity

WMH was evaluated a year after Phase II, on average, as part of GMBI. The presence of WMH in brain samples indicates leukoaraiosis, areas of ischemic damage to small vessels and surrounding areas. Brain magnetic resonance images were measured from MRI, using Signa 1.5T MRI scanners (GE Medical Systems, Waukesha, WI, USA) at Mayo Clinic.[157] WMH and total brain volume in the coronaradiata and periventricular zone were quantified from axial fluid-attenuated inversion recovery (FLAIR) images.[158] WMH in the coronaradiata and periventricular zone, as well as central gray infarcts (i.e., lacunes), were included in the leukoaraiosis measurements. Brain scans with cortical infarctions were excluded from the analyses because of

the distortion of WMH volume estimates that would be introduced in the automated segmentation algorithm. For additional details, see Smith et al.[159]

1.4.3 Genetic (SNP) data

A. Genome-wide chip data

Blood samples were genotyped using the Affymetrix® Genome-Wide Human SNP Array 6.0 or the Illumina 1M Duo. Samples and SNPs with a call rate <95%, samples with mismatch sex, and duplicate samples were removed. Genotypes were imputed using the 1000 Genomes Project Phase I integrated variant set (v.3) in NCBI build 37 (hg19) coordinates (released in March 2012). SNPs with high imputation quality will be assessed (r^2>0.7). Genetic principal components were calculated from genotyped SNPs and included in regression models to control for population stratification.

B. *Apolipoprotein E (APOE)* ε2 and ε4 alleles

To evaluate confounding and/or effect modification by *APOE* isoforms known to influence dementia risk, we measured rs7412 (to capture the *APOE* ε2 allele) and rs429359 (to capture the *APOE* ε4 allele) using a TaqMan assay and ABI Prism© Sequence Detection (Applied Biosystems, Foster City, CA) in 1544 participants. Participants were classified as having 0, 1, or 2 copies of ε2 (rs7412 T allele) and/or ε4 (rs429359 C allele).

1.4.4 DNA methylation data

Genomic data was extracted from stored peripheral blood leukocytes from 1,106 AA participants from Phase I and 304 AA participants from Phase II using the AutoGen FlexStar (AutoGen, Holliston, MA). Bisulfite conversion was performed with the EZ DNA Methylation Kit (Zymo Research, Irvine, CA), and methylation was measured using the Illumina HumanMethylationEPIC BeadChip. The raw intensity data was visualized using the shinyMethyl R package [160] to identify sex mismatches and outliers, which were removed. Samples with incomplete bisulfite conversion were identified using Qcinfo in the *Enmix* R package[161] and removed. Background correction and dye-bias normalization were performed using Noob in the *Minfi* R package.[162,163] We also checked sample identity using the 59 SNP probes on the EPIC chip, and mismatched samples were removed. Probe-type bias was adjusted using the Regression on Correlated Probes (RCP) method.[164] Probes with detection p-value $<10^{-16}$ were considered successfully detected, and probes and samples with detection rate<10% were removed.[165] After quality control, a total of 1,396 samples (N=1,100 from Phase I and N=294 from Phase II) and 857,121 CpG sites were available for analyses. For this analysis, all methylation data was from Phase I samples.

We used Illumina annotation to identify genes near each CpG site using the UCSC database and characterize each CpG site as being in a gene promoter, enhancer, DNAse I hypersensitive site (DHS), CpG Island (CGI), and/or CGI flanking shore/shelf.[166] A CpG site was considered to be in a promoter region if it was 0-1500 bases upstream of a transcriptional start site. A CpG site was assigned to CGI flanking shore/shelf if it was located within 4kb of a CGI. White blood cell proportions for CD8 T lymphocytes, CD4 T lymphocytes, natural killer cells, B cells, monocytes, and granulocytes were estimated using the Houseman method.[167] For each CpG site prior to analysis, the methylation beta value[168,169] was multiplied by 100 to

approximate the percent methylation at that site. Methylation beta values were pre-adjusted for batch effects (sample plate, row, and column) and white blood cell proportions using linear mixed modeling, and the resulting residuals were added to the mean values.

1.4.5 Gene expression data

Gene expression levels in transformed beta-lymphocyte cell lines from blood samples taken primarily at GENOA Phase II were measured using the Affymetrix Human Transcriptome Array 2.0. The Affymetrix Expression Console was used for quality control, and all array images passed visual inspection. Affymetrix Power Tool software was used to process raw intensity data.[170] We normalized Affymetrix CEL files using the Robust Multichip Average (RMA) algorithm, including background correction, quantile normalization, log_2-transformation and probe set summarization.[171] Linearity was also maintained using GC correction (GCCN), signal space transformation (SST), and gain lock (value=0.75). We used the Brainarray custom CDF[172] version 19 to map the probes to genes, specifically removing probes with non-unique matching cDNA/EST sequences that can be assigned to more than one gene cluster. As a result, the gene expression data processed through the custom CDF is expected to be free of mappability issues; however, alignment bias may still exist due to genetic variation, errors in the reference genome, and other complications.[173] After mapping, Combat was used to remove batch effects.[174]

1.5 Summary

This body of work will contribute to a better understanding of the risk factors that impact cognition among older AA adults and lend insight into how the interactions among multi-omic,

biological, and socio-contextual risk factors contribute to preclinical dementia in this population. Several genetic, epigenetic, medical and lifestyle factors are associated with dementia; however, the research has been primarily in overwhelmingly white populations. Focusing on data from one population and applying it to other populations (especially marginalized populations such as AA, individuals of low socioeconomic status, rural Americans, sexual and gender minorities, other racial and ethnic minorities, immigrants, and people with disabilities) is problematic because this research could lead to false conclusions.

We currently know much less about the social, biological, and multi-omic determinants of health in AA individuals. To better treat and prevent dementia and other diseases, we need to research the drivers of dementia in AA. It is worth noting that in addition to the multi-omics and socio-contextual factors that we study in this project, AD disparities for this population in particular have roots in structural and social determinants.[175] Considering that cognition is not only connected to dementia but also to healthcare utilization and quality and mortality, makes it an important focus of research in AA. By combining high-throughput "omics" technologies (e.g., genomics, transcriptomics, methylomics), and examining them within socio-contextual environments, we seek to provide deeper insight into the molecular features of cognition and dementia in this population.

1.6 References

1. Arvanitakis, Z., Shah, R.C., and Bennett, D.A. (2019). Diagnosis and Management of Dementia: Review. JAMA *322*, 1589–1599. 10.1001/jama.2019.4782.

2. World Health Organization (2018). Towards a dementia plan: a WHO guide (World Health Organization).

3. Dening, T., and Sandilyan, M.B. (2015). Dementia: definitions and types. Nursing Standard (2014+) *29*, 37.

4. Anor, C.J., O'Connor, S., Saund, A., Tang-Wai, D.F., Keren, R., and Tartaglia, M.C. (2017). Neuropsychiatric Symptoms in Alzheimer Disease, Vascular Dementia, and Mixed Dementia. Neurodegenerative Diseases *17*, 127–134. 10.1159/000455127.

5. O'Brien, J.T., and Thomas, A. (2015). Vascular dementia. The Lancet *386*, 1698–1706. 10.1016/S0140-6736(15)00463-8.

6. Iadecola, C. (2013). The pathobiology of vascular dementia. Neuron *80*, 844–866. 10.1016/j.neuron.2013.10.008.

7. Hof, P.R., Glannakopoulos, P., and Bouras, C. (1996). The neuropathological changes associated with normal brain aging. Histol Histopathol *11*, 1075–1088.

8. Perl, D.P. (2010). Neuropathology of Alzheimer's Disease. Mt Sinai J Med *77*, 32–42. 10.1002/msj.20157.

9. Cohuet, G., and Struijker-Boudier, H. (2006). Mechanisms of target organ damage caused by hypertension: Therapeutic potential. Pharmacology & Therapeutics *111*, 81–98. 10.1016/j.pharmthera.2005.09.002.

10. Hubbard, B.M., Fenton, G.W., and Anderson, J.M. (1990). A quantitative histological study of early clinical and preclinical Alzheimer's disease. Neuropathol Appl Neurobiol *16*, 111–121. 10.1111/j.1365-2990.1990.tb00940.x.

11. Sperling, R.A., Aisen, P.S., Beckett, L.A., Bennett, D.A., Craft, S., Fagan, A.M., Iwatsubo, T., Jack, C.R., Kaye, J., Montine, T.J., et al. (2011). Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement *7*, 280–292. 10.1016/j.jalz.2011.03.003.

12. Kunkle, B.W., Grenier-Boley, B., Sims, R., Bis, J.C., Damotte, V., Naj, A.C., Boland, A., Vronskaya, M., van der Lee, S.J., Amlie-Wolf, A., et al. (2019). Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. Nat Genet *51*, 414–430. 10.1038/s41588-019-0358-2.

13. Bagyinszky, E., Giau, V.V., and An, S.A. (2020). Transcriptomics in Alzheimer's Disease: Aspects and Challenges. International Journal of Molecular Sciences *21*, 3517. 10.3390/ijms21103517.

14. Deary, I.J., Johnson, W., and Houlihan, L.M. (2009). Genetic foundations of human intelligence. Hum Genet *126*, 215–232. 10.1007/s00439-009-0655-4.

15. Davies, G., Lam, M., Harris, S.E., Trampush, J.W., Luciano, M., Hill, W.D., Hagenaars, S.P., Ritchie, S.J., Marioni, R.E., Fawns-Ritchie, C., et al. (2018). Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. Nat Commun *9*, 2098. 10.1038/s41467-018-04362-x.

16. Davies, G., Armstrong, N., Bis, J.C., Bressler, J., Chouraki, V., Giddaluru, S., Hofer, E., Ibrahim-Verbaas, C.A., Kirin, M., Lahti, J., et al. (2015). Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N=53 949). Mol Psychiatry *20*, 183–192. 10.1038/mp.2014.188.

17. Clark, C., Dayon, L., Masoodi, M., Bowman, G.L., and Popp, J. (2021). An integrative multi-omics approach reveals new central nervous system pathway alterations in Alzheimer's disease. Alz Res Therapy *13*, 1–19. 10.1186/s13195-021-00814-7.

18. Karczewski, K.J., and Snyder, M.P. (2018). Integrative omics for health and disease. Nat Rev Genet *19*, 299–310. 10.1038/nrg.2018.4.

19. McCall, M.K., Stanfill, A.G., Skrovanek, E., Pforr, J.R., Wesmiller, S.W., and Conley, Y.P. (2018). Symptom Science: Omics Supports Common Biological Underpinnings Across Symptoms. Biological Research For Nursing *20*, 183–191. 10.1177/1099800417751069.

20. Matthews, K.A., Xu, W., Gaglioti, A.H., Holt, J.B., Croft, J.B., Mack, D., and McGuire, L.C. (2019). Racial and ethnic estimates of Alzheimer's disease and related dementias in the United States (2015–2060) in adults aged ≥65 years. Alzheimer's & Dementia *15*, 17–24. 10.1016/j.jalz.2018.06.3063.

21. Mayeda, E.R., Glymour, M.M., Quesenberry, C.P., and Whitmer, R.A. (2016). Inequalities in dementia incidence between six racial and ethnic groups over 14 years. Alzheimers Dement *12*, 216–224. 10.1016/j.jalz.2015.12.007.

22. Steenland, K., Goldstein, F.C., Levey, A., and Wharton, W. (2016). A Meta-Analysis of Alzheimer's Disease Incidence and Prevalence Comparing African-Americans and Caucasians. Journal of Alzheimer's Disease *50*, 71–76. 10.3233/JAD-150778.

23. Barnes, L.L., and Bennett, D.A. (2014). Alzheimer's disease in African Americans: risk factors and challenges for the future. Health Aff (Millwood) *33*, 580–586. 10.1377/hlthaff.2013.1353.

24. Brewster, P., Barnes, L., Haan, M., Johnson, J.K., Manly, J.J., Nápoles, A.M., Whitmer, R.A., Carvajal-Carmona, L., Early, D., Farias, S., et al. (2019). Progress and future challenges in aging and diversity research in the United States. Alzheimers Dement *15*, 995–1003. 10.1016/j.jalz.2018.07.221.

25. Rajan, K.B., Weuve, J., Barnes, L.L., Wilson, R.S., and Evans, D.A. (2019). Prevalence and incidence of clinically diagnosed Alzheimer's disease dementia from 1994 to 2012 in a population study. Alzheimers Dement *15*, 1–7. 10.1016/j.jalz.2018.07.216.

26. Weuve, J., Barnes, L.L., de Leon, C.F.M., Rajan, K.B., Beck, T., Aggarwal, N.T., Hebert, L.E., Bennett, D.A., Wilson, R.S., and Evans, D.A. (2018). Cognitive Aging in Black and White Americans: Cognition, Cognitive Decline, and Incidence of Alzheimer Disease Dementia. Epidemiology *29*, 151–159. 10.1097/EDE.0000000000000747.

27. Glymour, M.M., and Manly, J.J. (2008). Lifecourse social conditions and racial and ethnic patterns of cognitive aging. Neuropsychol Rev *18*, 223–254. 10.1007/s11065-008-9064-z.

28. Bell, M.L., and Ebisu, K. (2012). Environmental Inequality in Exposures to Airborne Particulate Matter Components in the United States. Environ Health Perspect *120*, 1699–1704. 10.1289/ehp.1205201.

29. Cullen, N.C., Leuzy, A., Janelidze, S., Palmqvist, S., Svenningsson, A.L., Stomrud, E., Dage, J.L., Mattsson-Carlgren, N., and Hansson, O. (2021). Plasma biomarkers of Alzheimer's disease improve prediction of cognitive decline in cognitively unimpaired elderly populations. Nat Commun *12*, 3555. 10.1038/s41467-021-23746-0.

30. Jansen, I.E., Savage, J.E., Watanabe, K., Bryois, J., Williams, D.M., Steinberg, S., Sealock, J., Karlsson, I.K., Hägg, S., Athanasiu, L., et al. (2019). Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. Nat Genet *51*, 404–413. 10.1038/s41588-018-0311-9.

31. Lambert, J.-C., Heath, S., Even, G., Campion, D., Sleegers, K., Hiltunen, M., Combarros, O., Zelenika, D., Bullido, M.J., Tavernier, B., et al. (2009). Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. Nature Genetics *41*, 1094–1099. 10.1038/ng.439.

32. Wightman, D.P., Jansen, I.E., Savage, J.E., Shadrin, A.A., Bahrami, S., Holland, D., Rongve, A., Børte, S., Winsvold, B.S., Drange, O.K., et al. (2021). A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. Nat Genet *53*, 1276–1282. 10.1038/s41588-021-00921-z.

33. Lardenoije, R., Iatrou, A., Kenis, G., Kompotis, K., Steinbusch, H.W.M., Mastroeni, D., Coleman, P., Lemere, C.A., Hof, P.R., van den Hove, D.L.A., et al. (2015). The epigenetics of aging and neurodegeneration. Prog. Neurobiol. *131*, 21–64. 10.1016/j.pneurobio.2015.05.002.

34. Noori, A., Mezlini, A.M., Hyman, B.T., Serrano-Pozo, A., and Das, S. (2021). Systematic review and meta-analysis of human transcriptomics reveals neuroinflammation, deficient energy metabolism, and proteostasis failure across neurodegeneration. Neurobiol. Dis. *149*, 105225. 10.1016/j.nbd.2020.105225.

35. Cacabelos, R., and Torrellas, C. (2015). Epigenetics of Aging and Alzheimer's Disease: Implications for Pharmacogenomics and Drug Response. International Journal of Molecular Sciences *16*, 30483–30543. 10.3390/ijms161226236.

36. Daniels, P.R., Kardia, S.L.R., Hanis, C.L., Brown, C.A., Hutchinson, R., Boerwinkle, E., Turner, S.T., and Genetic Epidemiology Network of Arteriopathy study (2004). Familial aggregation of hypertension treatment and control in the Genetic Epidemiology Network of Arteriopathy (GENOA) study. Am. J. Med. *116*, 676–681. 10.1016/j.amjmed.2003.12.032.

37. Reitz, C., Jun, G., Naj, A., and Rajbhandary, R. (2013). Variants in the ATP-Binding Cassette Transporter (ABCA7), Apolipoprotein E ϵ4,and the Risk of Late-Onset Alzheimer Disease in African Americans. JAMA *309*, 1483–1492. 10.1001/jama.2013.2973.

38. De Roeck, A., Van Broeckhoven, C., and Sleegers, K. (2019). The role of ABCA7 in Alzheimer's disease: evidence from genomics, transcriptomics and methylomics. Acta Neuropathol *138*, 201–220. 10.1007/s00401-019-01994-1.

39. Nelson, L., and Tabet, N. (2015). Slowing the progression of Alzheimer's disease; what works? Ageing Research Reviews *23*, 193–209. 10.1016/j.arr.2015.07.002.

40. Cassarino, M., and Setti, A. (2015). Environment as 'Brain Training': A review of geographical and physical environmental influences on cognitive ageing. Ageing Research Reviews *23*, 167–182. 10.1016/j.arr.2015.06.003.

41. Besser, L.M., McDonald, N.C., Song, Y., Kukull, W.A., and Rodriguez, D.A. (2017). Neighborhood Environment and Cognition in Older Adults: A Systematic Review. American Journal of Preventive Medicine *53*, 241–251. 10.1016/j.amepre.2017.02.013.

42. Wainberg, M., Sinnott-Armstrong, N., Mancuso, N., Barbeira, A.N., Knowles, D.A., Golan, D., Ermel, R., Ruusalepp, A., Quertermous, T., Hao, K., et al. (2019). Opportunities and challenges for transcriptome-wide association studies. Nat Genet *51*, 592–599. 10.1038/s41588-019-0385-z.

43. Gallagher, M.D., and Chen-Plotkin, A.S. (2018). The Post-GWAS Era: From Association to Function. The American Journal of Human Genetics *102*, 717–730. 10.1016/j.ajhg.2018.04.002.

44. Li, Z., Zhao, W., Shang, L., Mosley, T.H., Kardia, S.L.R., Smith, J.A., and Zhou, X. (2022). METRO: Multi-ancestry transcriptome-wide association studies for powerful gene-trait association detection. The American Journal of Human Genetics *109*, 783–801. 10.1016/j.ajhg.2022.03.003.

45. James, B.D., and Bennett, D.A. (2019). Causes and Patterns of Dementia: An Update in the Era of Redefining Alzheimer's Disease. Annu. Rev. Public Health *40*, 65–84. 10.1146/annurev-publhealth-040218-043758.

46. Jack, C.R., Bennett, D.A., Blennow, K., Carrillo, M.C., Dunn, B., Haeberlein, S.B., Holtzman, D.M., Jagust, W., Jessen, F., Karlawish, J., et al. (2018). NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. Alzheimer's & Dementia *14*, 535–562. 10.1016/j.jalz.2018.02.018.

47. Johnson, D.K., Storandt, M., Morris, J.C., and Galvin, J.E. (2009). Longitudinal Study of the Transition From Healthy Aging to Alzheimer Disease. Archives of Neurology *66*, 1254–1259. 10.1001/archneurol.2009.158.

48. Donohue, M.C., Sperling, R.A., Salmon, D.P., Rentz, D.M., Raman, R., Thomas, R.G., Weiner, M., Aisen, P.S., and for the Australian Imaging, B., and Lifestyle Flagship Study of

Ageing; the Alzheimer's Disease Neuroimaging Initiative; and the Alzheimer's Disease Cooperative Study (2014). The Preclinical Alzheimer Cognitive Composite: Measuring Amyloid-Related Decline. JAMA Neurology *71*, 961–970. 10.1001/jamaneurol.2014.803.

49. Jessen, F. (2014). Subjective and objective cognitive decline at the pre-dementia stage of Alzheimer's disease. Eur Arch Psychiatry Clin Neurosci *264*, 3–7. 10.1007/s00406-014-0539-z.

50. Arvanitakis, Z., Shah, R.C., and Bennett, D.A. (2019). Diagnosis and Management of Dementia: Review. JAMA *322*, 1589–1599. 10.1001/jama.2019.4782.

51. Dubois, B., Hampel, H., Feldman, H.H., Scheltens, P., Aisen, P., Andrieu, S., Bakardjian, H., Benali, H., Bertram, L., Blennow, K., et al. (2016). Preclinical Alzheimer's disease: Definition, natural history, and diagnostic criteria. Alzheimer's & Dementia *12*, 292–323. 10.1016/j.jalz.2016.02.002.

52. Brand, C., and Modgil, C. (1987). The importance of human intelligence. Arthur Jensen: Consensus and Controversy.

53. Deary, I.J. (2012). Intelligence. Annu. Rev. Psychol. *63*, 453–482. 10.1146/annurev-psych-120710-100353.

54. Cervilla, J., Prince, M., Joels, S., Lovestone, S., and Mann, A. (2004). Premorbid cognitive testing predicts the onset of dementia and Alzheimer's disease better than and independently of APOE genotype. Journal of Neurology, Neurosurgery & Psychiatry *75*, 1100–1106.

55. Ivanovic, D.M., María del Pilar, N.R., Pérez, H.T., Alvear, J.A., Almagia, A.F., Toro, T.D., Urrutia, M.S.C., Cruz, A.L., and Ivanovic, R.M. (2009). Impact of nutritional status at the onset of elementary school on academic aptitude test achievement at the end of high school in a multicausal approach. British Journal of Nutrition *102*, 142–154.

56. Gottfredson, L.S. (2002). Where and why g matters: Not a mystery. Human performance *15*, 25–46.

57. Buss, D.M. (1989). Sex differences in human mate preferences: Evolutionary hypotheses tested in 37 cultures. Behavioral and brain sciences *12*, 1–14.

58. Whalley, L.J., and Deary, I.J. (2001). Longitudinal cohort study of childhood IQ and survival up to age 76. Bmj *322*, 819.

59. Batty, G.D., Mortensen, L.H., Gale, C.R., Shipley, M.J., Roberts, B.A., and Deary, I.J. (2009). IQ in late adolescence/early adulthood, risk factors in middle age, and later cancer mortality in men: the Vietnam Experience Study. Psycho-Oncology: Journal of the Psychological, Social and Behavioral Dimensions of Cancer *18*, 1122–1126.

60. Batty, G.D., Shipley, M.J., Dundas, R., Macintyre, S., Der, G., Mortensen, L.H., and Deary, I.J. (2009). Does IQ explain socio-economic differentials in total and cardiovascular disease

mortality? Comparison with the explanatory power of traditional cardiovascular disease risk factors in the Vietnam Experience Study. European heart journal *30*, 1903–1909.

61. Liu-Seifert, H., Siemers, E., Price, K., Han, B., Selzler, K.J., Henley, D., Sundell, K., Aisen, P., Cummings, J., Raskin, J., et al. (2015). Cognitive Impairment Precedes and Predicts Functional Impairment in Mild Alzheimer's Disease. Journal of Alzheimer's Disease *47*, 205–214. 10.3233/JAD-142508.

62. Johansson, B., and Zarit, S.H. (1997). Early Cognitive Markers of the Incidence of Dementia and Mortality: A Longitudinal Population-Based Study of the Oldest Old. International Journal of Geriatric Psychiatry *12*, 53–59. 10.1002/(SICI)1099-1166(199701)12:1<53::AID-GPS507>3.0.CO;2-M.

63. Karlamangla, A.S., Miller-Martinez, D., Aneshensel, C.S., Seeman, T.E., Wight, R.G., and Chodosh, J. (2009). Trajectories of Cognitive Function in Late Life in the United States: Demographic and Socioeconomic Predictors. Am J Epidemiol *170*, 331–342. 10.1093/aje/kwp154.

64. Yaffe, K., Falvey, C., Harris, T.B., Newman, A., Satterfield, S., Koster, A., Ayonayon, H., and Simonsick, E. (2013). Effect of socioeconomic disparities on incidence of dementia among biracial older adults: prospective study. BMJ *347*, f7051. 10.1136/bmj.f7051.

65. Katz, M.J., Lipton, R.B., Hall, C.B., Zimmerman, M.E., Sanders, A.E., Verghese, J., Dickson, D.W., and Derby, C.A. (2012). 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 Disease & Associated Disorders *26*, 335–343. 10.1097/WAD.0b013e31823dbcfc.

66. Manly, J.J., Tang, M.-X., Schupf, N., Stern, Y., Vonsattel, J.-P.G., and Mayeux, R. (2008). Frequency and course of mild cognitive impairment in a multiethnic community. Annals of Neurology *63*, 494–506. 10.1002/ana.21326.

67. Gross, A.L., Mungas, D.M., Crane, P.K., Gibbons, L.E., MacKay-Brandt, A., Manly, J.J., Mukherjee, S., Romero, H., Sachs, B., Thomas, M., et al. (2015). Effects of education and race on cognitive decline: An integrative analysis of generalizability versus study-specific results. Psychol Aging *30*, 863–880. 10.1037/pag0000032.

68. Early, D.R., Widaman, K.F., Harvey, D., Beckett, L., Park, L.Q., Farias, S.T., Reed, B.R., DeCarli, C., and Mungas, D. (2013). Demographic Predictors of Cognitive Change in Ethnically Diverse Older Persons. Psychol Aging *28*, 633–645. 10.1037/a0031645.

69. Manly, J.J., Schupf, N., Tang, M.-X., and Stern, Y. (2005). Cognitive decline and literacy among ethnically diverse elders. J Geriatr Psychiatry Neurol *18*, 213–217. 10.1177/0891988705281868.

70. Wilson, R.S., Capuano, A.W., Sytsma, J., Bennett, D.A., and Barnes, L.L. (2015). Cognitive Aging in Older Black and White Persons. Psychol Aging *30*, 279–285. 10.1037/pag0000024.

71. Castora-Binkley, M., Peronto, C.L., Edwards, J.D., and Small, B.J. (2015). A Longitudinal Analysis of the Influence of Race on Cognitive Performance. The Journals of Gerontology: Series B *70*, 512–518. 10.1093/geronb/gbt112.

72. Masel, M.C., and Peek, M.K. (2009). Ethnic differences in cognitive function over time. Ann Epidemiol *19*, 778–783. 10.1016/j.annepidem.2009.06.008.

73. Bennett, D.A., Wilson, R.S., Schneider, J.A., Evans, D.A., Leon, C.F.M. de, Arnold, S.E., Barnes, L.L., and Bienias, J.L. (2003). Education modifies the relation of AD pathology to level of cognitive function in older persons. Neurology *60*, 1909–1915. 10.1212/01.WNL.0000069923.64550.9F.

74. Almeida, R.P., Schultz, S.A., Austin, B.P., Boots, E.A., Dowling, N.M., Gleason, C.E., Bendlin, B.B., Sager, M.A., Hermann, B.P., Zetterberg, H., et al. (2015). Effect of Cognitive Reserve on Age-Related Changes in Cerebrospinal Fluid Biomarkers of Alzheimer Disease. JAMA Neurology *72*, 699–706. 10.1001/jamaneurol.2015.0098.

75. Boots, E.A., Schultz, S.A., Almeida, R.P., Oh, J.M., Koscik, R.L., Dowling, M.N., Gallagher, C.L., Carlsson, C.M., Rowley, H.A., Bendlin, B.B., et al. (2015). Occupational Complexity and Cognitive Reserve in a Middle-Aged Cohort at Risk for Alzheimer's Disease. Archives of Clinical Neuropsychology *30*, 634–642. 10.1093/arclin/acv041.

76. Schultz, S.A., Larson, J., Oh, J., Koscik, R., Dowling, M.N., Gallagher, C.L., Carlsson, C.M., Rowley, H.A., Bendlin, B.B., Asthana, S., et al. (2015). Participation in cognitively-stimulating activities is associated with brain structure and cognitive function in preclinical Alzheimer's disease. Brain Imaging and Behavior *9*, 729–736. 10.1007/s11682-014-9329-5.

77. Crowe, M., Sartori, A., Clay, O.J., Wadley, V.G., Andel, R., Wang, H.-X., Sawyer, P., and Allman, R.M. (2010). Diabetes and Cognitive Decline: Investigating the Potential Influence of Factors Related to Health Disparities. J Aging Health *22*, 292–306. 10.1177/0898264309357445.

78. Turner, A.D., James, B.D., Capuano, A.W., Aggarwal, N.T., and Barnes, L.L. (2017). Perceived Stress and Cognitive Decline in Different Cognitive Domains in a Cohort of Older African Americans. The American Journal of Geriatric Psychiatry *25*, 25–34. 10.1016/j.jagp.2016.10.003.

79. Shih, R.A., Glass, T.A., Bandeen-Roche, K., Carlson, M.C., Bolla, K.I., Todd, A.C., and Schwartz, B.S. (2006). Environmental lead exposure and cognitive function in community-dwelling older adults. Neurology *67*, 1556–1562. 10.1212/01.wnl.0000239836.26142.c5.

80. Chojdak-Łukasiewicz, J., Dziadkowiak, E., Zimny, A., and Paradowski, B. (2021). Cerebral small vessel disease: A review. Adv Clin Exp Med *30*, 349–356. 10.17219/acem/131216.

81. Markus, H.S., Hunt, B., Palmer, K., Enzinger, C., Schmidt, H., and Schmidt, R. (2005). Markers of Endothelial and Hemostatic Activation and Progression of Cerebral White Matter Hyperintensities. Stroke *36*, 1410–1414. 10.1161/01.STR.0000169924.60783.d4.

82. Williams, O.A., Zeestraten, E.A., Benjamin, P., Lambert, C., Lawrence, A.J., Mackinnon, A.D., Morris, R.G., Markus, H.S., Barrick, T.R., and Charlton, R.A. (2019). Predicting Dementia in Cerebral Small Vessel Disease Using an Automatic Diffusion Tensor Image Segmentation Technique. Stroke *50*, 2775–2782. 10.1161/STROKEAHA.119.025843.

83. Prins, N.D., and Scheltens, P. (2015). White matter hyperintensities, cognitive impairment and dementia: an update. Nat Rev Neurol *11*, 157–165. 10.1038/nrneurol.2015.10.

84. Turner, S.T., and Boerwinkle, E. (2000). Genetics of hypertension, target-organ complications, and response to therapy. Circulation *102*, IV40-45. 10.1161/01.cir.102.suppl_4.iv-40.

85. Hajjar, I., Quach, L., Yang, F., Chaves, P.H.M., Newman, A.B., Mukamal, K., Longstreth, W., Inzitari, M., and Lipsitz, L.A. (2011). Hypertension, White Matter Hyperintensities, and Concurrent Impairments in Mobility, Cognition, and Mood. Circulation *123*, 858–865. 10.1161/CIRCULATIONAHA.110.978114.

86. Norton, S., Matthews, F.E., Barnes, D.E., Yaffe, K., and Brayne, C. (2014). Potential for primary prevention of Alzheimer's disease: an analysis of population-based data. Lancet Neurol *13*, 788–794. 10.1016/S1474-4422(14)70136-X.

87. Peterson, R.L., Fain, M.J., A. Butler, E., Ehiri, J.E., and Carvajal, S.C. (2020). The role of social and behavioral risk factors in explaining racial disparities in age-related cognitive impairment: a structured narrative review. Aging, Neuropsychology, and Cognition *27*, 173–196. 10.1080/13825585.2019.1598539.

88. Forrester, S.N., Gallo, J.J., Whitfield, K.E., and Thorpe, R.J. (2019). A Framework of Minority Stress: From Physiological Manifestations to Cognitive Outcomes. Gerontologist *59*, 1017–1023. 10.1093/geront/gny104.

89. Levasseur, M., Généreux, M., Bruneau, J.-F., Vanasse, A., Chabot, É., Beaulac, C., and Bédard, M.-M. (2015). Importance of proximity to resources, social support, transportation and neighborhood security for mobility and social participation in older adults: results from a scoping study. BMC Public Health *15*, 1–19. 10.1186/s12889-015-1824-0.

90. Aneshensel, C.S., Ko, M.J., Chodosh, J., and Wight, R.G. (2011). The Urban Neighborhood and Cognitive Functioning in Late Middle Age. J Health Soc Behav *52*, 163–179. 10.1177/0022146510393974.

91. James, B.D. Social engagement and cognitive decline in older adults: Pathways and neighborhood context.

92. Gatz, M., Reynolds, C.A., Fratiglioni, L., Johansson, B., Mortimer, J.A., Berg, S., Fiske, A., and Pedersen, N.L. (2006). Role of Genes and Environments for Explaining Alzheimer Disease. Archives of General Psychiatry *63*, 168–174. 10.1001/archpsyc.63.2.168.

93. Papassotiropoulos, A., and de Quervain, D.J.-F. (2011). Genetics of human episodic memory: dealing with complexity. Trends in Cognitive Sciences *15*, 381–387. 10.1016/j.tics.2011.07.005.

94. Migliore, L., and Coppedè, F. (2009). Genetics, environmental factors and the emerging role of epigenetics in neurodegenerative diseases. Mutat Res *667*, 82–97. 10.1016/j.mrfmmm.2008.10.011.

95. Plomin, R., and Deary, I.J. (2015). Genetics and intelligence differences: five special findings. Mol Psychiatry *20*, 98–108. 10.1038/mp.2014.105.

96. Tucker-Drob, E.M., Briley, D.A., and Harden, K.P. (2013). Genetic and environmental influences on cognition across development and context. Current directions in psychological science *22*, 349–355.

97. Haworth, C.M., Wright, M.J., Luciano, M., Martin, N.G., de Geus, E.J., van Beijsterveldt, C.E., Bartels, M., Posthuma, D., Boomsma, D.I., and Davis, O.S. (2010). The heritability of general cognitive ability increases linearly from childhood to young adulthood. Molecular psychiatry *15*, 1112–1120.

98. Reynolds, C.A., and Finkel, D. (2015). A Meta-analysis of Heritability of Cognitive Aging: Minding the “Missing Heritability” Gap. Neuropsychol Rev *25*, 97–112. 10.1007/s11065-015-9280-2.

99. Strittmatter, W.J., Saunders, A.M., Schmechel, D., Pericak-Vance, M., Enghild, J., Salvesen, G.S., and Roses, A.D. (1993). Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. Proceedings of the National Academy of Sciences *90*, 1977–1981. 10.1073/pnas.90.5.1977.

100. Strittmatter, W.J., and Roses, A.D. (1995). Apolipoprotein E and Alzheimer disease. Proceedings of the National Academy of Sciences *92*, 4725–4727. 10.1073/pnas.92.11.4725.

101. Corder, E.H., Saunders, A.M., Strittmatter, W.J., Schmechel, D.E., Gaskell, P.C., Small, G.W., Roses, A.D., Haines, J.L., and Pericak-Vance, M.A. (1993). Gene Dose of Apolipoprotein E Type 4 Allele and the Risk of Alzheimer’s Disease in Late Onset Families. Science *261*, 921–923. 10.1126/science.8346443.

102. Lambert, J.-C., Ibrahim-Verbaas, C.A., Harold, D., Naj, A.C., Sims, R., Bellenguez, C., Jun, G., DeStefano, A.L., Bis, J.C., Beecham, G.W., et al. (2013). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer’s disease. Nat Genet *45*, 1452–1458. 10.1038/ng.2802.

103. Harold, D., Abraham, R., Hollingworth, P., Sims, R., Gerrish, A., Hamshere, M.L., Pahwa, J.S., Moskvina, V., Dowzell, K., Williams, A., et al. (2009). Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer’s disease. Nat Genet *41*, 1088–1093. 10.1038/ng.440.

104. Seshadri, S., Fitzpatrick, A.L., Ikram, M.A., DeStefano, A.L., Gudnason, V., Boada, M., Bis, J.C., Smith, A.V., Carrasquillo, M.M., Lambert, J.C., et al. (2010). Genome-wide Analysis of Genetic Loci Associated With Alzheimer Disease. JAMA *303*, 1832–1840. 10.1001/jama.2010.574.

105. Naj, A.C., Jun, G., Beecham, G.W., Wang, L.-S., Vardarajan, B.N., Buros, J., Gallins, P.J., Buxbaum, J.D., Jarvik, G.P., Crane, P.K., et al. (2011). Common variants in MS4A4/MS4A6E, CD2uAP, CD33, and EPHA1 are associated with late-onset Alzheimer's disease. Nat Genet *43*, 436–441. 10.1038/ng.801.

106. Hollingworth, P., Harold, D., Sims, R., Gerrish, A., Lambert, J.-C., Carrasquillo, M.M., Abraham, R., Hamshere, M.L., Pahwa, J.S., Moskvina, V., et al. (2011). Common variants in ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. Nat Genet *43*, 429–435. 10.1038/ng.803.

107. Lee, J.H., Cheng, R., Barral, S., Reitz, C., Medrano, M., Lantigua, R., Jiménez-Velazquez, I.Z., Rogaeva, E., St. George-Hyslop, P.H., and Mayeux, R. (2011). Identification of Novel Loci for Alzheimer Disease and Replication of CLU, PICALM, and BIN1 in Caribbean Hispanic Individuals. Archives of Neurology *68*, 320–328. 10.1001/archneurol.2010.292.

108. Miyashita, A., Koike, A., Jun, G., Wang, L.-S., Takahashi, S., Matsubara, E., Kawarabayashi, T., Shoji, M., Tomita, N., Arai, H., et al. (2013). SORL1 Is Genetically Associated with Late-Onset Alzheimer's Disease in Japanese, Koreans and Caucasians. PLOS ONE *8*, e58618. 10.1371/journal.pone.0058618.

109. Bertram, L., Lange, C., Mullin, K., Parkinson, M., Hsiao, M., Hogan, M.F., Schjeide, B.M.M., Hooli, B., DiVito, J., Ionita, I., et al. (2008). Genome-wide Association Analysis Reveals Putative Alzheimer's Disease Susceptibility Loci in Addition to APOE. The American Journal of Human Genetics *83*, 623–632. 10.1016/j.ajhg.2008.10.008.

110. Jun, G., Asai, H., Zeldich, E., Drapeau, E., Chen, C., Chung, J., Park, J.-H., Kim, S., Haroutunian, V., Foroud, T., et al. (2014). PLXNA4 is associated with Alzheimer disease and modulates tau phosphorylation. Annals of Neurology *76*, 379–392. 10.1002/ana.24219.

111. Wijsman, E.M., Pankratz, N.D., Choi, Y., Rothstein, J.H., Faber, K.M., Cheng, R., Lee, J.H., Bird, T.D., Bennett, D.A., Diaz-Arrastia, R., et al. (2011). Genome-Wide Association of Familial Late-Onset Alzheimer's Disease Replicates BIN1 and CLU and Nominates CUGBP2 in Interaction with APOE. PLOS Genetics *7*, e1001308. 10.1371/journal.pgen.1001308.

112. Bellenguez, C., Küçükali, F., Jansen, I.E., Kleineidam, L., Moreno-Grau, S., Amin, N., Naj, A.C., Campos-Martin, R., Grenier-Boley, B., Andrade, V., et al. (2022). New insights into the genetic etiology of Alzheimer's disease and related dementias. Nat Genet, 1–25. 10.1038/s41588-022-01024-z.

113. Dumitrescu, L., Mahoney, E.R., Mukherjee, S., Lee, M.L., Bush, W.S., Engelman, C.D., Lu, Q., Fardo, D.W., Trittschuh, E.H., Mez, J., et al. (2020). Genetic variants and functional

pathways associated with resilience to Alzheimer's disease. Brain *143*, 2561–2575. 10.1093/brain/awaa209.

114. Jaenisch, R., and Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat Genet *33*, 245–254. 10.1038/ng1089.

115. Ladd-Acosta, C., and Fallin, M.D. (2016). The role of epigenetics in genetic and environmental epidemiology. Epigenomics *8*, 271–283. 10.2217/epi.15.102.

116. Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet *13*, 484–492. 10.1038/nrg3230.

117. Bakulski, K.M., and Fallin, M.D. (2014). Epigenetic epidemiology: promises for public health research. Environmental and molecular mutagenesis *55*, 171–183.

118. Marioni, R.E., McRae, A.F., Bressler, J., Colicino, E., Hannon, E., Li, S., Prada, D., Smith, J.A., Trevisi, L., Tsai, P.-C., et al. (2018). Meta-analysis of epigenome-wide association studies of cognitive abilities. Mol Psychiatry *23*, 2133–2144. 10.1038/s41380-017-0008-y.

119. Yang, A.W., Sachs, A.J., and Nystuen, A.M. (2015). Deletion of Inpp5a causes ataxia and cerebellar degeneration in mice. Neurogenetics *16*, 277–285. 10.1007/s10048-015-0450-4.

120. Andrews, S.J., Das, D., Anstey, K.J., and Easteal, S. (2017). Late Onset Alzheimer's Disease Risk Variants in Cognitive Decline: The PATH Through Life Study. JAD *57*, 423–436. 10.3233/JAD-160774.

121. Chouliaras, L., Mastroeni, D., Delvaux, E., Grover, A., Kenis, G., Hof, P.R., Steinbusch, H.W.M., Coleman, P.D., Rutten, B.P.F., and van den Hove, D.L.A. (2013). Consistent decrease in global DNA methylation and hydroxymethylation in the hippocampus of Alzheimer's disease patients. Neurobiology of Aging *34*, 2091–2099. 10.1016/j.neurobiolaging.2013.02.021.

122. Mastroeni, D., Grover, A., Delvaux, E., Whiteside, C., Coleman, P.D., and Rogers, J. (2010). Epigenetic changes in Alzheimer's disease: Decrements in DNA methylation. Neurobiology of Aging *31*, 2025–2037. 10.1016/j.neurobiolaging.2008.12.005.

123. Rao, J.S., Keleshian, V.L., Klein, S., and Rapoport, S.I. (2012). Epigenetic modifications in frontal cortex from Alzheimer's disease and bipolar disorder patients. Transl Psychiatry *2*, e132. 10.1038/tp.2012.55.

124. Smith, R.G., Pishva, E., Shireby, G., Smith, A.R., Roubroeks, J.A.Y., Hannon, E., Wheildon, G., Mastroeni, D., Gasparoni, G., Riemenschneider, M., et al. (2021). A meta-analysis of epigenome-wide association studies in Alzheimer's disease highlights novel differentially methylated loci across cortex. Nat Commun *12*, 3517. 10.1038/s41467-021-23243-4.

125. De Jager, P.L., Srivastava, G., Lunnon, K., Burgess, J., Schalkwyk, L.C., Yu, L., Eaton, M.L., Keenan, B.T., Ernst, J., McCabe, C., et al. (2014). Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. Nat Neurosci *17*, 1156–1163. 10.1038/nn.3786.

126. Coppieters, N., Dieriks, B.V., Lill, C., Faull, R.L.M., Curtis, M.A., and Dragunow, M. (2014). Global changes in DNA methylation and hydroxymethylation in Alzheimer's disease human brain. Neurobiology of Aging *35*, 1334–1344. 10.1016/j.neurobiolaging.2013.11.031.

127. Li, Q.S., Sun, Y., and Wang, T. (2020). Epigenome-wide association study of Alzheimer's disease replicates 22 differentially methylated positions and 30 differentially methylated regions. Clin Epigenet *12*, 1–14. 10.1186/s13148-020-00944-z.

128. Yu, L., Chibnik, L.B., Srivastava, G.P., Pochet, N., Yang, J., Xu, J., Kozubek, J., Obholzer, N., Leurgans, S.E., Schneider, J.A., et al. (2015). Association of Brain DNA Methylation in SORL1, ABCA7, HLA-DRB5, SLC24A4, and BIN1 With Pathological Diagnosis of Alzheimer Disease. JAMA Neurology *72*, 15–24. 10.1001/jamaneurol.2014.3049.

129. Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet *10*, 57–63. 10.1038/nrg2484.

130. Genomics, F.L., and Gunn, S. (2020). TWAS: Transcriptome-wide association study. Front Line Genomics. https://frontlinegenomics.com/twas-transcriptome-wide-association-study/.

131. Harries, L.W., Bradley-Smith, R.M., Llewellyn, D.J., Pilling, L.C., Fellows, A., Henley, W., Hernandez, D., Guralnik, J.M., Bandinelli, S., Singleton, A., et al. (2012). Leukocyte CCR2 Expression Is Associated with Mini-Mental State Examination Score in Older Adults. Rejuvenation Research *15*, 395–404. 10.1089/rej.2011.1302.

132. Nygaard, M., Larsen, M.J., Thomassen, M., McGue, M., Christensen, K., Tan, Q., and Christiansen, L. (2019). Global expression profiling of cognitive level and decline in middle-aged monozygotic twins. Neurobiology of Aging *84*, 141–147. 10.1016/j.neurobiolaging.2019.08.019.

133. Tang, S., Buchman, A.S., Jager, P.L.D., Bennett, D.A., Epstein, M.P., and Yang, J. (2021). Novel Variance-Component TWAS method for studying complex human diseases with applications to Alzheimer's dementia. PLOS Genetics *17*, e1009482. 10.1371/journal.pgen.1009482.

134. Nagpal, S., Meng, X., Epstein, M.P., Tsoi, L.C., Patrick, M., Gibson, G., De Jager, P.L., Bennett, D.A., Wingo, A.P., Wingo, T.S., et al. (2019). TIGAR: An Improved Bayesian Tool for Transcriptomic Data Imputation Enhances Gene Mapping of Complex Traits. The American Journal of Human Genetics *105*, 258–266. 10.1016/j.ajhg.2019.05.018.

135. Blalock, E.M., Geddes, J.W., Chen, K.C., Porter, N.M., Markesbery, W.R., and Landfield, P.W. (2004). Incipient Alzheimer's disease: microarray correlation analyses reveal major

transcriptional and tumor suppressor responses. Proceedings of the National Academy of Sciences *101*, 2173–2178.

136. Berchtold, N.C., Sabbagh, M.N., Beach, T.G., Kim, R.C., Cribbs, D.H., and Cotman, C.W. (2014). Brain gene expression patterns differentiate mild cognitive impairment from normal aged and Alzheimer's disease. Neurobiology of Aging *35*, 1961–1972. 10.1016/j.neurobiolaging.2014.03.031.

137. Colangelo, V., Schurr, J., Ball, M.J., Pelaez, R.P., Bazan, N.G., and Lukiw, W.J. (2002). Gene expression profiling of 12633 genes in Alzheimer hippocampal CA1: transcription and neurotrophic factor down-regulation and up-regulation of apoptotic and pro-inflammatory signaling. Journal of neuroscience research *70*, 462–473.

138. Maes, O.C., Xu, S., Yu, B., Chertkow, H.M., Wang, E., and Schipper, H.M. (2007). Transcriptional profiling of Alzheimer blood mononuclear cells by microarray. Neurobiology of aging *28*, 1795–1809.

139. Fessel, J. (2020). The paradox of opposite directions of gene expressions in MCI and AD suggests possible therapy to prevent progression of MCI to AD. Alzheimer's & Dementia: Translational Research & Clinical Interventions *6*, e12003. 10.1002/trc2.12003.

140. Baguette, M., Legrand, D., and Stevens, V.M. (2015). An Individual-Centered Framework For Unravelling Genotype-Phenotype Interactions. Trends in Ecology & Evolution *30*, 709–711. 10.1016/j.tree.2015.10.003.

141. Hindorff, L.A., Bonham, V.L., Brody, L.C., Ginoza, M.E.C., Hutter, C.M., Manolio, T.A., and Green, E.D. (2018). Prioritizing diversity in human genomics research. Nat Rev Genet *19*, 175–185. 10.1038/nrg.2017.89.

142. Popejoy, A.B., and Fullerton, S.M. (2016). Genomics is failing on diversity. Nature *538*, 161–164. 10.1038/538161a.

143. Daniels, P.R., Kardia, S.L.R., Hanis, C.L., Brown, C.A., Hutchinson, R., Boerwinkle, E., Turner, S.T., and Genetic Epidemiology Network of Arteriopathy study (2004). Familial aggregation of hypertension treatment and control in the Genetic Epidemiology Network of Arteriopathy (GENOA) study. Am. J. Med. *116*, 676–681. 10.1016/j.amjmed.2003.12.032.

144. Folstein, M.F., Folstein, S.E., and McHugh, P.R. (1975). "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. Journal of Psychiatric Research *12*, 189–198. 10.1016/0022-3956(75)90026-6.

145. Tombaugh, T.N. (2004). Trail Making Test A and B: Normative data stratified by age and education. Archives of Clinical Neuropsychology *19*, 203–214. 10.1016/S0887-6177(03)00039-8.

146. Small, B.J., Viitanen, M., and Backman, L. (1997). Mini-Mental State Examination Item Scores as Predictors of Alzheimer's Disease: Incidence Data From the Kungsholmen

Project, Stockholm. The Journals of Gerontology: Series A *52A*, M299–M304. 10.1093/gerona/52A.5.M299.

147. Lezak, P. of N.P. and N.M.D., Lezak, M.D., Howieson, A.P. of N. and P.D.B., Howieson, D.B., Loring, P. of N.D.W., Loring, D.W., and Fischer, J.S. (2004). Neuropsychological Assessment (Oxford University Press).

148. Smith, J.A., Mosley Jr, T.H., Turner, S.T., and Kardia, S.L. (2012). Shared Genetic Effects among Measures of Cognitive Function and Leukoaraiosis. Edited by Amit Agrawal, 39.

149. Jaeger, J. (2018). Digit Symbol Substitution Test. J Clin Psychopharmacol *38*, 513–519. 10.1097/JCP.0000000000000941.

150. Sutin, A.R., Stephan, Y., Luchetti, M., and Terracciano, A. (2019). Five-factor model personality traits and cognitive function in five domains in older adulthood. BMC Geriatrics *19*, 343. 10.1186/s12877-019-1362-1.

151. Malek-Ahmadi, M., Small, B.J., and Raj, A. (2012). The Diagnostic Value of Controlled Oral Word Association Test-FAS and Category Fluency in Single-Domain Amnestic Mild Cognitive Impairment. Dement Geriatr Cogn Disord *32*, 235–240. 10.1159/000334525.

152. Estévez-González, A., Kulisevsky, J., Boltes, A., Otermín, P., and García-Sánchez, C. (2003). Rey verbal learning test is a useful tool for differential diagnosis in the preclinical phase of Alzheimer's disease: comparison with mild cognitive impairment and normal aging. International Journal of Geriatric Psychiatry *18*, 1021–1028. 10.1002/gps.1010.

153. Scarpina, F., and Tagini, S. (2017). The Stroop Color and Word Test. Frontiers in Psychology *8*.

154. Jensen, A.R., and Rohwer, W.D. (1966). The stroop color-word test: A review. Acta Psychologica *25*, 36–93. 10.1016/0001-6918(66)90004-7.

155. Kane, M.J., and Engle, R.W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. Journal of Experimental Psychology: General *132*, 47–70. 10.1037/0096-3445.132.1.47.

156. Stebbins, G. (2007). Neuropsychological Testing. Textbook of Clinical Neurology: Third Edition, 539–557. 10.1016/B978-141603618-0.10027-X.

157. Jack, C.R., Twomey, C.K., Zinsmeister, A.R., Sharbrough, F.W., Petersen, R.C., and Cascino, G.D. (1989). Anterior temporal lobes and hippocampal formations: normative volumetric measurements from MR images in young adults. Radiology *172*, 549–554. 10.1148/radiology.172.2.2748838.

158. Jack, C.R., O'Brien, P.C., Rettman, D.W., Shiung, M.M., Xu, Y., Muthupillai, R., Manduca, A., Avula, R., and Erickson, B.J. (2001). FLAIR histogram segmentation for

measurement of leukoaraiosis volume. J Magn Reson Imaging *14*, 668–676. 10.1002/jmri.10011.

159. Smith, J.A., Turner, S.T., Sun, Y.V., Fornage, M., Kelly, R.J., Mosley, T.H., Jack, C.R., Kullo, I.J., and Kardia, S.L. (2009). Complexity in the genetic architecture of leukoaraiosis in hypertensive sibships from the GENOA Study. BMC Med Genomics *2*, 16. 10.1186/1755-8794-2-16.

160. Fortin, J.-P., Fertig, E., and Hansen, K. (2014). shinyMethyl: interactive quality control of Illumina 450k DNA methylation arrays in R. 10.12688/f1000research.4680.2.

161. Xu, Z., Niu, L., Li, L., and Taylor, J.A. (2016). ENmix: a novel background correction method for Illumina HumanMethylation450 BeadChip. Nucleic Acids Research *44*, e20–e20. 10.1093/nar/gkv907.

162. Fortin, J.-P., Triche, T.J., Jr, and Hansen, K.D. (2017). Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. Bioinformatics *33*, 558–560. 10.1093/bioinformatics/btw691.

163. Aryee, M.J., Jaffe, A.E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A.P., Hansen, K.D., and Irizarry, R.A. (2014). Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics *30*, 1363–1369. 10.1093/bioinformatics/btu049.

164. Niu, L., Xu, Z., and Taylor, J.A. (2016). RCP: a novel probe design bias correction method for Illumina Methylation BeadChip. Bioinformatics *32*, 2659–2663. 10.1093/bioinformatics/btw285.

165. Lehne, B., Drong, A.W., Loh, M., Zhang, W., Scott, W.R., Tan, S.-T., Afzal, U., Scott, J., Jarvelin, M.-R., Elliott, P., et al. (2015). A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. Genome Biology *16*, 37. 10.1186/s13059-015-0600-x.

166. Hansen, K. IlluminaHumanMethylationEPICanno.ilm10b2.hg19: Annotation for Illumina's EPIC methylation arrays. R package version 0.6.0.

167. Houseman, E.A., Accomando, W.P., Koestler, D.C., Christensen, B.C., Marsit, C.J., Nelson, H.H., Wiencke, J.K., and Kelsey, K.T. (2012). DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics *13*, 86. 10.1186/1471-2105-13-86.

168. Du, P., Zhang, X., Huang, C.-C., Jafari, N., Kibbe, W.A., Hou, L., and Lin, S.M. (2010). Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics *11*, 587. 10.1186/1471-2105-11-587.

169. Weisenberger, C.D.J., and Laird, P.W. Comprehensive DNA Methylation Analysis on the Illumina® Infinium® Assay Platform. 4.

170. Lockstone, H.E. (2011). Exon array data analysis using Affymetrix power tools and R statistical software. Briefings in Bioinformatics *12*, 634–644. 10.1093/bib/bbq086.

171. Irizarry, R.A., Bolstad, B.M., Collin, F., Cope, L.M., Hobbs, B., and Speed, T.P. (2003). Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Research *31*, e15. 10.1093/nar/gng015.

172. Dai, M., Wang, P., Boyd, A.D., Kostov, G., Athey, B., Jones, E.G., Bunney, W.E., Myers, R.M., Speed, T.P., Akil, H., et al. (2005). Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. Nucleic Acids Research *33*, e175. 10.1093/nar/gni179.

173. Saha, A., and Battle, A. (2018). False positives in trans-eQTL and co-expression analyses arising from RNA-sequencing alignment errors. 10.12688/f1000research.17145.1.

174. Johnson, W.E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics *8*, 118–127. 10.1093/biostatistics/kxj037.

175. Williams, D.R. (1997). Race and health: basic questions, emerging directions. Ann Epidemiol *7*, 322–333. 10.1016/s1047-2797(97)00051-3.

Chapter 2 . SNP-by-CpG Interactions in *ABCA7* are Associated with Cognition in Older African Americans

2.1 Abstract

SNPs in *ABCA7* confer the largest genetic risk for Alzheimer's Disease (AD) in African Americans (AA) after APOE ε4. However, the relationship between *ABCA7* and cognitive function has not been thoroughly examined. We investigated the effects of five known AD risk SNPs and 72 CpGs in *ABCA7*, as well as their interactions, on general cognitive function (cognition) in 634 older AA without dementia from Genetic Epidemiology Network of Arteriopathy (GENOA). Using linear mixed models, no SNP or CpG was associated with cognition at FDR q<0.1, but five CpGs were nominally associated (P<0.05). Four SNP-by-CpG interactions were associated with cognition (FDR q<0.1). Contrast tests show that methylation is associated with cognition in some genotype groups (P<0.05): a 1% increase at cg00135882 and cg22271697 is associated with a 0.68 SD decrease and 0.14 SD increase in cognition for those with the rs3764647 GG/AG (P=0.004) and AA (P=0.0002) genotypes, respectively. Also, a 1% increase at cg06169110 and cg17316918 is associated with a 0.37 SD decrease (P=0.0002) and 0.33 SD increase (P=0.004), respectively, in cognition for those with the rs115550680 GG/AG genotype. While AD risk SNPs in *ABCA7* are not associated with cognition in this sample, some have interactions with proximal methylation on cognition.

2.2 Introduction

Alzheimer's disease (AD) is a progressive neurodegenerative disease characterized by the dysregulation of the amyloid-β (Aβ) pathway leading to Aβ plaques[1] and the aggregation of tau tangles.[2] AD accounts for 60-80% of dementia cases in the elderly.[3–5] Approximately 6.2 million Americans age 65 and older are living with AD, and this estimate is projected to rise to 13.8 million by 2060.[3] AD risk differs by race, with African Americans (AA) twice as likely to develop AD compared to European Americans (EA).[6] Because this health disparity places a greater burden of personal and medical care on AA, it is crucial to better understand AD and its development in this population.

AD is a multifactorial disease that is likely influenced by interactions between genetic, environmental, and epigenetic factors, along with age-related neurodegeneration.[7] In addition to age, genetic variants in the apolipoprotein E (*APOE*) gene are the largest risk factor for AD in AA,[8] with one copy of the *APOE* $\varepsilon 4$ allele increasing AD risk by 3-5 fold.[9–11] *ABCA7* is the second largest genetic risk factor for AD in AA, with genetic variants increasing AD risk by 70-80%.[8] The *ABCA7* gene encodes the ATP-binding cassette (ABC) transporter A7 that regulates homeostasis of phospholipids and cholesterol in the central nervous system and peripheral tissues.[12–14] This gene is mostly expressed in the brain, spleen, lungs, and adrenal gland.[15] Studies suggested that mutations in *ABCA7* are associated with AD susceptibility through the dysregulation of lipid metabolism which facilitates $A\beta$ clearance.[16,17]

Though *ABCA7* is a risk locus for AD in both EA and AA, the specific risk variants differ across groups.[18] In EA, three *ABCA7* SNPs, rs3764650, rs3752246 and rs4147929, are associated with AD. They represent two independent signals as rs3752246 and rs4147929 are in nearly complete linkage disequilibrium (LD) in EA. Although rs3764650 shows the strongest

association with AD in EA, it is only nominally associated in AA.[18,19] In AA, two additional *ABCA7* SNPs, rs3764647 and rs3752239, have stronger associations with AD,[19] with rs3764647 being in the same LD block as rs3764650 in AA. Interestingly, another independent SNP in *ABCA7*, rs115550680, which is monomorphic in EA, is strongly associated with AD in AA. In particular, the G allele of rs115550680 confers an AD risk comparable to *APOE* ε4 (OR=1.79) in AA.[8]

Epigenetic modifications, such as DNA methylation, are potential molecular mechanisms that can modulate the effect of genetic risk factors.[20] When methylation sites (CpGs) are clustered together as a CpG island (CGI), it often serves as a hub for gene expression regulation. CGIs in the promoter region usually suppress transcription whereas CGIs in the intragenic region can interact with multiple regulatory elements to have a variety of impacts on gene expression (e.g., influence mRNA isoforms, promote enhancer function).[21] Given the regulatory role of DNA methylation on gene expression, there has been a growing interest in understanding the extent to which DNA methylation contributes to AD risk.[22–26] In particular, recent studies of post-mortem brain tissue found evidence of association between DNA methylation in *ABCA7* and both AD and AD-related pathologies, including Aβ load and tau tangle density.[23,24] This evidence suggests that methylation in *ABCA7* has a non-trivial functional role that is worthy of further investigation.

Although the relationships between AD and *ABCA7* SNPs are well-characterized, there are limited studies on the association between genetic variation in *ABCA7* and measures of cognitive function and/or cognitive decline prior to the development of dementia. An imaging study showed that *ABCA7* SNPs were associated with amyloidosis among cognitively healthy individuals and those with mild cognitive impairment (MCI), but not among those with AD,

suggesting an early effect of *ABCA7* on cognition and cognitive decline.[27] A few studies in EA found inconsistent results for the effect of *ABCA7* SNPs on cognition, with associations varying by sex, *APOE* status, and disease progression.[28] For example, in healthy older adults, a longitudinal study found association between rs3764650 and cognitive decline, but only in females.[29] Also, interactions between *APOE* ε4 allele and SNPs rs3764650 and rs3752246 were associated with three cognitive factor scores related to Verbal Learning and Memory, Working Memory, and Intermediate Memory, in a genotype dependent manner: in the absence of *ABCA7* minor alleles, each additional ε4 allele was associated with lower memory scores; and conversely, in the presence of *ABCA7* minor alleles, each additional ε4 allele was associated with better memory scores.[30] Lastly, rs3764650 was significantly associated with increased rates of memory decline among individuals with MCI or AD.[31]

To our knowledge, no study has investigated the relationship between *ABCA7* genetic variation and cognition in cognitively healthy AA. Further, few studies have examined the relationship between DNA methylation in *ABCA7* and/or its interaction with genetic variants on general cognitive function. In this study, we investigate whether previously identified risk SNPs (referred to as sentinel SNPs) in *ABCA7*, DNA methylation in *ABCA7*, and their interactions are associated with general cognitive function in older AA without dementia. To better understand the functional consequence of these risk factors at the molecular level, we also evaluated whether identified epigenetic or genetic risk factors are associated with transcript level *ABCA7* gene expression in transformed beta lymphocytes from the same cohort. A thorough investigation of the relationship between these multi-omic layers and later-life cognition can help characterize the underlying genetic architecture of cognition in older adulthood, prior to dementia onset. This

may allow the identification of targets for intervention and treatment, especially in populations that are most at risk.[32]

2.3 Materials and Methods

2.3.1 Sample

The Genetic Epidemiology Network of Arteriopathy (GENOA) study is a community-based longitudinal study aimed at examining the genetic effects of hypertension and related target organ damage.[33] European American (EA) and African American (AA) hypertensive sibships were recruited if at least two siblings were clinically diagnosed with hypertension before age 60. All other siblings were invited to participate, regardless of hypertension status. Exclusion criteria included secondary hypertension, alcoholism or drug abuse, pregnancy, insulin-dependent diabetes mellitus, active malignancy, or serum creatinine levels >2.5mg/dL. In Phase I (1996-2001), 1,854 AA participants (Jackson, MS) and 1,583 EA participants (Rochester, MN) were recruited.[33] In Phase II (2000-2004), 1,482 participants AA participants and 1,239 EA participants were successfully followed up, and their potential target organ damage from hypertension was measured. Demographics, medical history, clinical characteristics, information on medication use, and blood samples were collected in each phase. Methylation levels were measured only in AA participants using blood samples collected in Phases I and II. In an ancillary study (2001-2006), 1010 AA and 967 EA GENOA participants underwent a battery of established neurocognitive tests to assess several measures of cognitive function, including learning, memory, attention, concentration, and language. Written informed consent was obtained from all participants, and approval was granted by participating institutional review boards (University of Michigan, University of Mississippi Medical Center, and Mayo Clinic).

A total of 850 AA participants had non-missing genetic and demographic data. Since participants with a history of stroke or dementia may have changes in general cognitive function that differ from non-pathological cognitive aging, we excluded those who had a history of stroke (N=43) and/or preliminary evidence of dementia as indicated by a score of <24 on the Mini-Mental State Examination (MMSE) (N=76).[34] We also excluded participants younger than age 45 (N=16). A total of 634, 494 and 429 participants were available for SNP, methylation, and gene expression analyses, respectively (Figure S2-3).

2.3.2 Measures

A. General cognitive function

General cognitive function was calculated using five neurocognitive measures evaluated at Phase II:[34,35]

6. Weschler Adult Intelligence Scale-Revised: Digit Symbol Substitution Test (DSST) measured complex visual attention, sustained and focused concentration, response speed and visuomotor coordination. DSST relates to the executive function of working memory in cognition.[36] The score comprised the number of symbols correctly matched within 90 seconds.
7. The Controlled Oral Word Association Test (COWA-FAS) tested for verbal fluency (phonetic association) and language. This required participants to generate as many words as possible that start with F, A, and S in 1 minute. The score consisted of the total number of admissible words generated.
8. Rey Auditory Verbal Learning Test (RAVLT) measured delayed recall, relating to the cognitive functions of new learning, immediate memory span and vulnerability to

interference in learning and recognition memory. Its score was determined by the number of words recalled after a 30-minute delay. Scores ranged from 0 to 15.

9. Stroop Color-Word Test (SCWT) assessed concentration effectiveness by taking the sum of the color words that were correctly stated in 45 seconds. Specifically, the ability to shift perceptual sets in response to novel stimuli, was tested.
10. Trail Making Test A (TMTA) evaluated visual conceptual tracking as participants need to connect a set of 25 circles quickly and accurately. TMTA provides information on the cognitive functions of visual search, scanning, processing speed and executive functions. The TMTA score was measured as the amount of time (seconds) the participants took to complete the task. The maximum time allowed was 240 seconds. Prior to analysis, TMTA scores were natural log transformed and recoded so that higher scores indicate better cognitive function.

General cognitive function, a measure of overall cognitive performance, can be quantified as a summary measure of cognitive tests in multiple cognitive domains.[37] In this study, general cognitive function was calculated as the first unrotated principal component (FUPC) from a principal component analysis (PCA) of the five neurocognitive measures in the full sample (N=634). The FUPC accounted for 53% of the total variance in the neurocognitive measures and loading values of the five measures ranged from 0.52 to 0.87.

B. Demographic data

Age was assessed at cognitive testing. Educational attainment, measured at Phase II, was categorized into a three-level variable of (1) less than high school degree (reference group), (2) high school degree or GED, and (3) at least some college. Smoking has been shown to have

substantial impact on the epigenome [38], so we used smoking data from the same timepoint as the DNA methylation measures (Phase I). Participants were categorized as current, former, or never smokers (reference group).

C. Genetic data

Blood samples were genotyped using the Affymetrix® Genome-Wide Human SNP Array 6.0 or the Illumina 1M Duo. Samples and SNPs with a call rate <95%, samples with mismatch sex, and duplicate samples were removed. Genotypes were imputed using the 1000 Genomes Project phase I integrated variant set (v.3) (Hg19, released in March 2012). Of the six SNPs of interest identified from existing literature (rs3764647, rs3764650, rs115550680, rs3752246, rs3752239 and rs4147929), five had high imputation quality ($r^2>0.7$), and one (rs3752239) was excluded due to low imputation quality ($r^2=0.49$). SNPs were coded as the dosage of the corresponding AD risk allele as specified in the previous literature. Genetic principal components were calculated from genotyped SNPs and included in regression models to control for population stratification. To evaluate confounding and/or effect modification by *APOE* isoforms known to influence dementia risk, we measured rs7412 (to capture the *APOE* ε2 allele) and rs429359 (to capture the *APOE* ε4 allele) using a TaqMan assay and ABI Prism© Sequence Detection (Applied Biosystems, Foster City, CA) in 1544 participants. Participants were classified as having 0, 1, or 2 copies of ε2 (rs7412 T allele) and/or ε4 (rs429359 C allele).

D. Methylation measures

Genomic data was extracted from stored peripheral blood leukocytes from 1,106 AA participants from Phase I and 304 AA participants from Phase II using the AutoGen FlexStar

(AutoGen, Holliston, MA). Bisulfite conversion was performed with the EZ DNA Methylation Kit (Zymo Research, Irvine, CA), and methylation was measured using the Illumina HumanMethylationEPIC BeadChip. The raw intensity data was visualized using the shinyMethyl R package [39] to identify sex mismatches and outliers, which were removed. Samples with incomplete bisulfite conversion were identified using Qcinfo in the *Enmix* R package [40] and removed. Background correction and dye-bias normalization were performed using Noob in the *Minfi* R package.[41,42] We also checked sample identity using the 59 SNP probes on the EPIC chip, and mismatched samples were removed. Probe-type bias was adjusted using the Regression on Correlated Probes (RCP) method.[43] Probes with detection p-value $<10^{-16}$ were considered successfully detected, and probes and samples with detection rate<10% were removed.[44] After quality control, a total of 1,396 samples (N=1,100 from Phase I and N=294 from Phase II) and 857,121 CpG sites were available for analyses. For this analysis, all methylation data was from Phase I samples.

We selected all CpG sites within 5kb of the *ABCA7* gene (a total of 72 CpG sites within the *ABCA7* region: chr19, 1040102–1065570, hg19). We used Illumina annotation [45] to characterize each CpG site as being in a promoter region and/or CGI, CGI shore, or CGI shelf. White blood cell proportions for CD8+ T lymphocytes, CD4+ T lymphocytes, natural killer cells, B cells, monocytes, and granulocytes were estimated using the Houseman method.[46] For each CpG site prior to analysis, the methylation beta value [47,48] was multiplied by 100 to approximate the percent methylation at that site. Methylation beta values were pre-adjusted for batch effects (sample plate, row, and column) and white blood cell proportions using linear mixed modelling, and the resulting residuals were added to the mean values.

E. Gene expression measures

Gene expression levels in transformed beta-lymphocyte cell lines from blood samples taken primarily at GENOA Phase II were measured using the Affymetrix Human Transcriptome Array 2.0. The Affymetrix Expression Console was used for quality control, and all array images passed visual inspection. Affymetrix Power Tool software was used to process raw intensity data.[49] We normalized Affymetrix CEL files using the Robust Multichip Average (RMA) algorithm, including background correction, quantile normalization, $\log_2$-transformation and probe set summarization.[50] Linearity was also maintained using GC correction (GCCN), signal space transformation (SST), and gain lock (value=0.75). We used the Brainarray custom CDF [51] version 19 to map the probes to genes, specifically removing probes with non-unique matching cDNA/EST sequences that can be assigned to more than one gene cluster. As a result, the gene expression data processed through the custom CDF is expected to be free of mappability issues; however, alignment bias may still exist due to genetic variation, errors in reference genome, and other complications.[52] After mapping, Combat was used to remove batch effects.[53]

2.3.3 Statistical analysis

A. Genetic analysis

We first calculated Pearson correlations between sentinel SNPs. Next, the association between *ABCA7* sentinel SNPs and general cognitive function was analyzed using linear mixed models with random effects to adjust for relatedness. Model 1 adjusted for age at cognitive testing, sex, and the first four genetic principal components (PC1-4), with family as a random effect to account for sibships. Model 2 additionally adjusted for educational attainment. Model 3 further adjusted for *APOE* ε2 and ε4. For any SNPs that were significantly associated with

general cognitive function, we further examined the association between those SNPs and each of the five neurocognitive measures to identify the domain(s) that most strongly drive the association. Since prior studies suggest that the effect of *ABCA7* SNPs may vary by sex, education and/or *APOE* status, we also assessed the interaction between the sentinel SNPs and sex, education or *APOE* (ε2 and ε4) on cognitive outcomes.

B. Epigenetic analysis

Pearson correlations were calculated among all 72 CpG sites. Next, linear mixed models were used to test the associations between each of the 72 CpG sites and general cognitive function. Model 1 adjusted for basic covariates including age at cognitive testing, sex, four genetic principal components, age difference between methylation and cognition measurements, smoking status, and family as a random effect to account for sibships. Model 2 additionally adjusted for educational attainment, and Model 3 further adjusted for *APOE* ε2 and ε4. The coMET package was used to create a regional plot to visualize association P-values, correlations, and Ensembl genes.[54] BioRender was used to annotate and format the figure.[55] For any CpGs that were significantly associated with general cognitive function, we further examined the association between those CpGs and each of the five neurocognitive measures to identify the domain(s) that most strongly drive the association.

C. Genetic-epigenetic interaction analysis

We next examined the interaction between each CpG site and sentinel *ABCA7* SNPs in association with general cognitive function. In this analysis, we adjusted for age at cognitive testing, sex, four genetic principal components, age difference between methylation and

cognition measurements, smoking status, and *APOE* ε2 and ε4, with family as a random effect to account for sibships (Model 4). Models 1-4 that are used to assess genetic, epigenetic and genetic-epigenetic interaction associations with general cognitive function are shown in Figure S2-4. To improve interpretability, we mean-centered methylation so that the estimated betas reflect the effect sizes for those with average methylation in the population. For any identified significant interaction, we stratified the genotypes by number of risk alleles (0, 1, or 2 risk alleles) and estimated the marginal means for linear trend (Emtrends function) using the Emmeans [56] package in R. Contrast tests were also conducted to obtain the effect size of the CpG associated with general cognitive function in each genotype group. Minor homozygote genotype groups that were <5% of the sample size were grouped with heterozygous genotype groups to increase power as appropriate. Plots of SNP-by-CpG interactions on general cognitive function were generated using the effects [57] and ggplot2 [58] packages in R. Any identified SNP-by-CpG interactions significantly associated with general cognitive function were also tested for association with each of the five neurocognitive measures.

As a sensitivity analysis for significant interactions (FDR q<0.1), we tested the association after excluding outlying CpG values that were more than four standard deviations from the mean (Model 4). We then assessed whether the SNP-by-CpG interactions (FDR q<0.1) were driven by potential SNP-CpG correlations by testing the association between each SNP and its corresponding CpG, adjusting for age at methylation measurement, sex, first four genetic principal components, with family as a random effect. If the SNP and CpG were associated at P<0.05, we adjusted out the effect of the SNP from the CpG site and re-tested the interaction (Model 4).

D. Gene expression analysis

Among the 494 participants with methylation and genetic data, 429 participants also had gene expression data. Figure S2-5 presents a graphical depiction of *ABCA7* transcripts observed in the Genotype Tissue Expression (GTEx) project,[59] which assesses gene expression levels in a variety of cell types. A total of 17 transcripts, along with a measure of overall *ABCA7* gene expression, were available for analysis in our study. For SNPs, CpGs, or interactions that were significantly associated with general cognitive function, we assessed their association with *ABCA7* gene-level expression and transcripts (Model 5) using linear mixed models. Model 5 adjusted for age at which gene expression data was generated (age at blood draw), sex, first four genetic principal components, and family as a random effect. For models that included CpG sites, Model 5 also included the age difference between methylation and gene expression measurements. Contrast tests were conducted to obtain the effect size in each genotype group. Minor homozygote genotype groups (<5% sample size) were grouped with heterozygous genotype groups to increase power as appropriate.

We next evaluated whether the identified CpG sites within the *ABCA7* region, including within the promoter region (chr19, 1,037,800-1,043,201),[60] correlate with gene expression of *ABCA7* and/or nearby genes. For this, we used *cis*-eQTM results from peripheral blood mononuclear cells (PBMCs) and three specific white blood cell types (CD4+T lymphocytes, monocytes and neutrophils) in the iMETHYL database,[61,62] which integrates genotype, methylation, and gene expression data from 102 individuals. We also examined gene expression levels of *ABCA7* in different cell types available from the Genotype Tissue Expression (GTEx) project.[59]

E. Multiple testing correction

All statistical analyses were conducted in R (version 3.6).[63] For genetic analysis, Bonferroni corrected p-value cut off ($p<0.05/5$) was used to claim significance. For all other analyses, false discovery rate (FDR) correction was applied to each model, and FDR $q<0.1$ was considered significant. Since the SNPs, CpG sites, and transcripts in *ABCA7* are all correlated, applying stringent multiple testing corrections might be too conservative, thus any nominal associations are also noted.

2.4 Results

2.4.1 Sample characteristics

The sample included 634 AA without dementia (Table 2-1). Overall, the participant age ranged from 45 to 85 years (mean=63.3 years), and the mean age difference between Phase I methylation and cognitive measurements was 6.0 years (SD=1.3). More than half of participants (74.9%) were female, and 47.3% had at least some college education. General cognitive function was normally distributed. Mean RAVLT score was 7.1 (SD=3.3) words recalled, mean DSST score was 34.4 (SD=12.6) symbols, mean COWA-FAS score was 29.7 (SD=11.6) words, mean SCWT score was 22.5 (SD=9.8) items, and mean TMTA score was 61.6 (SD=32.0) seconds to completion.

2.4.2 Correlation among six cognitive outcomes

Pearson correlations (r) among the six cognitive outcomes (general cognitive function and the five individual neurocognitive measures) are shown in Table S2-4. The five neurocognitive

measures were moderately correlated (Pearson r ranged from 0.24 to 0.66), with the highest correlation between DSST and TMTA ($r=0.66$, $P<0.001$).

2.4.3 Correlation among ABCA7 SNPs

Pearson correlations among the five sentinel *ABCA7* SNPs are shown in Table S2-5. Rs3764647 was strongly correlated with rs3764650 ($r=0.84$, $P<0.001$), and rs3752246 was highly correlated with rs4147929 ($r=0.96$, $P<0.001$). Otherwise, the other sentinel SNP pairs had low but significant correlations ranging from -0.14 to -0.004 ($p<0.05$).

2.4.4 Genetic associations

In Models 1 and 2, there were no *ABCA7* SNPs that met the nominal significance threshold ($p<0.05$, Table S2-6). Although *APOE* is not part of the primary analysis, *APOE* ε2 and ε4 were analyzed separately as exposures in Models 1 and 2. *APOE* ε4 was associated with general cognitive function in both models in the expected direction (higher dosage of ε4 was associated with lower cognitive function), but only met the Bonferroni-corrected significance threshold in Model 2. After adjusting for educational attainment and *APOE ε2* and *ε4* in Model 3, no sentinel SNPs were significantly associated with general cognitive function. There were no observed significant interactions between SNPs and sex, *APOE* isoforms, or educational attainment on general cognitive function.

2.4.5 Epigenetic associations

Among the 72 CpG sites examined, six were nominally associated with general cognitive function in at least one of the three Models (Table S2-7). After adjusting for educational attainment and *APOE* ε2 and ε4 (Model 3), five CpGs (cg22271697, cg00874873, cg11714200, cg26264438 and cg12082025) in the *ABCA7* region were nominally associated with general cognitive function. Figure 2-1 illustrates the regional plot of association P-values of the 72 CpGs in the *ABCA7* region with general cognitive function according to the chromosomal positions of CpG sites, as well as the correlations between the CpGs (Model 3).

2.4.6 Genetic-epigenetic associations

Since rs3764647 and rs3764650, as well as rs4147929 and rs3752246, are highly correlated with each other (Table S2-5), we removed one SNP from each pair and analyzed three independent risk SNPs (r<0.60) in the interaction analysis. Two of the independent SNPs we selected have previously been identified in AA GWAS (rs3764647 and rs115550680) [8,19] and the third one is the only *ABCA7* missense variant (p.Gly1527Ala) to be identified by GWAS (rs3752246).[64] We assessed the interaction between each of the three independent sentinel SNPs (rs3764647, rs115550680 and rs3752246) and 72 CpG sites on general cognitive function and identified four significant SNP-by-CpG interactions (FDR q<0.1) that were associated with general cognitive function (Table 2-2): rs3764647*cg00135882 (P=1.46E-04), rs3764647*cg22271697 (P=5.77E-04), rs115550680*cg06169110 (P=2.18E-04), rs115550680*cg17316918 (P=4.84E-04). The two SNPs and four CpGs that were involved in the four significant SNP-by-CpG interactions are shown in Figure 2-1 to highlight their positions with respect to neighboring genes, regulatory elements, and CGIs in the ABCA7 region. All

interactions with at least nominal significance are shown in Table S2-8. Notably, an additional seven CpG sites had nominally significant interactions with rs115550680, and one additional site had a nominally significant interaction with rs3764647. In Table S2-9, we present Pearson correlations among the *ABCA7* CpG sites that were nominally associated with general cognitive function (Table S2-7) and/or were involved in an FDR-significant SNP-by-CpG interaction (Table 2-2). The majority of these CpGs were weakly correlated or uncorrelated.

For interactions with FDR q<0.1, we performed contrast tests to estimate the effect size of the specific CpG site per genotype group. In all four cases, the minor homozygote genotype group had a small frequency (<5% of the sample size), thus we combined them with the corresponding heterozygote genotype group. Contrast tests show that methylation is associated with general cognitive function in some genotype groups, but not others (P<0.05; Table 2-3 and Figure 2-2).

Rs3764647 had significant interactions with two CpGs (cg00135882 and cg22271697). For those with the risk genotype (GG/AG), a 1% increase at cg00135882 is associated with a 0.68 SD decrease in general cognitive function (P=0.004, Figure 2-2A); whereas for those with the AA genotype, a 1% increase at cg22271697 is associated with a 0.14 SD increase in general cognitive function (P=2.00E-04, Figure 2-2B). Similarly, rs115550680 had interactions with two CpGs (cg06169110 and cg17316918). For those with the risk genotype (GG/AG), a 1% increase at cg06169110 is associated with a 0.37 SD decrease in general cognitive function (P=2.00E-04, Figure 2-2C), and a 1% increase at cg17316918 is associated with a 0.33 SD increase in general cognitive function (P=0.004, Figure 2-2D).

We performed a sensitivity analysis by excluding outlying CpG values beyond four standard deviations of mean methylation, and our results remained consistent (Table S2-10). To test

whether the interaction was driven by potential SNP-CpG correlation, we assessed the association between each SNP-CpG pair. We observed nominal associations between rs3764647 and cg22271697, as well as between rs115550680 and cg06169110. For these two SNP-CpG pairs, we regressed out the SNP effect from the corresponding CpGs and re-tested the interactions. The results remained consistent with those reported in Table 2-3 (Table S2-11). We also tested the association between all four significant interactions with each of the five neurocognitive domains. Similar interactions were observed for multiple neurocognitive measures, especially DSST and SCWT, in which all four interactions were significantly associated (Table S2-12).

2.4.7 Gene expression associations

To understand the functional effects of identified SNP-by-CpG interactions, we examined their interaction effects (Table S2-13 and S2-14) as well as marginal effects (Table S2-15 and Table S2-16) on *ABCA7* gene and transcript expression. At the gene level, none of the identified SNP-by-CpG interactions were associated with gene expression in our sample. However, we found a negative association between one of the SNPs, rs115550680, and ENSG00000064687: for each additional rs115550680 G allele, there is a 0.05 decrease in gene expression (P=0.027).

At the transcript level, two SNP-by-CpG interactions (rs115550680*cg17316918 and rs3764647*cg22271697) were nominally associated with two different transcripts (ENST00000525939 and ENST00000531467) (Table S2-13). ENST00000531467 (Chromosome 19: 1,062,261-1,063,945 forward strand) is a protein coding transcript with four coding exons (Figure S2-5). ENST00000525939 (Chromosome 19: 1,062,261-1,063,945 forward strand) is a retained intron, found primarily in the spleen, pituitary, whole blood and brain (cerebellum and

cerebellar hemisphere) (Figure S2-5). Although the interactions were only nominally significant, we performed contrast tests to estimate the effect size of the CpG site in each genotype group on each identified transcript. Contrast tests show that methylation at cg17316918 trends toward a positive association with ENST00000525939 among those with the rs115550680 risk genotype (GG/AG) but does not reach nominal significance (Table S2-14). We also assessed the marginal associations of the two SNPs and two CpGs involved in the interactions on each of the *ABCA7* transcripts (Table S2-15 and S2-16). We found that rs115550680 is negatively associated with 11 *ABCA7* transcripts, including ENST00000531467, after FDR correction (Table S2-15). Rs3764647 was positively associated with only ENST00000530703 (P=0.037; Table S2-15). Among CpGs involved in the interactions, cg06169110 was positively associated with two transcripts (Table S2-16).

The iMETHYL [60] *cis*-eQTM results for PBMCs and the three white blood cell types showed that there are CpGs within the *ABCA7* region, including within the promoter region, that regulate expression of both *ABCA7* and nearby genes. However, the CpGs identified in the significant SNP-by-CpG interactions in our study were not associated with gene expression of *ABCA7* or nearby genes at FDR q<0.05.

2.5 Discussion

While previous studies have implied that *ABCA7* is a causal gene for AD,[65–68] there is a dearth of studies examining the relationship between *ABCA7* and cognitive function. AD is a gradual neurodegenerative disease, characterized by noticeable cognitive impairment in areas of episodic memory, semantic memory, and executive function, with pathophysiology preceding the illness decades prior.[69,70] Studying the relationship between SNPs and CpGs in *ABCA7* and

cognition may enhance our understanding of cognitive health and further elucidate the role of *ABCA7* in cognitive aging preceding AD. To our knowledge, this study is the first assessment of the association, and interaction, between DNA methylation and genetic risk factors in *ABCA7* on cognition in AA without dementia.

In this study, we found no association between known AD-associated SNPs and cognitive measures. This is perhaps not surprising, as previous studies have been inconsistent regarding the association between *ABCA7* SNPs and cognition. Most of the studies, however, have been conducted in primarily European ancestral populations.[29–31,71] For example, the Three-City Dijon study found no association between *ABCA7* common variants and global cognition, as well as other cognitive outcomes.[71] Other studies in EA show that SNPs may be associated with cognition in subgroups stratified on gender,[29] *APOE* status[30] or disease progression.[31] In light of this, we also assessed whether *ABCA7* SNP associations are modified by sex, *APOE* major isoforms, and/or education status. Unlike prior studies,[29,30] we did not find any evidence of interaction. Lack of association with cognition for the sentinel SNP-by-sex and SNP-by-*APOE* interactions may be due to differences in ancestry or to small sample size as those studies have a sample size ranges from 1,153 to 3,267. Our study also did not find SNP-by-education associations interactions on cognition. This is consistent with another study that observed no interaction between education and *ABCA7* variants on memory performance in either EA or AA; however a weak signal was observed for memory decline in AA, which is a cognitive measure more related to AD and dementia.[72]

Other lines of evidence also suggest that the *ABCA7* risk variants may not be highly relevant to the neurological pathways underlying normal cognitive function and/or cognitive reserve. For example, previous GWAS for general cognitive function and AD have shown few overlapping

loci.[37,73] Further, studies of cognitively "resilient" individuals who live to an older age with intact cognitive function, despite the presence of AD neuropathology, have found the genetic architecture of cognitive resilience to be distinct from that of AD.[74] At this point, relatively little is known about the pathways involving genetic variants and cognitive aging in those without dementia. Thus, studying variants that affect general cognitive function in those without dementia may identify novel pathways for therapeutic targets.

Only one epigenome-wide association study (EWAS) has examined the association between all CpG sites across the genome, including CpGs in *ABCA7* gene, and general cognitive function in participants from multi-ethnic backgrounds.[75] This study did not identify any significant associations between *ABCA7* and general cognitive function. However, due to large numbers of CpG sites tested, the EWAS could have missed signals with smaller effect sizes. Moreover, the EWAS sample was mostly comprised of EA. Our study, which focuses on CpG sites in *ABCA7* in an AA cohort, would give us more power to detect an association in this region among AA. Nevertheless, we also failed to detect any associations between CpGs and general cognitive function after multiple testing correction, although six CpGs were associated at a nominal level. Importantly, we examined methylation levels in whole blood leukocytes, which is not the most relevant tissue for brain function. A study in post-mortem brain tissue found associations between CpGs in *ABCA7* and AD as well as increased burden of pathologies (e.g., Aβ load and tau tangle density), whereas another study failed to demonstrate differential methylation in peripheral blood between AD patients and controls.[23] Although methylation patterns differ between blood and brain tissues,[25,76] blood cells touch every cell bed that affects the brain, and are related to chronic inflammation and oxidative stress, which are linked to cognitive

performance.[77,78] Studying methylation in blood also allows us to study epigenetic associations with cognition in living participants in an inexpensive and non-invasive manner.

Although *ABCA7* sentinel SNPs and CpG sites were not associated with general cognitive function, we did see evidence of SNP-by-CpG interactions. Four interactions reached FDR significance (rs3764647*cg00135882, rs3764647*cg22271697, rs115550680*cg06169110, and rs115550680*cg17316918). Further, a total of nine CpG sites had nominally significant interactions with rs115550680 on cognition function. For participants who are homozygous for the rs115550680 major allele (AA), local methylation does not seem to have an effect on cognitive function. However, for the participants who carry the risk allele (GG/AG), methylation at local CpG may play an important role on cognition. This might be related to the different *ABCA7* transcripts that are involved in each case. Rs115550680 is located in an LD block that spans several introns and exons.[8] A prior study suggests that there is a 44-base pair exonic deletion (rs142076058, p.Arg578 fs) among rs115550680 G carriers, which could cause a frameshift in the *ABCA7*-coding sequence resulting in the formation of a premature termination codon (PTC).[79] Indeed, our gene expression analysis found that the risk allele (G) at rs115550680 was strongly associated with decreased expression of 11 *ABCA7* transcripts. Taken together, this suggests that this SNP might influence the major isoforms that are expressed, and the expressed alternative transcripts may influence cognitive function. Furthermore, alternative transcripts that are expressed in those carrying the risk allele may be further modulated by methylation level at local CpG sites, which may lead to differences in cognitive function in this group. Consistent with this hypothesis, methylation at cg17316918 was associated with transcript ENST00000525939 in rs115550680 risk allele carriers (GG/AG) only. Interestingly, this transcript is largely expressed in the brain. However, there is no prior evidence to show an

association between this transcript and AD and/or cognition. Nonetheless, alternative splicing of *ABCA7* is likely to play a similar important role in cognition as has been demonstrated in AD.[80,81]

The other SNP that had significant interactions with *ABCA7* CpG sites, rs3764647, is a missense mutation where the risk allele (G) leads to the amino acid change p.His395Arg in the first extracellular loop of the *ABCA7* protein.[18] One CpG site (cg00135882) is associated with cognitive function in participants who carry the risk allele (GG/AG) and another CpG site (cg22271697) is associated with cognitive function in those who do not carry risk allele (AA). This differential pattern may be due to different functions of the two transcripts instead of alternative splicing. Consistently, we did not observe a direct association between this SNP or CpG with expression of ABCA7 transcripts. Notably, three of the CpGs (cg00135882, cg22271697, and cg06169110) in the significant SNP-by-CpG interactions are either flanking or within CGIs. Active intragenic CGIs may change the major isoforms that are expressed by interfering with splicing and/or polyadenylation. Alternatively, they may promote enhancer function or act directly as an enhancer to regulate gene expression.[21] Consistent with this hypothesis, all four CpGs are located in regions that contain at least one important regulatory element (i.e., promoters, enhancers and/or CTCF binding sites). Taken together, these results suggest that SNPs and CpG sites in *ABCA7* may interact to modulate the expression and/or function of *ABCA7* transcripts, and that some of the affected transcripts may influence cognitive function in older AA.

Indeed, recent literature suggests that SNP-by-CpG interactions might be an important mechanism underlying human complex diseases.[82–84] Similar SNP-by-CpG interactions have been identified in association with complex human disorders, such as breast cancer,[85] type 2

diabetes,[86] alcohol dependence[87] and suicide attempt in schizophrenia.[88] One thing to note, though, is that SNPs could have a cis-regulatory effects on local CpGs, which could cause a spurious interaction. However, our sensitivity analysis demonstrates that the interactions we observed were not solely due to SNP-CpG correlations. In summary, we demonstrate that a complicated interplay between genetic and epigenetic risk factors in the *ABCA7* region may play an important role in cognitive function. Future studies are needed to disentangle this complicated relationship.

Our study is not without limitations. First, our gene expression measures were taken from transformed beta-lymphocytes from immortalized cell lines. While transformed beta-lymphocytes are a convenient source of DNA, the transformation process causes epigenetic changes to the immortalized cells that are not fully understood.[89] However, they provide a unique and efficient way to examine the functional effects of genetic and epigenetic variation on gene expression since the environmental conditions of the cells are the same across individuals. Second, our findings need to be replicated in a larger sample of AA. Further studies in animal and cellular models are also warranted to confirm our findings and reveal how SNPs and methylation jointly contribute to cognitive function. Finally, due to the cross-sectional nature of our study, we cannot infer causality of our findings or quantify how the SNP-by-CpG interactions alone impact cognition. To that end, longitudinal studies are necessary to investigate how cognitive function changes over time. Also, previous cis-eQTM studies in white blood cells have shown that at least some CpGs within the *ABCA7* region promote or repress gene expression of *ABCA7* and nearby genes, but we did not observe eQTM relationships with those same CpGs in our study. One reason for this may be that our methylation was measured in blood and included a mix of white blood cells, while our gene expression was measured in transformed

beta-lymphocytes. Additional work is needed to understand how *ABCA7* CpGs and their interactions with SNPs influence proximal gene expression in a variety of white blood cell types to further shed light on the complicated biological mechanisms that contribute to cognitive function. However, to our knowledge, our study is the first to take a multi-omic approach to investigate the relations between the *ABCA7* gene region and cognitive function in a population-based cohort of older adults without diagnosed dementia. Our study was also conducted in AA, an understudied population with a higher prevalence of AD[3,5] and higher conferred risk of AD from *ABCA7* compared to EA.[8] Additionally, with comprehensive cognition measures, we were able to assess associations with multiple neurocognitive domains, as well as general cognitive function.

2.6 Conclusion

In the present study, we evaluated the association between *ABCA7* genetic, epigenetic, and transcriptomic markers and cognitive function in 634 AA participants without preliminary evidence of dementia. We found that DNA methylation levels at local CpG sites modify the relationship between genetic variants and general cognitive function. Specifically, two SNPs in the *ABCA7* gene region (rs3764647 and rs115550680) may regulate the effects of methylation on cognition. Differential gene expression analysis further highlighted the potentially causal transcripts. In conclusion, our findings suggest that a complicated interplay between genetic and epigenetic factors in *ABCA7* may influence cognition in older AA without dementia.

2.7 References

1. Hampel, H., Hardy, J., Blennow, K., Chen, C., Perry, G., Kim, S.H., Villemagne, V.L., Aisen, P., Vendruscolo, M., Iwatsubo, T., et al. (2021). The Amyloid-β Pathway in Alzheimer's Disease. Mol Psychiatry *26*, 5481–5503. 10.1038/s41380-021-01249-0.

2. Grundke-Iqbal, I., Iqbal, K., Tung, Y.C., Quinlan, M., Wisniewski, H.M., and Binder, L.I. (1986). Abnormal phosphorylation of the microtubule-associated protein tau (tau) in Alzheimer cytoskeletal pathology. Proc Natl Acad Sci U S A *83*, 4913–4917.

3. 2021 Alzheimer's disease facts and figures (2021). Alzheimer's & Dementia *17*, 327–406. 10.1002/alz.12328.

4. Barnes, L.L., and Bennett, D.A. (2014). Alzheimer's disease in African Americans: risk factors and challenges for the future. Health Aff (Millwood) *33*, 580–586. 10.1377/hlthaff.2013.1353.

5. Tang, M.X., Cross, P., Andrews, H., Jacobs, D.M., Small, S., Bell, K., Merchant, C., Lantigua, R., Costa, R., Stern, Y., et al. (2001). Incidence of AD in African-Americans, Caribbean Hispanics, and Caucasians in northern Manhattan. Neurology *56*, 49–56. 10.1212/wnl.56.1.49.

6. Ridge, P.G., Mukherjee, S., Crane, P.K., Kauwe, J.S.K., and Consortium, A.D.G. (2013). Alzheimer's Disease: Analyzing the Missing Heritability. PLOS ONE *8*, e79771. 10.1371/journal.pone.0079771.

7. Migliore, L., and Coppedè, F. (2009). Genetics, environmental factors and the emerging role of epigenetics in neurodegenerative diseases. Mutat Res *667*, 82–97. 10.1016/j.mrfmmm.2008.10.011.

8. Reitz, C., Jun, G., Naj, A., and Rajbhandary, R. (2013). Variants in the ATP-Binding Cassette Transporter (ABCA7), Apolipoprotein E ϵ4,and the Risk of Late-Onset Alzheimer Disease in African Americans. JAMA *309*, 1483–1492. 10.1001/jama.2013.2973.

9. Strittmatter, W.J., Saunders, A.M., Schmechel, D., Pericak-Vance, M., Enghild, J., Salvesen, G.S., and Roses, A.D. (1993). Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. Proceedings of the National Academy of Sciences *90*, 1977–1981. 10.1073/pnas.90.5.1977.

10. Strittmatter, W.J., and Roses, A.D. (1995). Apolipoprotein E and Alzheimer disease. Proceedings of the National Academy of Sciences *92*, 4725–4727. 10.1073/pnas.92.11.4725.

11. Corder, E.H., Saunders, A.M., Strittmatter, W.J., Schmechel, D.E., Gaskell, P.C., Small, G.W., Roses, A.D., Haines, J.L., and Pericak-Vance, M.A. (1993). Gene Dose of

Apolipoprotein E Type 4 Allele and the Risk of Alzheimer's Disease in Late Onset Families. Science *261*, 921–923. 10.1126/science.8346443.

12. Lamartinière, Y., Boucau, M.-C., Dehouck, L., Krohn, M., Pahnke, J., Candela, P., Gosselet, F., and Fenart, L. (2018). ABCA7 Downregulation Modifies Cellular Cholesterol Homeostasis and Decreases Amyloid-β Peptide Efflux in an in vitro Model of the Blood-Brain Barrier. Journal of Alzheimer's Disease *64*, 1195–1211. 10.3233/JAD-170883.

13. Aikawa, T., Holm, M.-L., and Kanekiyo, T. (2018). ABCA7 and Pathogenic Pathways of Alzheimer's Disease. Brain Sci *8*. 10.3390/brainsci8020027.

14. Zhao, Q.-F., Yu, J.-T., Tan, M.-S., and Tan, L. (2015). ABCA7 in Alzheimer's Disease. Mol Neurobiol *51*, 1008–1016. 10.1007/s12035-014-8759-9.

15. Wang, N., Lan, D., Gerbod-Giannone, M., Linsel-Nitschke, P., Jehle, A.W., Chen, W., Martinez, L.O., and Tall, A.R. (2003). ATP-binding Cassette Transporter A7 (ABCA7) Binds Apolipoprotein A-I and Mediates Cellular Phospholipid but Not Cholesterol Efflux. J. Biol. Chem. *278*, 42906–42912. 10.1074/jbc.M307831200.

16. Aikawa, T., Holm, M.-L., and Kanekiyo, T. (2018). ABCA7 and Pathogenic Pathways of Alzheimer's Disease. Brain Sci *8*. 10.3390/brainsci8020027.

17. Zhao, Q.-F., Yu, J.-T., Tan, M.-S., and Tan, L. (2015). ABCA7 in Alzheimer's Disease. Mol Neurobiol *51*, 1008–1016. 10.1007/s12035-014-8759-9.

18. Logue, M.W., Lancour, D., Farrell, J., Simkina, I., Fallin, M.D., Lunetta, K.L., and Farrer, L.A. (2018). Targeted Sequencing of Alzheimer Disease Genes in African Americans Implicates Novel Risk Variants. Frontiers in Neuroscience *12*.

19. N'Songo, A., Carrasquillo, M.M., Wang, X., and Burgess, J.D. (2017). African American exome sequencing identifies potential risk variants at Alzheimer disease loci. Neurology Genetics *3*, e141. 10.1212/NXG.0000000000000141.

20. Jaenisch, R., and Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat Genet *33*, 245–254. 10.1038/ng1089.

21. Cain, J.A., Montibus, B., and Oakey, R.J. (2022). Intragenic CpG Islands and Their Impact on Gene Regulation. Frontiers in Cell and Developmental Biology *10*.

22. Li, Q.S., Sun, Y., and Wang, T. (2020). Epigenome-wide association study of Alzheimer's disease replicates 22 differentially methylated positions and 30 differentially methylated regions. Clin Epigenet *12*, 1–14. 10.1186/s13148-020-00944-z.

23. Yu, L., Chibnik, L.B., Srivastava, G.P., Pochet, N., Yang, J., Xu, J., Kozubek, J., Obholzer, N., Leurgans, S.E., Schneider, J.A., et al. (2015). Association of Brain DNA Methylation in SORL1, ABCA7, HLA-DRB5, SLC24A4, and BIN1 With Pathological Diagnosis of Alzheimer Disease. JAMA Neurology *72*, 15–24. 10.1001/jamaneurol.2014.3049.

24. De Jager, P.L., Srivastava, G., Lunnon, K., Burgess, J., Schalkwyk, L.C., Yu, L., Eaton, M.L., Keenan, B.T., Ernst, J., McCabe, C., et al. (2014). Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. Nat Neurosci *17*, 1156–1163. 10.1038/nn.3786.

25. Lunnon, K., Smith, R., Hannon, E., De Jager, P.L., Srivastava, G., Volta, M., Troakes, C., Al-Sarraj, S., Burrage, J., Macdonald, R., et al. (2014). Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. Nat Neurosci *17*, 1164–1170. 10.1038/nn.3782.

26. Yamazaki, K., Yoshino, Y., Mori, T., Yoshida, T., Ozaki, Y., Sao, T., Mori, Y., Ochi, S., Iga, J., and Ueno, S. (2017). Gene Expression and Methylation Analysis of ABCA7 in Patients with Alzheimer's Disease. Journal of Alzheimer's Disease *57*, 171–181. 10.3233/JAD-161195.

27. Apostolova, L.G., Risacher, S.L., Duran, T., Stage, E.C., Goukasian, N., West, J.D., Do, T.M., Grotts, J., Wilhalme, H., Nho, K., et al. (2018). Associations of the Top 20 Alzheimer Disease Risk Variants With Brain Amyloidosis. JAMA Neurology *75*, 328–341. 10.1001/jamaneurol.2017.4198.

28. De Roeck, A., Van Broeckhoven, C., and Sleegers, K. (2019). The role of ABCA7 in Alzheimer's disease: evidence from genomics, transcriptomics and methylomics. Acta Neuropathol *138*, 201–220. 10.1007/s00401-019-01994-1.

29. Nettiksimmons, J., Tranah, G., Evans, D.S., Yokoyama, J.S., and Yaffe, K. (2016). Gene-based aggregate SNP associations between candidate AD genes and cognitive decline. Age (Dordr) *38*, 41. 10.1007/s11357-016-9885-2.

30. Engelman, C.D., Koscik, R.L., Jonaitis, E.M., Okonkwo, O.C., Hermann, B.P., La Rue, A., and Sager, M.A. (2013). Interaction between two cholesterol metabolism genes influences memory: findings from the Wisconsin Registry for Alzheimer's Prevention. J. Alzheimers Dis. *36*, 749–757. 10.3233/JAD-130482.

31. Carrasquillo, M.M., Crook, J.E., Pedraza, O., Thomas, C.S., Pankratz, V.S., Allen, M., Nguyen, T., Malphrus, K.G., Ma, L., Bisceglio, G.D., et al. (2015). Late-onset Alzheimer's risk variants in memory decline, incident mild cognitive impairment, and Alzheimer's disease. Neurobiology of Aging *36*, 60–67. 10.1016/j.neurobiolaging.2014.07.042.

32. Cacabelos, R., and Torrellas, C. (2015). Epigenetics of Aging and Alzheimer's Disease: Implications for Pharmacogenomics and Drug Response. International Journal of Molecular Sciences *16*, 30483–30543. 10.3390/ijms161226236.

33. Daniels, P.R., Kardia, S.L.R., Hanis, C.L., Brown, C.A., Hutchinson, R., Boerwinkle, E., Turner, S.T., and Genetic Epidemiology Network of Arteriopathy study (2004). Familial aggregation of hypertension treatment and control in the Genetic Epidemiology Network of Arteriopathy (GENOA) study. Am. J. Med. *116*, 676–681. 10.1016/j.amjmed.2003.12.032.

34. Lezak, P. of N.P. and N.M.D., Lezak, M.D., Howieson, A.P. of N. and P.D.B., Howieson, D.B., Loring, P. of N.D.W., Loring, D.W., and Fischer, J.S. (2004). Neuropsychological Assessment (Oxford University Press).

35. Smith, J.A., Mosley Jr, T.H., Turner, S.T., and Kardia, S.L. (2012). Shared Genetic Effects among Measures of Cognitive Function and Leukoaraiosis. Edited by Amit Agrawal, 39.

36. Jaeger, J. (2018). Digit Symbol Substitution Test. J Clin Psychopharmacol *38*, 513–519. 10.1097/JCP.0000000000000941.

37. Davies, G., Armstrong, N., Bis, J.C., Bressler, J., Chouraki, V., Giddaluru, S., Hofer, E., Ibrahim-Verbaas, C.A., Kirin, M., Lahti, J., et al. (2015). Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N=53 949). Mol Psychiatry *20*, 183–192. 10.1038/mp.2014.188.

38. Gao, X., Jia, M., Zhang, Y., Breitling, L.P., and Brenner, H. (2015). DNA methylation changes of whole blood cells in response to active smoking exposure in adults: a systematic review of DNA methylation studies. Clin Epigenet *7*, 113. 10.1186/s13148-015-0148-3.

39. Fortin, J.-P., Fertig, E., and Hansen, K. (2014). shinyMethyl: interactive quality control of Illumina 450k DNA methylation arrays in R. 10.12688/f1000research.4680.2.

40. Xu, Z., Niu, L., Li, L., and Taylor, J.A. (2016). ENmix: a novel background correction method for Illumina HumanMethylation450 BeadChip. Nucleic Acids Research *44*, e20–e20. 10.1093/nar/gkv907.

41. Fortin, J.-P., Triche, T.J., Jr, and Hansen, K.D. (2017). Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. Bioinformatics *33*, 558–560. 10.1093/bioinformatics/btw691.

42. Aryee, M.J., Jaffe, A.E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A.P., Hansen, K.D., and Irizarry, R.A. (2014). Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics *30*, 1363–1369. 10.1093/bioinformatics/btu049.

43. Niu, L., Xu, Z., and Taylor, J.A. (2016). RCP: a novel probe design bias correction method for Illumina Methylation BeadChip. Bioinformatics *32*, 2659–2663. 10.1093/bioinformatics/btw285.

44. Lehne, B., Drong, A.W., Loh, M., Zhang, W., Scott, W.R., Tan, S.-T., Afzal, U., Scott, J., Jarvelin, M.-R., Elliott, P., et al. (2015). A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. Genome Biology *16*, 37. 10.1186/s13059-015-0600-x.

45. Hansen, K. IlluminaHumanMethylationEPICanno.ilm10b2.hg19: Annotation for Illumina's EPIC methylation arrays. R package version 0.6.0.

46. Houseman, E.A., Accomando, W.P., Koestler, D.C., Christensen, B.C., Marsit, C.J., Nelson, H.H., Wiencke, J.K., and Kelsey, K.T. (2012). DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics *13*, 86. 10.1186/1471-2105-13-86.

47. Du, P., Zhang, X., Huang, C.-C., Jafari, N., Kibbe, W.A., Hou, L., and Lin, S.M. (2010). Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics *11*, 587. 10.1186/1471-2105-11-587.

48. Weisenberger, C.D.J., and Laird, P.W. Comprehensive DNA Methylation Analysis on the Illumina® Infinium® Assay Platform. 4.

49. Lockstone, H.E. (2011). Exon array data analysis using Affymetrix power tools and R statistical software. Briefings in Bioinformatics *12*, 634–644. 10.1093/bib/bbq086.

50. Irizarry, R.A., Bolstad, B.M., Collin, F., Cope, L.M., Hobbs, B., and Speed, T.P. (2003). Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Research *31*, e15. 10.1093/nar/gng015.

51. Dai, M., Wang, P., Boyd, A.D., Kostov, G., Athey, B., Jones, E.G., Bunney, W.E., Myers, R.M., Speed, T.P., Akil, H., et al. (2005). Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. Nucleic Acids Research *33*, e175. 10.1093/nar/gni179.

52. Saha, A., and Battle, A. (2018). False positives in trans-eQTL and co-expression analyses arising from RNA-sequencing alignment errors. 10.12688/f1000research.17145.1.

53. Johnson, W.E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics *8*, 118–127. 10.1093/biostatistics/kxj037.

54. Martin, T.C., Hardiman, T., Yet, I., Tsai, P.-C., and Bell, J.T. (2022). coMET: coMET: visualisation of regional epigenome-wide association scan (EWAS) results and DNA co-methylation patterns. 10.18129/B9.bioc.coMET.

55. BioRender https://app.biorender.com/illustrations/634485ef72e6ef474cc65999.

56. Lenth, R.V., Buerkner, P., Herve, M., Love, J., Riebl, H., and Singmann, H. (2021). emmeans: Estimated Marginal Means, aka Least-Squares Means.

57. Fox, J., Weisberg, S., Price, B., Friendly, M., Hong, J., Andersen, R., Firth, D., Taylor, S., and R Core Team (2020). effects: Effect Displays for Linear, Generalized Linear, and Other Models.

58. ggplot2 package - RDocumentation https://www.rdocumentation.org/packages/ggplot2/versions/3.3.5.

59. THE GTEX CONSORTIUM (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science *369*, 1318–1330. 10.1126/science.aaz1776.

60. Gene: ABCA7 (ENSG00000064687) - Summary - Homo_sapiens - Ensembl genome browser 107 https://useast.ensembl.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG00000064687;r=19:1039997-1065572.

61. Komaki, S., Shiwa, Y., Furukawa, R., Hachiya, T., Ohmomo, H., Otomo, R., Satoh, M., Hitomi, J., Sobue, K., Sasaki, M., et al. (2018). iMETHYL: an integrative database of human DNA methylation, gene expression, and genomic variation. Hum Genome Var *5*, 1–4. 10.1038/hgv.2018.8.

62. Hachiya, T., Furukawa, R., Shiwa, Y., Ohmomo, H., Ono, K., Katsuoka, F., Nagasaki, M., Yasuda, J., Fuse, N., Kinoshita, K., et al. (2017). Genome-wide identification of inter-individually variable DNA methylation sites improves the efficacy of epigenetic association studies. npj Genomic Med *2*, 1–14. 10.1038/s41525-017-0016-5.

63. R: The R Project for Statistical Computing https://www.r-project.org/.

64. Naj, A.C., Jun, G., Beecham, G.W., Wang, L.-S., Vardarajan, B.N., Buros, J., Gallins, P.J., Buxbaum, J.D., Jarvik, G.P., Crane, P.K., et al. (2011). Common variants in MS4A4/MS4A6E, CD2uAP, CD33, and EPHA1 are associated with late-onset Alzheimer's disease. Nat Genet *43*, 436–441. 10.1038/ng.801.

65. Vasquez, J.B., Fardo, D.W., and Estus, S. (2013). ABCA7 expression is associated with Alzheimer's disease polymorphism and disease status. Neurosci Lett *556*, 58–62. 10.1016/j.neulet.2013.09.058.

66. Steinberg, S., Stefansson, H., Jonsson, T., Johannsdottir, H., Ingason, A., Helgason, H., Sulem, P., Magnusson, O.T., Gudjonsson, S.A., Unnsteinsdottir, U., et al. (2015). Loss-of-function variants in ABCA7 confer risk of Alzheimer's disease. Nat Genet *47*, 445–447. 10.1038/ng.3246.

67. De Roeck, A., Duchateau, L., Van Dongen, J., Cacace, R., Bjerke, M., Van den Bossche, T., Cras, P., Vandenberghe, R., De Deyn, P.P., Engelborghs, S., et al. (2018). An intronic VNTR affects splicing of ABCA7 and increases risk of Alzheimer's disease. Acta Neuropathol *135*, 827–837. 10.1007/s00401-018-1841-z.

68. De Roeck, A., Van den Bossche, T., van der Zee, J., Verheijen, J., De Coster, W., Van Dongen, J., Dillen, L., Baradaran-Heravi, Y., Heeman, B., Sanchez-Valle, R., et al. (2017). Deleterious ABCA7 mutations and transcript rescue mechanisms in early onset Alzheimer's disease. Acta Neuropathol *134*, 475–487. 10.1007/s00401-017-1714-x.

69. Hof, P.R., Glannakopoulos, P., and Bouras, C. (1996). The neuropathological changes associated with normal brain aging. Histol Histopathol *11*, 1075–1088.

70. Perl, D.P. (2010). Neuropathology of Alzheimer's Disease. Mt Sinai J Med *77*, 32–42. 10.1002/msj.20157.

71. Vivot, A., Glymour, M.M., Tzourio, C., Amouyel, P., Chêne, G., and Dufouil, C. (2015). Association of Alzheimer's related genotypes with cognitive decline in multiple domains: results from the Three-City Dijon study. Molecular Psychiatry *20*, 1173–1178. 10.1038/mp.2015.62.

72. Smith, J.A., Kho, M., Zhao, W., Yu, M., Mitchell, C., and Faul, J.D. (2021). Genetic effects and gene-by-education interactions on episodic memory performance and decline in an aging population. Social Science & Medicine *271*, 112039. 10.1016/j.socscimed.2018.11.019.

73. Davies, G., Lam, M., Harris, S.E., Trampush, J.W., Luciano, M., Hill, W.D., Hagenaars, S.P., Ritchie, S.J., Marioni, R.E., Fawns-Ritchie, C., et al. (2018). Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. Nat Commun *9*, 2098. 10.1038/s41467-018-04362-x.

74. Dumitrescu, L., Mahoney, E.R., Mukherjee, S., Lee, M.L., Bush, W.S., Engelman, C.D., Lu, Q., Fardo, D.W., Trittschuh, E.H., Mez, J., et al. (2020). Genetic variants and functional pathways associated with resilience to Alzheimer's disease. Brain *143*, 2561–2575. 10.1093/brain/awaa209.

75. Marioni, R.E., McRae, A.F., Bressler, J., Colicino, E., Hannon, E., Li, S., Prada, D., Smith, J.A., Trevisi, L., Tsai, P.-C., et al. (2018). Meta-analysis of epigenome-wide association studies of cognitive abilities. Mol Psychiatry *23*, 2133–2144. 10.1038/s41380-017-0008-y.

76. Yu, L., Chibnik, L.B., Yang, J., McCabe, C., Xu, J., Schneider, J.A., De Jager, P.L., and Bennett, D.A. (2016). Methylation profiles in peripheral blood CD4+ lymphocytes versus brain: The relation to Alzheimer's disease pathology. Alzheimers Dement *12*, 942–951. 10.1016/j.jalz.2016.02.009.

77. Ashraf-ganjouei, A., Moradi, K., Bagheri, S., and Aarabi, M.H. (2020). The association between systemic inflammation and cognitive performance in healthy adults. Journal of Neuroimmunology *345*, 577272. 10.1016/j.jneuroim.2020.577272.

78. Weinstein, G., Lutski, M., Goldbourt, U., and Tanne, D. (2017). C-reactive protein is related to future cognitive impairment and decline in elderly individuals with cardiovascular disease. Archives of Gerontology and Geriatrics *69*, 31–37. 10.1016/j.archger.2016.11.002.

79. Cukier, H.N., Kunkle, B.W., Vardarajan, B.N., Rolati, S., Hamilton-Nelson, K.L., Kohli, M.A., Whitehead, P.L., Dombroski, B.A., Booven, D.V., Lang, R., et al. (2016). ABCA7 frameshift deletion associated with Alzheimer disease in African Americans. Neurology Genetics *2*. 10.1212/NXG.0000000000000079.

80. Raj, T., Li, Y.I., Wong, G., Humphrey, J., Wang, M., Ramdhani, S., Wang, Y.-C., Ng, B., Gupta, I., Haroutunian, V., et al. (2018). Integrative transcriptome analyses of the aging

brain implicate altered splicing in Alzheimer's disease susceptibility. Nat Genet *50*, 1584–1592. 10.1038/s41588-018-0238-1.

81. Humphries, C., Kohli, M.A., Whitehead, P., Mash, D.C., Pericak-Vance, M.A., and Gilbert, J. (2015). Alzheimer disease (AD) specific transcription, DNA methylation and splicing in twenty AD associated loci. Molecular and Cellular Neuroscience *67*, 37–45. 10.1016/j.mcn.2015.05.003.

82. Tsuboi, K., Nagatomo, T., Gohno, T., Higuchi, T., Sasaki, S., Fujiki, N., Kurosumi, M., Takei, H., Yamaguchi, Y., Niwa, T., et al. (2017). Single CpG site methylation controls estrogen receptor gene transcription and correlates with hormone therapy resistance. J Steroid Biochem Mol Biol *171*, 209–217. 10.1016/j.jsbmb.2017.04.001.

83. Qiu, C., Shen, H., Fu, X., Xu, C., and Deng, H. (2018). Meta-Analysis of Genome-Wide Association Studies Identifies Novel Functional CpG-SNPs Associated with Bone Mineral Density at Lumbar Spine. Int J Genomics *2018*, 6407257. 10.1155/2018/6407257.

84. Gertz, J., Varley, K.E., Reddy, T.E., Bowling, K.M., Pauli, F., Parker, S.L., Kucera, K.S., Willard, H.F., and Myers, R.M. (2011). Analysis of DNA Methylation in a Three-Generation Family Reveals Widespread Genetic Influence on Epigenetic Regulation. PLoS Genet *7*, e1002228. 10.1371/journal.pgen.1002228.

85. Harlid, S., Ivarsson, M.I.L., Butt, S., Hussain, S., Grzybowska, E., Eyfjörd, J.E., Lenner, P., Försti, A., Hemminki, K., Manjer, J., et al. (2011). A candidate CpG SNP approach identifies a breast cancer associated ESR1-SNP. Int J Cancer *129*, 1689–1698. 10.1002/ijc.25786.

86. Dayeh, T.A., Olsson, A.H., Volkov, P., Almgren, P., Rönn, T., and Ling, C. (2013). Identification of CpG-SNPs associated with type 2 diabetes and differential DNA methylation in human pancreatic islets. Diabetologia *56*, 1036–1046. 10.1007/s00125-012-2815-7.

87. Taqi, M.M., Bazov, I., Watanabe, H., Sheedy, D., Harper, C., Alkass, K., Druid, H., Wentzel, P., Nyberg, F., Yakovleva, T., et al. (2011). Prodynorphin CpG-SNPs associated with alcohol dependence: elevated methylation in the brain of human alcoholics. Addict Biol *16*, 499–509. 10.1111/j.1369-1600.2011.00323.x.

88. Bani-Fatemi, A., Gonçalves, V.F., Zai, C., de Souza, R., Le Foll, B., Kennedy, J.L., Wong, A.H., and De Luca, V. (2013). Analysis of CpG SNPs in 34 genes: association test with suicide attempt in schizophrenia. Schizophr Res *147*, 262–268. 10.1016/j.schres.2013.04.018.

89. Fridman, A.L., and Tainsky, M.A. (2008). Critical pathways in cellular senescence and immortalization revealed by gene expression profiling. Oncogene *27*, 5975–5987. 10.1038/onc.2008.213.

2.8 Tables

Table 2-1. Sample characteristics of Genetic Epidemiology Network of Arteriopathy (GENOA) African Americans (N=634).

	Mean (SD) or N%
Age at cognition measurement (years)	63.31 (8.08)
Age difference between methylation and cognition measurements (years)[a]	6.03 (1.29)
Sex	
Female	475 (74.90%)
Male	159 (25.10%)
Educational attainment	
At least some college	300 (47.32%)
High school degree/GED	169 (26.66%)
Less than High School degree/GED	165 (26.03%)
Smoking Status	
Current Smoker	105 (16.56%)
Former Smoker	146 (23.03%)
Never Smoker	383 (60.41%)
General cognitive function	0.00 (1.00)
Delayed recall (RAVLT, number of words recalled)	7.05 (3.34)
Processing speed (DSST, number of symbols)	34.44 (12.62)
Word fluency (COWA-FAS, number of words)	29.73 (11.61)
Concentration effectiveness (SCWT, number of items)	22.53 (9.83)
Visual conceptual tracking (TMTA, seconds to test completion)	61.63 (31.96)

Abbreviations: HS, High School; RAVLT, Rey Auditory Verbal Learning Test; DSST, Digit Symbol Substitution Test; COWA-FAS, Controlled Oral Word Association Test; SCWT, Stroop Color-Word Test; TMTA, Trail Making Test A.

a. Subset sample (n=494) consists of subjects with available genotype and methylation data

Table 2-2. Interaction of *ABCA7* sentinel SNPs and CpG sites on general cognitive function (FDR q<0.1; N=494).

	SNP annotation				CpG site annotation				Main effects				Interaction	
SNP * CpG site Interaction	SNP	Position	Risk allele	RAF	CpG site	Position	Site Type	Relation to CpG Island	β_{SNP}	P-value	β_{CpG}	P-value	$\beta_{interaction}$	P-value
rs3764647 * cg00135882	rs3764647	1044712	G	0.20	cg00135882	1065783	Promoter	North Shore	-0.01	0.875	0.24	0.086	-0.80	1.46E-04**
rs3764647 * cg22271697	rs3764647	1044712	G	0.20	cg22271697	1042537	Promoter	North Shelf	-0.07	0.319	0.16	7.23E-06*	-0.18	5.77E-04**
rs115550680 * cg06169110	rs115550680	1050420	G	0.06	cg06169110	1046615	Gene Body	CG Island	-0.23	0.045*	0.06	0.143	-0.38	2.18E-04**
rs115550680 * cg17316918	rs115550680	1050420	G	0.06	cg17316918	1056930	Gene Body	Open Sea	-0.05	0.661	-0.06	0.164	0.41	4.84E-04**

Abbreviations: AA, African American; EA, European American; RAF, risk allele frequency in GENOA

Model 4: General cognitive function ~ SNP + CpG + SNP*CpG + age at cognitive testing + age difference between methylation and cognition measurements + sex + educational attainment + *APOE* ε2 + *APOE* ε4 + smoking status + PC1-4 + familial relatedness (random effect)

* p<0.05, ** q<0.1 (FDR-corrected significance level)

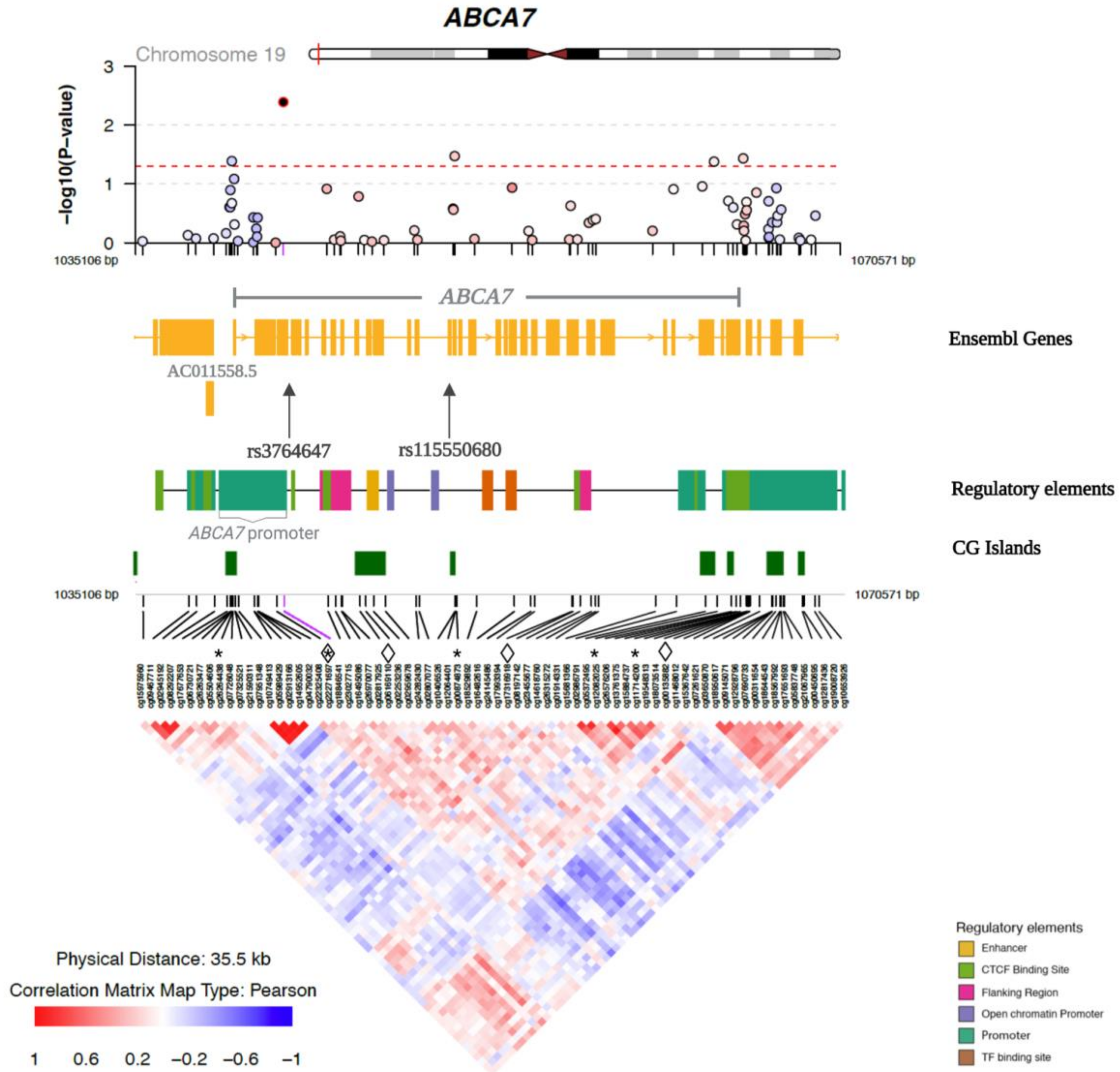
Table 2-3. Estimated effect of CpG site on general cognitive function for given *ABCA7* SNP genotype group (N=494).

SNP	CpG site	Genotype	β_{CpG}	P-value
rs3764647 [a]	cg00135882	AA	0.09	0.566
		GG/AG	-0.68	0.004*
rs3764647 [a]	cg22271697	AA	0.14	2.00E-04*
		GG/AG	-0.02	0.719
rs115550680 [b]	cg06169110	AA	0.05	0.221
		GG/AG	-0.37	2.00E-04*
rs115550680 [b]	cg17316918	AA	-0.06	0.202
		GG/AG	0.33	0.004*

a. GG (N=17) and AG (N=156) groups were combined in the GG/AG group (N=173).
b. GG (N=5) and AG (N=54) groups were combined in the GG/AG group (N=59).
Model 4: General cognitive function~ SNP + CpG + SNP*CpG + age at cognition measurement + age difference between methylation and cognition measurements + sex + educational status + *APOE* ε2 + *APOE* ε4 + smoking status + PC1-4 + familial relatedness (random effect)
* $p<0.05$

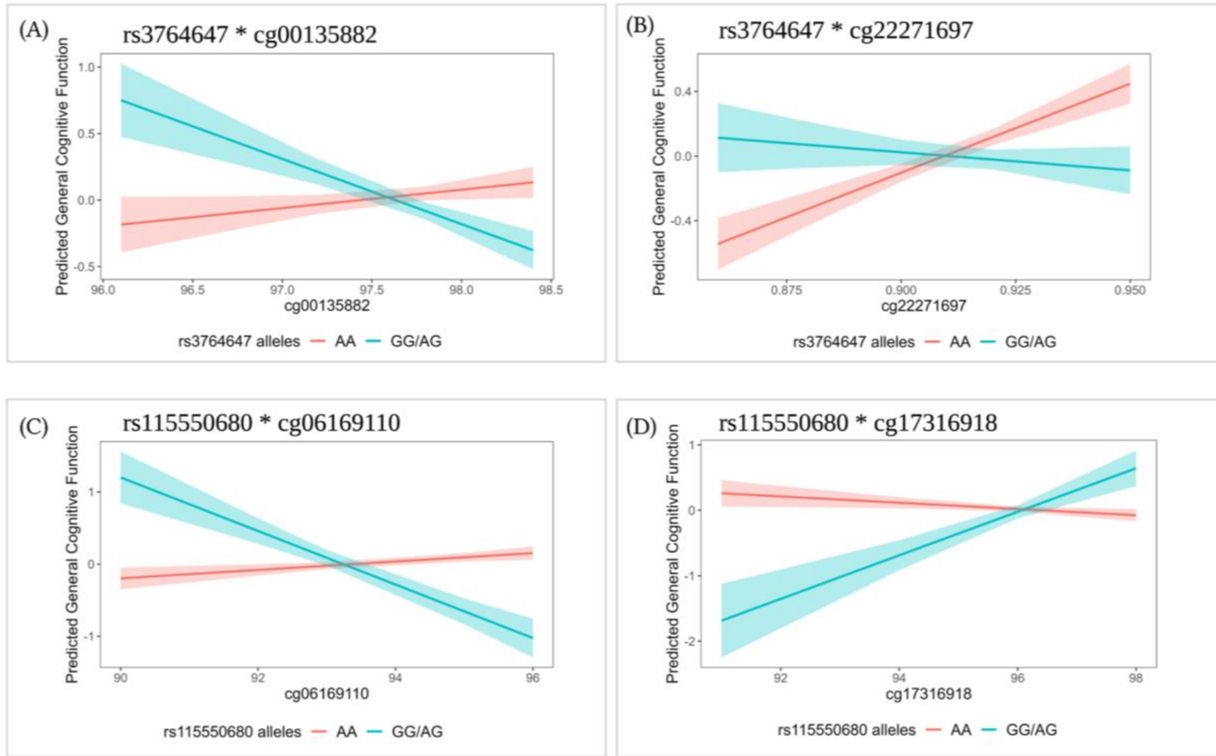
2.9 Figures

Figure 2-1. Regional plot of the association between DNA methylation in the ABCA7 region and general cognitive function.



The top panel shows -log10 (P value) for the association between methylation and general cognitive function, adjusting for age, sex, age difference between methylation and cognition measurements, educational attainment, APOE ε2, APOE ε4, smoking status, PC1-4, and familial relatedness (random effects; Model 3), according to chromosomal positions. Nominally significant (P<0.05) associations are above the dashed line. The middle panels show Ensembl genes, regulatory elements, and CpG islands (UCSC Genome Browser) in the ABCA7 region. The lower panel shows the correlations in the DNA methylation levels among the 72 CpG sites in this region. The five CpGs that have a nominal association with general cognitive function are marked by asterisks. The four CpGs and two intronic SNPs that were identified in the SNP-by-CpG interactions associated with general cognitive function are marked by diamond symbols (CpGs) and arrows (SNPs).

Figure 2-2. Linear prediction of CpG sites (% methylated) on general cognitive function for a given SNP genotype group in the *ABCA7* region.



Linear prediction of CpG sites (% methylated) on general cognitive function for a given SNP genotype group in the *ABCA7* region: (A) rs3764647*cg00135882, (B) rs3764647*cg22271697, (C) rs115550680*cg06169110, and (D) rs115550680*cg17316918. Models were adjusted for age, sex, age difference between methylation measurement and cognition measurement, educational attainment, *APOE* ε2, *APOE* ε4, smoking status, PC1-4, and familial relatedness as a random effect (Model 4). Regression lines are shown with standard error bands. For rs3764647, GG (N=17) and AG (N=156) groups were combined in the GG/AG group (N=173). For rs115550680, GG (N=5) and AG (N=54) groups were combined in the GG/AG group (N=59).

2.10 Supplementary Material

Table S2-4. Pearson's correlations among the six cognitive measures (n=634)

	RAVLT	DSST	COWA-FAS	SCWT	TMTA	General cognitive function
RAVLT	1.000					
DSST	0.365***	1.000				
COWA-FAS	0.248***	0.516***	1.000			
SCWT	0.251***	0.516***	0.336***	1.000		
TMTA	0.241***	0.663***	0.419***	0.432***	1.000	
General cognitive function	0.522***	0.874***	0.698***	0.704***	0.791***	1.000

Abbreviations: RAVLT, Rey Auditory Verbal Learning Test; DSST, Digit Symbol Substitution Task; COWA-FAS, Controlled Oral Word Association Test; SCWT, Stroop Color-Word Test; TMTA, Trail Making Test A

* $p<0.05$, **$p<0.01$, ***$p<0.001$

Table S2-5. Pearson's correlations among the five sentinel *ABCA7* SNPs (n=634)

	rs3764647	rs3764650	rs115550680	rs3752246	rs4147929
rs3764647	1.000				
rs3764650	0.843***	1.000			
rs115550680	-0.117**	-0.141***	1.000		
rs3752246	-0.139***	-0.004	-0.101*	1.000	
rs4147929	-0.140***	-0.026	-0.110**	0.956***	1.000

* $p<0.05$, ** $p<0.01$, *** $p<0.001$

Table S2-6. Association between *ABCA7* sentinel SNPs and general cognitive function (n=634)

PMID [a]	Ancestry [b]	SNP	Chr	Position	RA	RAF	Model 1		Model 2		Model 3	
							β_{SNP}	*P-value*	β_{SNP}	*P-value*	β_{SNP}	*P-value*
28480329	AA	rs3764647	19	1044712	G	0.20	-0.04	0.518	-0.02	0.786	-0.01	0.823
21460840	EA	rs3764650	19	1046520	G	0.25	-0.03	0.598	-0.02	0.768	-0.02	0.716
23571587	AA	rs115550680	19	1050420	G	0.06	-0.03	0.748	-0.01	0.928	-0.01	0.884
21460841	EA	rs3752246	19	1056492	G	0.04	0.21	0.088	0.15	0.180	0.15	0.186
24162737	EA	rs4147929	19	1063443	A	0.05	0.21	0.075	0.12	0.243	0.12	0.241
-	-	*APOE* ε2	19	45411941	T	0.12	0.07	0.317	0.11	0.087	-	-
-	-	*APOE* ε4	19	45412079	C	0.23	-0.11	0.046*	-0.12	0.022*	-	-

Abbreviations: PMID, Pubmed ID; AA, African American; EA, European American; Chr, chromosome; RA, risk allele; RAF, risk allele frequency in GENOA

a. Pubmed ID numbers for studies that identified sentinel SNPs in the *ABCA7* region in association with Alzheimer's disease.

b. Ancestry of cohorts in which significant associations were identified between sentinel SNPs in the *ABCA7* region and Alzheimer's disease

Model 1: General cognitive function ~ SNP + age at cognition measurement + sex + PC1-4 + familial relatedness (random effect)

Model 2: Model 1 + educational attainment

Model 3: Model 2 + *APOE ε2* + *APOE ε4*

*p<0.05; no associations were significant after Bonferroni correction at α = 0.05/5 = 0.01.

Table S2-7. Association of CpGs in the *ABCA7* region and general cognitive function (p<0.05; n=494)

CpG site	Position	Site Type	Relation to CpG Island	Model 1		Model 2		Model 3	
				β_{CpG}	*P*	β_{CpG}	*P*	β_{CpG}	*P*
cg22271697	1042537	Gene Body	North Shelf	0.08	0.009*	0.07	0.007*	0.08	0.004*
cg00874873	1051161	Gene Body	CG Island	0.12	0.074	0.13	0.025*	0.12	0.034*
cg11714200	1065689	Promoter	North Shore	0.06	0.101	0.08	0.030*	0.07	0.037*
cg26264438	1039942	Promoter	CG Island	0.53	0.236	0.84	0.039*	0.83	0.041*
cg12082025	1064219	Gene Body	CG Island	0.05	0.394	0.11	0.047*	0.11	0.042*
cg18644543	1067356	1st Exon; 5' UTR	CG Island	-0.51	0.031*	-0.33	0.132	-0.34	0.118

Model 1: General cognitive function ~ CpG site + sex + age at cognition measurement + age difference between methylation and cognition measurements + smoking status+ PC1-4 + familial relatedness (random effect)
Model 2: Model 1 + educational attainment
Model 3: Model 2 + *APOE* ε2 + *APOE* ε4
*$p<0.05$; No associations are significant at FDR q<0.1

Table S2-8. Interaction between *ABCA7* sentinel SNPs and CpG sites on general cognitive function (p<0.05; n=494)

SNP * CpG site Interaction	CpG position	Main effects				Interaction	
		β_{SNP}	*p*-value	β_{CpG}	*p*-value	$\beta_{interaction}$	*p*-value
rs3764647 * cg00135882	1065783	-0.01	0.875	0.24	0.086	-0.80	1.46×10^{-4} **
rs115550680 * cg06169110	1046615	-0.23	0.045*	0.06	0.143	-0.38	2.18×10^{-4}**
rs115550680 * cg17316918	1056930	-0.05	0.661	-0.06	0.164	0.41	4.84×10^{-4} **
rs3764647 * cg22271697	1042537	-0.07	0.319	0.16	7.23×10^{-6}*	-0.18	5.77×10^{-4} **
rs115550680 * cg05372495	1063625	-0.04	0.707	4.92×10^{-3}	0.837	0.17	0.008*
rs115550680 * cg02913166	1041178	-0.10	0.329	-0.02	0.244	0.17	0.010*
rs115550680 * cg09467711	1037732	-0.26	0.049*	-0.01	0.632	0.10	0.011*
rs115550680 * cg12817436	1068561	-0.01	0.961	-0.02	0.376	0.20	0.011*
rs115550680 * cg07726048	1039944	-0.03	0.799	0.30	0.031*	-1.02	0.012*
rs115550680 * cg07690733	1066986	-0.11	0.301	-0.14	0.60	2.27	0.014*
rs115550680 * cg07325521	1040062	-0.02	0.872	-0.12	0.627	-1.50	0.015*
rs3764647 * cg09467711	1037732	-0.07	0.315	0.03	0.109	-0.07	0.017*
rs3752246 * cg06169110	1046615	4.94×10^{-3}	0.967	-0.02	0.597	0.24	0.033*

Model 4: General cognitive function ~ SNP + CpG + SNP*CpG + age at cognition measurement + age difference between methylation and cognition measurements + sex + educational attainment + *APOE ε2* + *APOE ε4* + smoking status + PC1-4 + familial relatedness (random effect)

* *p*<0.05; ** FDR q<0.1

Table S2-9. Pearson's correlations among *ABCA7* CpG sites[a] (n=494)

	cg00135882	cg22271697	cg06169110	cg17316918	cg00874873	cg11714200	cg26264438	cg12082025	cg18644543
cg00135882	1.000								
cg22271697	0.243***	1.000							
cg06169110	0.273***	0.085	1.000						
cg17316918	0.114*	0.152***	-0.051	1.000					
cg00874873	0.056	0.166***	0.037	0.216***	1.000				
cg11714200	0.128**	0.173***	-0.039	0.213***	0.139**	1.000			
cg26264438	-0.291***	-0.119**	-0.105*	-0.259***	-0.091*	-0.104*	1.000		
cg12082025	0.400***	0.121**	0.217***	0.151***	0.043	0.070	-0.223***	1.000	
cg18644543	-0.220***	-0.130**	0.041	-0.380***	-0.235***	-0.152***	0.407***	-0.125**	1.000

a. CpG sites in this correlation matrix were chosen from Tables 2 and S4. Cg00135882, cg22271697, cg06169110 and cg17316918 are significant CpG sites in the SNP-by-CpG interactions on general cognitive function (FDR q<0.1; Table 2). Cg22271697, cg00874873, cg11714200, cg26264438, cg12082025 and cg18644543 are nominally associated with general cognitive function ($p<0.05$; Table S4).

* $p<0.05$, **$p<0.01$, ***$p<0.001$

Table S2-10. Estimated effect of CpG site on general cognitive function for given *ABCA7* SNP genotype group, after excluding outlying values for CpG sites[a]

SNP	CpG site	Genotype	β_{CpG}	*p*-value
rs3764647	cg00135882	AA	0.14	0.311
		GG/AG	-0.49	0.005*
rs3764647	cg22271697	AA	0.14	1.00×10^{-4} *
		GG/AG	-0.02	0.719
rs115550680	cg06169110	AA	0.06	0.130
		GG/AG	-0.37	2.00×10^{-4} *
rs115550680	cg17316918	AA	-0.05	0.238
		GG/AG	0.33	0.004*

a. Outliers greater or less than 4 standard deviations were excluded: 4 values were excluded for cg00135882 (n = 490), 2 values were excluded for cg22271697 (n = 492) and cg17316918 (n = 492), and 1 value was excluded for cg06169110 (n = 493)
Model 4: General cognitive function~ SNP + CpG + SNP*CpG + age at cognition measurement + age difference between methylation and cognition measurements + sex + educational status + *APOE* ε2 + *APOE* ε4 + smoking status + PC1-4 + familial relatedness (random effect)
* $p<0.05$

Table S2-11. Estimated effect of CpG[a] site on general cognitive function for given *ABCA7* SNP genotype group, after adjusting for SNP effect.

SNP	CpG site	Genotype	β_{CpG}	*p*-value
rs3764647	cg22271697	AA	0.15	1.00×10^{-4} *
		GG/AG	-0.02	0.571
rs115550680	cg06169110	AA	0.06	0.120
		GG/AG	-0.37	2.00×10^{-4} *

a. Sensitivity analysis was conducted on identified SNP-by-CpG interactions from Table 2 whose CpGs were associated with their corresponding SNPs ($p<0.05$). The SNP effect was adjusted out of the CpG site effect, and the interaction analysis was conducted using the adjusted CpG value

Model 4: General cognitive function~ SNP + CpG + SNP*CpG + age at cognition measurement + age difference between methylation and cognition measurements + sex + educational status + *APOE* ε2 + *APOE* ε4 + smoking status + PC1-4 + familial relatedness (random effect)
* $p<0.05$

Table S2-12. Interaction between *ABCA7* sentinel SNPs and CpG sites on neurocognitive measurements (n=494)

DSST	Main effects				Interaction	
	β_{SNP}	*p*-value	β_{CpG}	*p*-value	$\beta_{interaction}$	*p*-value
rs3764647 * cg00135882	-0.35	0.679	0.68	0.709	-7.73	0.005*
rs3764647 * cg22271697	-0.88	0.307	1.22	0.008*	-1.37	0.047*
rs115550680 * cg06169110	-2.19	0.145	0.49	0.340	-4.24	0.002*
rs115550680 * cg17316918	-0.07	0.959	-0.54	0.319	3.38	0.028*

COWA-FAS	Main effects				Interaction	
	β_{SNP}	*p*-value	β_{CpG}	*p*-value	$\beta_{interaction}$	*p*-value
rs3764647 * cg00135882	-0.20	0.828	2.85	0.143	-6.79	0.023*
rs3764647 * cg22271697	-0.64	0.488	1.15	0.021*	-1.04	0.158
rs115550680 * cg06169110	-1.67	0.300	0.01	0.978	-1.76	0.219
rs115550680 * cg17316918	-0.59	0.684	0.07	0.905	3.56	0.030*

RAVLT	Main effects				Interaction	
	β_{SNP}	*p*-value	β_{CpG}	*p*-value	$\beta_{interaction}$	*p*-value
rs3764647 * cg00135882	0.53	0.055	0.19	0.747	-0.87	0.346
rs3764647 * cg22271697	0.50	0.070	0.27	0.075	-0.48	0.036*
rs115550680 * cg06169110	-0.32	0.511	0.07	0.688	-0.71	0.107
rs115550680 * cg17316918	0.09	0.831	-0.06	0.737	1.20	0.017*

SCWT	Main effects				Interaction	
	β_{SNP}	*p*-value	β_{CpG}	*p*-value	$\beta_{interaction}$	*p*-value
rs3764647 * cg00135882	-0.55	0.498	2.92	0.089	-7.68	0.004*
rs3764647 * cg22271697	-1.06	0.187	1.68	1.21×10^{-4} *	-1.79	0.006*
rs115550680 * cg06169110	-2.70	0.058	0.93	0.056	-3.29	0.009*
rs115550680 * cg17316918	-1.23	0.340	-0.89	0.083	3.52	0.016*

TMTA	Main effects				Interaction	
	β_{SNP}	*p*-value	β_{CpG}	*p*-value	$\beta_{interaction}$	*p*-value
rs3764647 * cg00135882	-0.03	0.333	0.05	0.484	-0.23	0.043*
rs3764647 * cg22271697	-0.05	0.143	0.06	0.002*	-0.07	0.020*
rs115550680 * cg06169110	-0.08	0.187	0.02	0.423	-0.15	0.006*
rs115550680 * cg17316918	-0.01	0.903	-0.02	0.272	0.11	0.089

Key: DSST, Digit Symbol Substitution Task; COWA-FAS, Controlled Oral Word Association Test; RAVLT, Rey Auditory Verbal Learning Test; SCWT, Stroop Color-Word Test; TMTA, Trail Making Test A

Model 4: Cognitive test score ~ SNP+ CpG + SNP*CpG + age at cognition measurement + age difference between methylation and cognition measurements + sex + educational attainment + *APOE* ε2 + *APOE* ε4 + smoking status + PC1-4 + familial relatedness (random effect)

* $p<0.05$

Table S2-13. Interaction between *ABCA7* sentinel SNPs and CpG sites[a] on transcripts in the ABCA7 gene region (p<0.05; n = 429)

		Main effects				Interaction	
Transcript	**SNP * CpG site Interaction**	β_{SNP}	***p*-value**	β_{CpG}	***p*-value**	$\beta_{interaction}$	***p*-value**
ENST00000525939	rs115550680 * cg17316918	0.03	0.428	-9.82×10^{-3}	0.493	0.09	0.026*
ENST00000531467	rs3764647 * cg22271697	0.03	0.085	-0.012	0.270	0.03	0.046*

Model 5: Transcript ~ SNP + CpG + SNP*CpG + age at gene expression measurement + age difference between methylation and gene expression measurements + sex + PC1-4 + familial relatedness (random effect)

a. Significant SNP-by-CpG interactions in Table 2
* $p<0.05$; No associations are significant at FDR q<0.1

Table S2-14. Estimated effect of CpG site on *ABCA7* transcripts for given *ABCA7* SNP genotype group (n=429)

Transcript [a]	SNP	CpG site	Genotype	β_{CpG}	*p*-value
ENST00000531467	rs3764647 [b]	cg22271697	AA	-0.01	0.319
			GG/AG	0.02	0.120
ENST00000525939	rs115550680 [c]	cg17316918	AA	-7.6×10^{-3}	0.597
			GG/AG	0.07	0.054

a. Transcripts associated with previously identified SNP-by-CpG interactions in Table S10
b. GG (n = 15) and AG (n = 156) groups were combined in the GG/AG group (n = 151)
c. GG (n = 3) and AG (n = 47) were combined in the GG/AG group (n = 50)

Model 5: Transcript ~ SNP + CpG + SNP*CpG + age at gene expression measurement + age difference between methylation and gene expression measurements + sex + PC1-4 + familial relatedness (random effect)

No associations are significant at $p<0.05$

Table S2-15. Association of SNPs[a] on transcripts in the *ABCA7* gene region (p<0.05; n=429)

Transcript	SNP	β_{SNP}	*p*-value
ENST00000531467	rs115550680	-0.13	3.17×10^{-5}**
ENST00000527496	rs115550680	-0.13	2.14×10^{-4}**
ENST00000529442	rs115550680	-0.10	5.07×10^{-4}**
ENST00000524850	rs115550680	-0.09	0.001**
ENST00000526885	rs115550680	-0.06	0.008**
ENST00000532194	rs115550680	-0.07	0.009**
ENST00000433129	rs115550680	-0.06	0.012**
ENST00000525238	rs115550680	-0.06	0.012**
ENST00000263094	rs115550680	-0.05	0.015**
ENST00000530703	rs115550680	-0.06	0.024**
ENST00000435683	rs115550680	-0.05	0.026**
ENST00000530703	rs3764647	0.03	0.037*

Model 5: Transcript ~ SNP + age at gene expression measurement + sex + PC1-4 + familial relatedness (random effect)
a. SNPs shown were previously significant in the SNP-by-CpG interactions in Table 2
* $p<0.05$, ** FDR q<0.1

Table S2-16. Association of CpG sites[a] on transcripts in the *ABCA7* region (p<0.05; n=429)

Transcript	CpG Site	β_{CpG}	*p*-value
ENST00000531478	cg06169110	0.02	0.008*
ENST00000526885	cg06169110	0.02	0.037*

Model 5: Transcript ~ CpG + age + age difference between methylation measurement and gene expression measurement + sex + PC1-4 + familial relatedness (random effect)
a. CpG sites shown were previously significant in the SNP-by-CpG interactions in Table 2
*$p<0.05$; No associations are significant at FDR q<0.1

Figure S2-3. Flow diagram illustrating sample sizes for genetic (n = 634), epigenetic (n = 494), and transcriptomic (n = 429) analyses in GENOA AA.

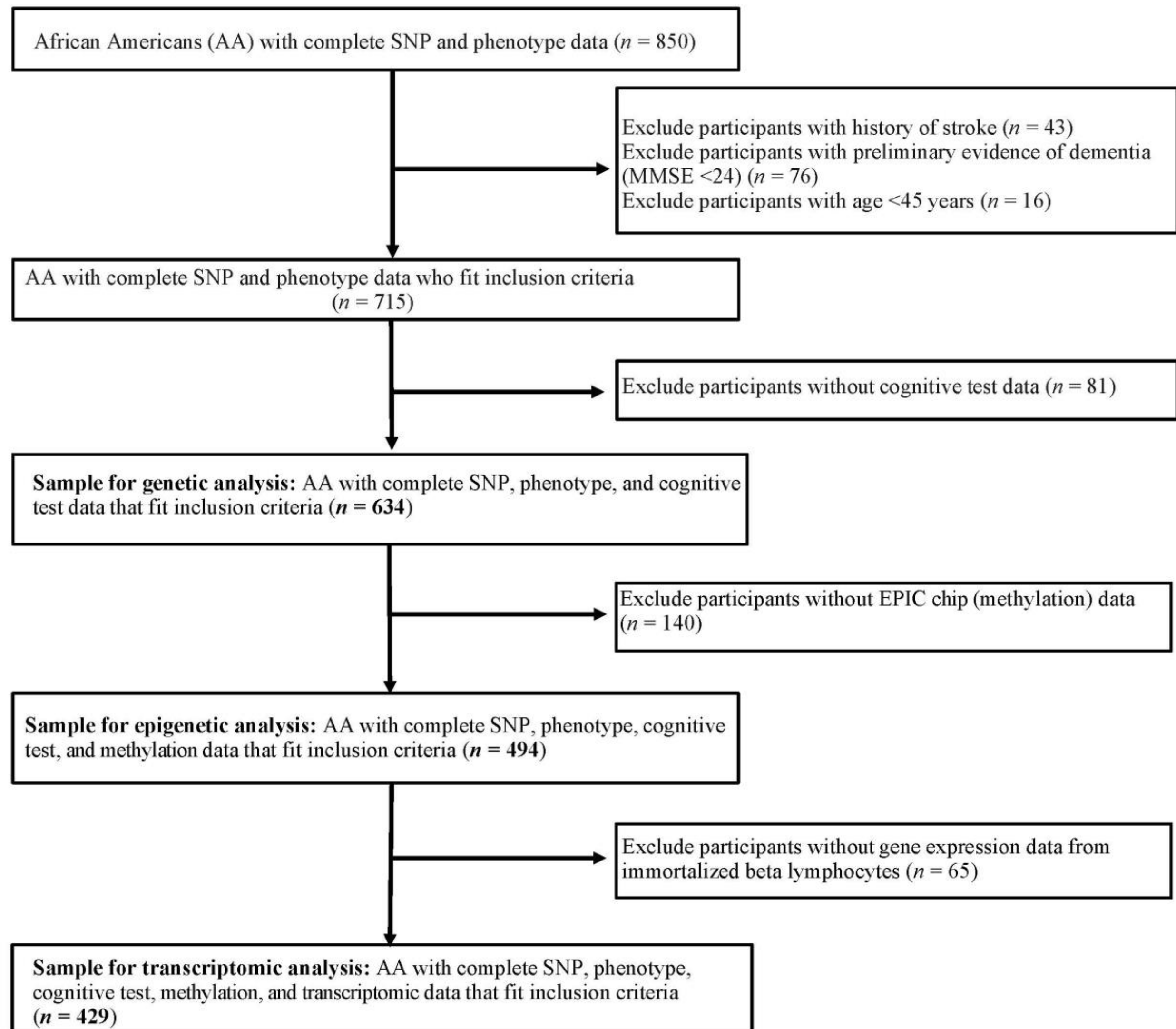


Figure S2-4. Models used to assess genetic, epigenetic and genetic-epigenetic interaction associations with general cognitive function.

<u>**Genetic associations**</u>

Model 1: General cognitive function ~ SNP + age at cognition measurement + sex + PC1-4+ familial relatedness (random effect)

Model 2: Model 1 + educational attainment

Model 3: Model 2 + *APOE ε2 + APOE ε4*

<u>**Epigenetic associations**</u>

Model 1: General cognitive function ~ CpG site + sex + age at cognition measurement + age difference between methylation and cognition measurements + smoking status + PC1-4 + familial relatedness (random effect)

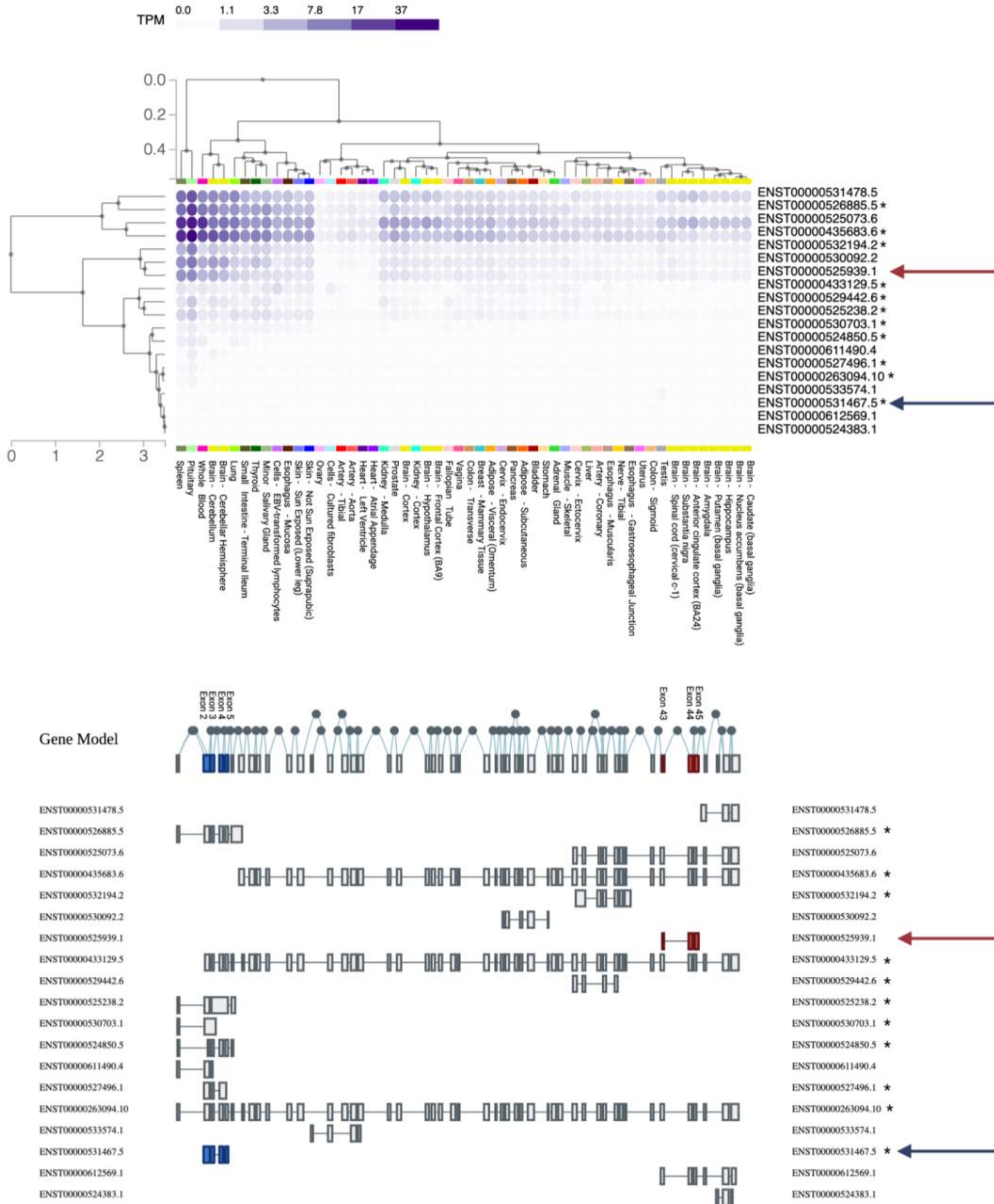
Model 2: Model 1 + educational attainment

Model 3: Model 2 + *APOE* ε2 + *APOE* ε4

<u>**Genetic-epigenetic interaction associations**</u>

Model 4: General cognitive function ~ SNP + CpG + SNP*CpG + age at cognition measurement + age difference between methylation and cognition measurements + sex + educational attainment + *APOE ε2 + APOE ε4* + smoking status + PC1-4 + familial relatedness (random effect)

Figure 2-5. Transcript expression of *ABCA7*: ENSG00000064687 (12 ATP binding cassette subfamily A member 7



Transcript expression of *ABCA7*: ENSG00000064687 (12 ATP binding cassette subfamily A member 7 [Source: HGNC Symbol; Acc:HGNC:37]). The upper panel shows the tissue expression levels for all *ABCA7* transcripts available in GTEx. The lower panel shows exonic positions of the *ABCA7* transcript isoforms. ENST00000525939 and ENST00000531467, which are associated with rs115550680*cg17316918 and rs3764647*cg22271697 interactions, respectively (Table S10), are indicated by red and blue arrows. Introns within the *ABCA7* gene that are included in each of the two transcripts are colored red and blue correspondingly. Transcripts that are associated with rs115550680 (Table S12) are indicated by asterisks. Figure adapted from https://www.gtexportal.org/home/gene/ENSG00000064687. Data Source: GTEx Analysis Release V8 (dbGaP Accession phs000424.v8.p2) [59].

Chapter 3 . Neighborhood Environment Associations with Cognitive Function and Structural Brain Measures in Older African Americans

3.1 Abstract

Since older adults spend a large proportion of their time in their neighborhood environment, factors such as neighborhood socioeconomic disadvantage, high racial segregation, low healthy food availability, low access to recreation, and minimal social engagement may have adverse effects on cognitive function and increase susceptibility to dementia. DNA methylation, which is associated with neighborhood characteristics as well as cognitive function and white matter hyperintensity (WMH), may act as a mediator between neighborhood characteristics and neurocognitive outcomes. In this study, we examined whether DNA methylation in peripheral blood leukocytes mediates the relationship between neighborhood characteristics and cognitive function (N=477) or WMH (N=404) in older AA participants without preliminary evidence of dementia from the Genetic Epidemiology Network of Arteriopathy (GENOA). For a 1-mile buffer around a participant's residence, each additional fast food destination or unfavorable food store with alcohol per square mile was associated with a 0.05 (p=0.04) and a 0.04 (p=0.04) second improvement in visual conceptual tracking score, respectively. Also, each additional alcohol drinking place per square mile was associated with a 0.62 word increase in delayed recall score (p=0.03), indicating better memory function. Although the presence of these destinations encourage unhealthy diet and behaviors, they may provide meeting places for community members that allow for greater interaction and stimulation of cognitive health. In this study, there

was no evidence that DNA methylation mediated the observed associations between neighborhood characteristics and cognitive function. Further examination of the potential pathways between the neighborhood environment and cognitive function/WMH may allow the development of potential behavioral, infrastructural, and pharmaceutical interventions to facilitate aging in place and healthy brain aging in older adults, especially in marginal populations that are most at risk.

3.2 Introduction

Dementia is preceded by a noticeable decline in cognitive abilities that becomes severe enough to interfere with daily functioning.[4] Among U.S. adults ages 65 and older, approximately 10% of have dementia and 22% have mild cognitive impairment (MCI).[1] Dementia, which includes Alzheimer's disease (AD), vascular dementia (VaD), and other types of dementia, places a substantial burden on family, friends, and healthcare systems.[2] To date, there are no effective treatments available to prevent or cure dementia. However some research suggests performing cognitively stimulating exercises and treating cardiovascular risk factors may delay or prevent the onset of dementia and reduce its associated pathology.[3,4] While individual-level factors, such as educational attainment, [5,6] smoking habits,[7] and physical activity,[8,9] are associated with cognitive function, there is growing interest in how neighborhood characteristics may shape health behaviors and health outcomes in older adults.[10,11]

Neighborhoods are defined as living and work environments that possess both physical and social attributes that may affect the health of their residents. Specifically, characteristics of the neighborhood social environment and neighborhood socioeconomic status (SES) are associated with cognitive function,[12–15] and higher incidence of ischemic stroke[16,17] in older

adults. Cerebral small vessel disease (SVD), detected on magnetic resonance imaging (MRI) as white matter hyperintensities (WMH), causes one quarter of all ischemic strokes and is associated with cognitive function[18] and VaD.[19–21] Since older adults spend a large proportion of their time in their neighborhood environment, factors such as neighborhood socioeconomic disadvantage,[22] high racial segregation,[23–26] low healthy food availability,[27] low access to recreation,[28,29] and minimal social engagement[30] may have adverse effects on cognitive function and SVD and may also increase susceptibility to dementia. As such, specific neighborhood infrastructures may support or hinder cognitive health among older adults aging in place. Understanding how neighborhood environments impact dementia pathology may allow us to develop better interventions to prevent disease onset.

Previous studies have linked several individual- and neighborhood-level social disadvantage indicators, including low adult socioeconomic status (SES)[31,32] and living in disadvantaged neighborhoods,[33–35] to DNA methylation patterns. After adjusting for individual SES, neighborhood socioeconomic disadvantage and social environment were also associated with DNA methylation in stress- and inflammation-related genes.[34] In addition, epigenome-wide association studies (EWAS) have shown associations between methylation and cognitive function[36,37] and WMH.[38,39] Since DNA methylation has been associated with both neighborhood-level factors and cognitive function/WMH, it may act as a mediator between neighborhood-level risk factors and cognitive outcomes. To date, a handful of studies have examined whether epigenome-wide markers mediate the effects of social disadvantage on health outcomes and risk factors. For example, in the New England Family Study, epigenetic markers from adipose tissue partially mediated the association between individual-level social disadvantage and body mass index (BMI) in adulthood.[40,41] In the Multi-Ethnic Study of

Atherosclerosis (MESA), methylation from monocytes partially mediated the associations between adult SES and/or neighborhood socioeconomic disadvantage and several CVD risk factors.[42] To our knowledge, no studies have examined epigenetic mediation in the association between neighborhood characteristics and cognitive function/WMH.

African Americans (AA) have a greater burden of and risk for developing dementia,[43–46] and stroke,[47] compared to Non-Hispanic Whites (NHW). Underlying causes of these disparities remain poorly understood but are likely due to multifactorial and multilevel factors that occur over the life-course. For example, differences in cognitive performance and dementia risk in AA may in part be caused by racial disparities in education (amount and quality), availability of material and social resources, access to favorable food and physical activity environments, exposure to discrimination, and neurotoxicants.[48,49] While studies have examined individual-level risk factors as explanations for racial/ethnic disparities (e.g., socioeconomic, psychosocial, genetic, epigenetic, biological), there is increasing interest in the role of the neighborhood on health outcomes in AA populations. Altogether, AA are more likely to live in neighborhoods with social factors that may affect their stress levels (e.g., higher discrimination, lower educational attainment, and lower SES) that over time may result in physiological dysregulation[25] that ultimately leads to hypertension, diabetes, coronary heart disease, and depression. Dysregulation of neurocognitive processes may also lead to cognitive decline or dementia.

To better understand the mechanisms underlying relationships between neighborhood environment and dementia risk factors in older AA, we used high-dimensional mediation methods to identify DNA methylation sites (CpGs) in peripheral blood leukocytes that may mediate the relationship between neighborhood-level factors and cognitive function or WMH in

the Genetic Epidemiology Network of Arteriopathy (GENOA) study. To better understand the functional consequences of identified CpG mediators at the molecular level, we also examined whether gene-level expression in transformed beta lymphocytes mediates CpG associations with cognitive function or WMH in the same cohort.

3.3 Materials and Methods

3.3.1 Sample

The Genetic Epidemiology Network of Arteriopathy (GENOA) is a community-based longitudinal study intended to examine the genetic effects of hypertension and related target organ damage.[50] European American (EA) and African American (AA) hypertensive sibships were recruited if at least 2 siblings were clinically diagnosed with hypertension before age 60. All other siblings were invited to participate, regardless of hypertension status. Exclusion criteria included secondary hypertension, alcoholism or drug abuse, pregnancy, insulin-dependent diabetes mellitus, active malignancy, or serum creatinine levels >2.5mg/dL. In Phase I (1996-2001), 1,854 AA participants (Jackson, MS) and 1,583 EA participants (Rochester, MN) were recruited.[50] In Phase II (2000-2004), 1,482 participants AA participants and 1,239 EA participants were successfully followed up, and their potential target organ damage from hypertension was measured. Demographics, medical history, clinical characteristics, medication use, and blood samples were collected in each phase. Methylation levels were measured only in AA participants using blood samples collected in Phases I and II.

In an ancillary study, the Genetics of Microangiopathic Brain Injury (GMBI; 2001-2006), 1,010 AA and 967 EA GENOA participants underwent a battery of established cognitive tests to assess measures of cognitive function.[51,52] White matter hyperintensity (WMH) was also

measured using brain Magnetic Resonance Imaging (MRI). The GMBI exam occurred approximately one year after the participant completed Phase II (mean time between Phase II and GMBI = 1.1 years, SD=1.0 year). Written informed consent was obtained from all participants, and approval was granted by participating institutional review boards (University of Michigan, University of Mississippi Medical Center, and Mayo Clinic).

A total of 710 AA participants had non-missing demographic, cognitive, and methylation data. Since participants with a history of stroke or dementia may have had changes in general cognitive function that differed from non-pathological cognitive aging, we excluded those with a history of stroke (n=31) and/or preliminary evidence of dementia indicated by a Mini-Mental State Examination Score (MMSE) of <24 (n=38). Participants younger than age 45 were also excluded (n=28). A total of 542 and 477 participants were available with neighborhood spatial (density measures) and neighborhood socioeconomic disadvantage analyses, respectively (Figure S3-4).

A total of 602 AA participants had non-missing demographic, WMH, and methylation data. Participants with a history of stroke (n=17), and/or preliminary evidence of dementia indicated by a Mini-Mental State Examination Score (MMSE) of <24 (n=23) were excluded. Participants younger than age 45 were also excluded (n=17). A total of 466 and 404 participants were available for neighborhood spatial (density measures) and neighborhood socioeconomic disadvantage analyses, respectively (Figure S3-5).

3.3.2 Measures

A. Measures of cognitive function

The following four cognitive domains were evaluated: delayed recall (Rey Auditory Verbal Learning Test (RAVLT)), processing speed (Digit Symbol Substitution Test (DSST)), word fluency (Controlled Oral Word Association Test (COWA-FAS)) and visual conceptual tracking (Trail Making Test A (TMTA)).[51,52] All cognitive domains were coded so that a higher score corresponds to better cognitive function. See Supplementary Methods for additional details.

In addition to analyzing individual cognitive domains, we assessed a summary measure of general cognitive function, which is often quantified using cognitive tests in multiple cognitive domains.[53] In this study, general cognitive function was calculated as the first unrotated principal component (FUPC) from a principal component analysis (PCA) of the four cognitive domains in the full sample (N=542). The FUPC accounted for 57% of the total variance in the cognitive measures and loading factors of the four measures were 0.61 for delayed recall (RAVLT), 0.88 for processing speed (DSST), 0.70 for word fluency (COWA-FAS) and 0.81 for visual conceptual tracking (TMTA).

B. White matter hyperintensity

Presence of WMH in brain samples indicates areas of ischemic damage to small vessels and surrounding areas. Brain magnetic resonance images were measured from magnetic resonance imaging (MRI), using Signa 1.5T MRI scanners (GE Medical Systems, Waukesha, WI, USA) at Mayo Clinic.[54] For additional details, see Smith et al.[55] WMH and total brain volume in the coronaradiata and periventricular zone were quantified from axial fluid-attenuated

inversion recovery (FLAIR) images.[56] Brain scans with cortical infarctions were excluded from the analyses because of the distortion of WMH volume estimates that would be introduced in the automated segmentation algorithm. Models assessing WMH were adjusted for total intracranial volume (TIV). Distributional plots indicated that the measures of WMH are right-skewed, so the WMH variable was transformed as ln(WMH + 1).

C. DNA methylation measures

Genomic data was extracted from stored peripheral blood leukocytes from 1,106 AA GENOA participants from Phase I and 304 AA participants from Phase II using the AutoGen FlexStar (AutoGen, Holliston, MA). Bisulfite conversion was performed with the EZ DNA Methylation Kit (Zymo Research, Irvine, CA), and methylation was measured using the Illumina HumanMethylationEPIC BeadChip. The raw intensity data was visualized using the shinyMethyl R package [57] to identify sex mismatches and outliers, which were removed. Samples with incomplete bisulfite conversion were identified using Qcinfo in the *Enmix* R package[58] and removed. Background correction and dye-bias normalization were performed using Noob in the *Minfi* R package.[59,60] Sample identity was verified using 59 SNP probes on the EPIC array, and mismatched samples were removed. Probe-type bias was adjusted using the Regression on Correlated Probes (RCP) method.[61] Probes with detection p-value $<10^{-16}$ were considered successfully detected, and probes and samples with detection rate<10% were removed.[62] We also excluded cross-reactive probes[63] and probes with a SNP at the target CpG site or within a single-base extension. After quality control, a total of 1,396 samples (N=1,100 from Phase I and N=294 from Phase II) and 857,121 CpG sites were available for analysis. For this analysis, all methylation data were from Phase I samples. White blood cell proportions for CD8+ T lymphocytes, CD4+ T lymphocytes, natural killer cells, B cells, monocytes, and granulocytes

were estimated using the Houseman method.[64] For each CpG site prior to analysis, the methylation beta-values[65,66] were pre-adjusted for batch effects (sample plate, row, and column) and white blood cell proportions using linear mixed modeling, and the resulting residuals were added to the mean values.

D. Gene expression measures

Gene expression levels in transformed beta-lymphocyte cell lines from blood samples taken primarily at GENOA Phase II were measured using the Affymetrix Human Transcriptome Array 2.0. The Affymetrix Expression Console was used for quality control, and all array images passed visual inspection. Affymetrix Power Tool software was used to process raw intensity data.[67] We normalized Affymetrix CEL files using the Robust Multichip Average (RMA) algorithm, including background correction, quantile normalization, log_2-transformation, and probe set summarization.[68] Linearity was also maintained using GC correction (GCCN), signal space transformation (SST), and gain lock (value=0.75). We used the Brainarray custom CDF[69] version 19 to map the probes to genes, specifically removing probes with non-unique matching cDNA/EST sequences that can be assigned to more than one gene cluster. As a result, the gene expression data processed through the custom CDF is expected to be free of mappability issues; however, alignment bias may still exist due to genetic variation, errors in the reference genome, and other complications.[70] After mapping, Combat was used to remove batch effects.[71] A total of 17,616 gene-level expression values were available for analysis.

E. Individual-level measures

Age was assessed at cognitive testing. Adult socioeconomic status (SES) was indicated by the respondent's highest level of educational attainment, categorized as: (1) less than high school degree/GED (reference group), (2) high school degree or GED, and (3) at least four years of college or trade/technical school. Smoking has a substantial impact on the epigenome[72] so we used smoking data from the same timepoint as the DNA methylation measures (Phase I). Participants were categorized as current, former, or never smokers (reference group).

F. Neighborhood characteristics

i. GIS-based measures

Neighborhood density characteristics were derived from Geographic Information System (GIS)[73] data (1996-2015). Simple densities per square mile were created for ½-mile, 1-mile, and 3-mile buffer sizes around home addresses of GENOA participants at Phase I using ArcGIS V.9.3 (ESRI, Inc., Redlands, California)[74,75] We used 1-mile buffer in our primary analysis, as previous studies have done,[76,77] and examined ½- and 3-mile buffers in sensitivity analysis. Kernel densities per square mile, with greater weighting towards destinations located closer to the home of a participant, were also created for GENOA participants using the kernel density command in ArcGIS V.9.3[74,75] for the same buffer sizes; these were also explored in sensitivity analysis.

For each participant, simple densities were estimated for the following 12 destinations: fast food restaurants (including both chain and non-chain), total physical activity facilities, total social engagement destinations, and alcohol outlets. Summary density measures were also created for densities of unfavorable food stores with and without alcohol, healthy (favorable) food stores, popular walking destinations, total stores, and total food stores. The modified retail

food environment index (MRFEI) was calculated from the number of healthy and less healthy food retailers within census tracts across states, based on typical food offerings in specific retail stores.[78] See Supplementary Methods for additional details.

ii. Census measures

Briefly, neighborhood socioeconomic disadvantage was assessed using data collected in the 2000 U.S. Census,[79,80] American Community Survey (ACS) 2005-2009,[81] and ACS 2007-2011.[82] Data was linked to GENOA participant data (Phase I; 1995-2000) by census tract using Census and ACS estimates for the closest time period. To derive neighborhood socioeconomic disadvantage, we used six variables that reflected aspects of wealth and income, education, and occupation for each census tract.[83] Z-scores for each census tract were estimated for each variable, and neighborhood socioeconomic disadvantage was defined as the sum of Z-scores from the six variables, with higher scores indicating more disadvantage. See Supplementary Methods for additional details.

3.3.3 Statistical analysis

We first calculated Pearson correlations among the five cognitive outcomes (general cognitive function and the four cognitive domains), and among the 13 neighborhood characteristics (12 density measures and neighborhood socioeconomic disadvantage). Since areas of increased population density (e.g., urban neighborhoods) generally have a higher absolute number of destinations, we next examined the neighborhood characteristics after pre-adjusting for census tract population density using linear modeling. Correlations were calculated among

the neighborhood characteristics for simple and kernel densities per square mile for 1-mile buffer sizes.

Associations between neighborhood measures and cognitive function/WMH

To identify which exposures and outcomes have a significant total effect, we tested for association between each neighborhood characteristic (exposure) and general cognitive function, each cognitive domain, or WMH (outcome), and assessed significance at alpha=0.05. We first tested for association between a neighborhood characteristic (socioeconomic disadvantage or simple density measures) and general cognitive function, adjusting for age at cognitive function measurement, sex, current smoking status, the first 5 genetic principal components (PCs) of ancestry, and family relatedness as a random effect (Model 1a). In Model 1b, we tested for association between each neighborhood characteristic and WMH, adjusting for the same covariates as Model 1a and TIV. In Models 2a/2b, we additionally adjusted for census tract population density in 2000 and included census tract as a random effect. We also tested for associations between each neighborhood characteristic and each of the four cognitive domains using Model 2a. Associations between neighborhood characteristics and cognitive function/WMH that were significant at P<0.05 in Models 1a/1b or 2a/2b were selected for mediation analysis. In sensitivity analysis, we tested the same associations using simple densities at ½- and 3-mile buffers as well as kernel densities at all 3 buffers. The total effects model is outlined below:

$$Y_{2jk} = \beta_0 + \boldsymbol{\omega} X_{1jk} + \alpha C_{1jk} + W_k + \varepsilon_{jk}$$

β_0: intercept value; cognitive function/WMH value when neighborhood characteristic (exposure) equals zero

ω: effect estimate of neighborhood characteristic (exposure) on cognitive function/WMH
X_{1jk}: neighborhood characteristic (exposure) for participant j in sibship k at Phase I
C_{1jk}: set of covariates (age at cognitive function/WMH measurement, sex, and genetic principal components at Phase I; and TIV for WMH outcome).
W_k: random effect (familial relatedness).
ε_{jk}: residual error (independent and normal distribution) for participant j in sibship k.
Y_{2jk}: cognitive function/WMH for participant j in sibship k at Phase II

Mediation analysis

If a significant association (total effect) was identified between a neighborhood characteristic and a cognitive/WMH outcome, we conducted an epigenome-wide high-dimensional mediation analysis to identify CpG sites that may partially mediate the relationship. We used a cross-product-based mediation approach in which the mediation effect is obtained by multiplying the exposure-mediator effect (β_1) and the mediator-outcome effect (β_3; see Equations 1 and 2 below). We obtained these parameters for each exposure and outcome tested using linear mixed models to separately estimate the association between neighborhood characteristics with DNA methylation (mediator), while adjusting for covariates (Equation 1), and the association between DNA methylation and cognitive function/WMH, while adjusting for the corresponding exposure tested and the same set of covariates (Equation 2). The covariate sets in Equations 1 and 2 are the same as in Models 1a/b and 2a/b. The specified models (Equations 1 and 2) for a given exposure-outcome association are outlined below:

$$M_{jk} = \beta_0 + \boldsymbol{\beta_1} X_{1jk} + \alpha V_{1jk} + W_k + \varepsilon_{jk} \qquad \text{(Equation 1)}$$

$$Y_{2jk} = \beta_0 + \beta_2 X_{1jk} + \boldsymbol{\beta_3} M_{jk} + \alpha V_{1jk} + W_k + \varepsilon_{jk} \qquad \text{(Equation 2)}$$

β_0: intercept value; cognitive function/WMH value when neighborhood characteristic (exposure) equals zero
M_{jk}: DNA methylation (mediator; beta-value) for participant j in sibship k
X_{1jk}: neighborhood characteristic (exposure) for participant j in sibship k at Phase I

V_{1jk}: adjustment covariates for participant j in sibship k at Phase I
W_k: random effect for each sibship which accounts for the multiple siblings within sibships
ε_{jk}: residual error (independent and normal distribution) for participant j in sibship k
Y_{2jk}: cognitive function/WMH (outcome) for participant j in sibship k at Phase II
β_1: effect estimate of neighborhood characteristic (exposure) on DNA methylation (mediator)
β_2: direct effect estimate of the neighborhood characteristic (exposure) on cognitive function/WMH (outcome)
β_3: effect estimate of DNA methylation (mediator) on cognitive function/WMH (outcome), adjusting for the direct effect (β_2)

Using Equations 1 and 2 above, the epigenetic mediation effect was tested using the following:

$$H_0: \beta_1\beta_3 = 0$$

$$H_A: \beta_1\beta_3 \neq 0$$

The null hypothesis was comprised of three sub-hypotheses: (1) H_{01}: $\beta_1 = 0, \beta_3 \neq 0$; (2) H_{10}: $\beta_1 \neq 0, \beta_3 = 0$; and (3) H_{00}: $\beta_1 = \beta_3 = 0$. To that end, π_{01}, π_{10} and π_{00} are the true proportions of ($\beta_1 = 0, \beta_3 \neq 0$), ($\beta_1 \neq 0, \beta_3 = 0$) and ($\beta_1 = \beta_3 = 0$) among all *J* tests. Figure 3-1 shows a directed acyclic graph (DAG) of the hypothesized associations. To test for the mediation effect, we used the Sobel-comp[84] method in the *medScan* package in R, which uses a corrected mixture reference distribution for Sobel's test statistic according to the composite structure of the null hypothesis. We corrected for multiple testing using the false discovery rate (FDR)[85] on the mediation p-values (FDR q<0.10).[85]

3.4 Results

3.4.1 Sample Characteristics

The sample included 542 AA without dementia (Table 3-1). Participant age ranged from 45 to 83 years (mean = 62.5 years). More than half of participants (73%) were female. A total of 25.0% had less than a high school degree/GED, 46.5% attained a high school degree/GED, and

28.6% completed at least four years of college or trade school. General cognitive function was normally distributed (Figure 3-2). Mean delayed recall (RAVLT) score was 7.0 (SD=3.3) words recalled, mean processing speed (DSST) was 33.8 (SD=13.0) symbols, mean word fluency (COWA-FAS) score was 29.4 (SD=11.6) words, and mean visual conceptual tracking (TMTA) score was 63.8 (SD=35.2) seconds to completion. Participants had a mean WMH of 9.42 cm^3 (SD=9.19). WMH distribution was severely right skewed but had a normal distribution after log transformation.

3.4.2 Correlation among cognitive and WMH outcomes

The four cognitive domains were moderately correlated (Pearson r ranged from 0.21 to 0.68), with the highest correlation among processing speed (DSST) and visual conceptual tracking (TMTA) (r=0.68, p<0.001, Table S3-6). WMH was negatively and weakly correlated with all the cognitive measures except COWA-FAS (Pearson r ranged from -0.27 to -0.34 for significant correlations).

3.4.3 Correlation among the neighborhood exposures

Pearson correlations among the neighborhood exposures, including neighborhood socioeconomic disadvantage and the 12 neighborhood simple density measures per square mile for 1-mile buffer size, are shown in Table S3-7. Neighborhood exposures were moderately correlated (Pearson r ranged from -0.237 to 0.995), with the highest correlation between the simple densities of total social engagement and MRFEI with alcohol (r = 0.995, p<0.001). Neighborhood socioeconomic disadvantage was positively, but weakly, correlated with

unfavorable food stores without alcohol, total social engagement destinations, total popular walking destinations and alcoholic drinking places.

After adjusting for census tract population density, the correlations between neighborhood socioeconomic disadvantage and neighborhood characteristics increased in magnitude in the positive direction for all measures except fast food destinations, alcoholic drinking places, and the MRFEI measures. For instance, neighborhood socioeconomic disadvantage was negatively correlated with fast food destinations ($r=-0.20$, $p<0.01$) and unfavorable food stores with alcohol ($r=-0.21$, $p<0.001$); however, after adjusting for census tract population density, fast food destinations were weakly correlated with neighborhood disadvantage ($r=-0.02$, $p<0.001$) and positively correlated with unfavorable food stores ($r=0.92$, $p<0.001$; Table S3-8). The simple and kernel densities of each neighborhood characteristic are strongly and positively correlated with each other (Pearson r ranged from 0.702 to 0.934; Table S3-9).

3.4.4 Associations between neighborhood characteristics and cognitive/WMH outcomes

A. Neighborhood socioeconomic disadvantage associations

Neighborhood socioeconomic disadvantage was not associated with general cognitive function or WMH either before (Models 1a/1b) or after adjusting for census tract population density and census tracts as a random effect (Models 2a/2b, Table 3-2). Further, neighborhood socioeconomic disadvantage was not associated with any of the four cognitive domains (Model 2a, Table 3-3).

A. Density associations

There was no association between the 12 neighborhood simple density exposures at 1-mile buffer size and cognitive/WMH outcomes either before (Models 1a/1b) or after adjusting for census tract population density and census tracts as a random effect (Models 2a/2b; Table 3-4). The associations between simple neighborhood densities per square mile for ½- and 3-mile buffer sizes and cognitive function/WMH are reported in Table S3-10. One additional alcoholic drinking place per square mile for the 3-mile buffer size was associated with a 0.71 SD decrease in general cognitive function after adjusting for census tract population density and census tracts as a random effect (p=0.03; Model 2a; Table S3-10).

We also tested the association between the 12 neighborhood simple density exposures examined at 1- mile buffer region with each of the four cognitive domains (Model 2a; Table 3-5). One additional fast food destination or unfavorable food store with alcohol per square mile was associated with a 0.05 (p=0.04) and a 0.04 (p=0.04) second increase in visual conceptual tracking score, respectively, indicating that more of these destinations was associated with better visual conceptual tracking. In addition, one additional alcohol drinking place per square mile was associated with a 0.62 word (p=0.03) increase in delayed recall score (Table 3-5), indicating better memory function. The associations between simple neighborhood densities per square mile for ½- and 3-mile buffer sizes and cognitive/WMH measures are also reported in Tables S3-10 and S3-11.

We also tested the association between the 12 neighborhood kernel density exposures at ½-, 1- and 3- mile buffer sizes with cognitive function/WMH (Table S3-12) and the cognitive domains (Table S3-13). There were no associations between the kernel density neighborhood exposures and general cognitive function or WMH in Models 1a/2a and 1b/2b (Table S3-12). At

the 1-mile buffer, kernel density of fast food destinations and unfavorable food stores with alcohol were both associated with better visual conceptual tracking, consistent with the simple density associations; however, the association between kernel density of alcohol drinking places and delayed recall score was not. We also found that at the 1-mile buffer, kernel densities of unfavorable food stores without alcohol, total popular walking destinations, and total food stores were all associated with better visual conceptual tracking as well. The associations between kernel neighborhood densities per square mile for ½- and 3-mile buffer sizes and cognitive/WMH measures are also reported in Tables S3-12 and S3-12.

3.4.5 Mediation analysis

When the total effect of a neighborhood characteristic (simple density at 1-mile buffer) and cognitive function/WMH was significant at $p<0.05$, we conducted epigenome-wide high-dimensional mediation analysis to identify possible CpG sites that may partially mediate the relationship between the neighborhood exposure and corresponding outcome using Model 2a in 477 participants with complete data. The following exposure-outcome combinations were investigated: (a) alcohol drinking places and delayed recall, (b) fast food destinations and visual conceptual tracking, and (c) unfavorable food stores with alcohol and visual conceptual tracking. Figure 3-3 shows quantile-quantile (QQ) plots for the 5 exposure-outcome relationships using Sobel-Comp. The p-values from Sobel-Comp test were deflated, potentially due to the large number of zero exposure-mediator (β_1) and mediator-outcome (β_3) estimates and the small sample size (Figure 3-3). No associations were significant at FDR $q<0.1$.

3.5 Discussion

As the aging population rapidly grows, a better understanding of how the neighborhood environment may affect cognitive health is needed to mitigate the future burden of dementia in the U.S. While there are studies showing the effect of individual factors, such as lifestyle, genetics and biomarkers on cognitive function, there is little research on the association between neighborhood characteristics and cognitive function to date.[86] Further, only a few studies have examined the potential molecular mechanisms linking neighborhood environment and cognitive health.[12,87] To our knowledge, this study is the first assessment of whether DNA methylation partially mediates the association between various neighborhood environment characteristics and cognitive function in AA without dementia. This cross-sectional study suggests that greater simple densities of alcohol drinking places may be associated with better memory as measured by delayed recall (RAVLT), and greater densities of fast-food destination and unfavorable food stores with alcohol with better attention and task switching as measured by visual conceptual tracking (TMTA) in cognitively normal AA. However, we did not find associations between neighborhood characteristics and WMH. We also were unable to detect mediating effects of DNA methylation on the associations between these neighborhood characteristics on cognitive function and cognitive measures in this sample.

We initially expected higher densities of unfavorable food stores to be associated with worse cognitive function, suggesting that increased access to unhealthy food and drink may encourage unhealthy dietary choices that lead to lower cognitive health. Instead, we found that greater densities of alcohol drinking places, fast-food, and unfavorable stores with alcohol that may encourage unhealthy dietary choices were associated with better cognitive function as measured by delayed recall and visual conceptual tracking after adjustment for population

density. Considering that Jackson, MS does not have a highly dense population (approximately 1,300 people per square mile in 2010), the presence of these walking destinations may provide meeting places for community members, allowing for greater interaction and stimulation of cognitive health, regardless of their impact on unhealthy diet and behaviors. As such, these meeting hubs may contribute to better cognitive function through increased access to community residents, neighborhood community resources, and proximal walking destinations that improve cognitive health by increasing physical activity levels, social engagement, mental health or quality of life.[88]

To date, results from previous studies examining similar characteristics of the neighborhood environment and cognitive function have been mixed. In the Chicago Health and Aging Project (CHAP), increasing densities of social and walking destinations such as community centers were associated with slower cognitive decline,[89] yet a study in the Multi-Ethnic Study of Atherosclerosis (MESA) showed an inverse association between these same measures and cognitive function, and most noticeably in individuals of non-white race.[90] Also, closer access to community resources has been associated with better cognitive function in NHW, but worse cognitive function in AA,[91] while other studies showed no association between the presence of neighborhood built environment characteristics, such as recreation centers and institutional resources (e.g., libraries, schools and community centers) and cognitive function.[89,91,92] In our study, the plausible mechanisms and direction or presence of neighborhood-cognitive function association may depend on the neighborhood characteristic and cognitive domain being studied, and more than one mechanism may be at play.

Different underlying mechanisms of neighborhood environment on cognitive function have been examined to understand how interventions can prevent dementia onset. In MESA,

increasing social destination density, walking destination density, and intersection density were associated with worse cognitive function, and increasing proportion of land dedicated to retail was associated with better processing speed.[93] While we did not observe similar patterns among simple densities, we did observe greater kernel densities of total popular walking destinations per square mile (for ½- and 1-mile buffer sizes) were associated with higher visual conceptual tracking and greater kernel densities of total social engagement destinations per square mile (½-mile buffer) were associated with higher delayed recall. Access to a safe and walkable neighborhood environment may help older adults age in place and delay the onset of cognitive impairment and decline prior to dementia.[92,94,95] In addition, the positive relationship between proportion of land dedicated to retail and processing speed may be explained by increased utilitarian physical activity and social engagement, or increased cognitive stimulation that contributes to the cognitive reserve.[92] Also, fast-food outlets and local retail food environments may play a role in providing social and community engagement, connectedness, emotional support and cognitive stimulation for older adults outside of more formal or age-graded settings such as doctor's office, church or senior center.[96,97]

Other studies have found inverse relationships between neighborhood characteristics and cognitive function that may be related to cognitive overload among older adults due to stress from greater number of destination choices or navigation of traffic. It is possible that highly dense areas consisting of social and walking destinations and street intersections have increased vehicular pollutant exposure due to decreased distances to busy roadways and decreased air ventilation created by buildings.[98] Airborne pollutants have been associated with worse cognitive function and brain structure in older adults.[98] Factors such as neighborhood socioeconomic disadvantage,[22] low healthy food availability,[27] low access to recreation,[28,29] high racial

segregation,[23–26] and minimal social engagement[30] may have adverse effects on cognitive function and increase susceptibility to dementia as well. These mixed results from other studies may be affected by residual confounding from unmeasured factors. Thus, additional research on the many confounders and mechanisms related to the relationship between the neighborhood environment and cognitive function is necessary.

In addition, we found correlations between favorable and unfavorable destinations, even after adjusting for population density, which may further illuminate our findings in the context of cognitive health and behaviors. For example, greater densities of fast-food destinations were associated with greater densities of favorable food stores, physical activity destinations, and MRFEI (the proportion of favorable food stores to total food stores), even after adjusting for population density. These correlations in Jackson may be attributed to a complex interplay of socioeconomic, urban planning, cultural, historical and policy-related factors and confounders. Further, socioeconomic disparities often lead to variations in access to health-promoting resources, with neighborhoods of lower SES facing limited access to healthy options and an increased prevalence of unhealthy alternatives. The availability of favorable food stores may reflect the demand from residents, according to their purchasing power, who can afford healthier options. To account for this discrepancy, we adjusted for neighborhood socioeconomic disadvantage in our associations. The city's urban planning, historical development (e.g., redlining and discriminatory housing practices in the past) and government policies may play crucial roles in shaping the distribution of health-related destinations. Another possibility is that areas with higher commercial zoning may attract both fast food establishments and favorable food stores, creating clusters of businesses in certain neighborhoods. Additionally, cultural preferences and consumer demand influence the types of businesses and amenities in specific

neighborhoods. For example, the high correlation between favorable and unfavorable food store density may be due to a micro-cultural artifact at play Jackson that encourages increased densities of fast food in Black neighborhoods.[99] This micro-culture, which results from shared race/ethnicity, beliefs, styles, skills, and habits of residents of a particular area, may disfavor physical activity and other healthy behaviors, even in the presence of features that allow for them.[100,101]

Considering that the neighborhood context has the potential to influence cognitive function, it is important to clarify the potential biological mechanisms linking neighborhood characteristics and cognitive function to shed light on the etiology and causal mechanisms driving health disparities. DNA methylation may help us better understand the pathways that mediate or interact with the environment and cognitive function. Previous studies have shown that the neighborhood context affects DNA methylation, even after adjusting for individual level factors, and that DNA methylation patterns in stress and inflammatory pathways may be responsive to interventions.[34] EWAS have also found multiple CpGs related to neurodegeneration associated with cognitive function.[36,37] Considering these factors and that past studies have found CpGs mediating the relationship between neighborhood socioeconomic disadvantage and various cardiovascular risk factors,[40–42] which are potential upstream factors of cognitive function and dementia, we expected to detect mediating CpG sites in the associations between neighborhood characteristics and cognitive function/WMH.

One reason that we may not have observed epigenetic mediation is the choice of mediation model implemented. Sobel-Comp[84] is a more powerful extension of high-dimensional mediation hypothesis testing (HDMT) [36] that is preferred when almost all exposure-mediator and mediator-outcome associations are equal to 0 (π_{00} is close to 1), and there are almost no non-zero

exposure-mediator or mediator-outcome associations (π_{01} and π_{10} are close to 0). One limitation is that Sobel-Comp is conservative under these conditions, compared to other high-dimensional mediation methods such as JT-Comp;[102] however Sobel-Comp has the advantages of using the correct mixture reference distribution for Sobel's test statistic, maintain a false positive rate (FPR) close to the nominal level, and it yielding larger true positive rates (TPRs). In this study, Sobel-Comp was the appropriate method because π_{00} was bounded away from 1 for all associations tested, but we did not detect significant mediation effects due to a potentially large number of zero exposure-mediator (β_1) and mediator-outcome (β_3) estimates, deflated p-values and small sample size. In addition, DNAm levels of proximal CpGs in the same biological pathways may be correlated, resulting in properties that are not desirable for TPR and FPR.[65] When there correlated mediators, single-mediator hypothesis testing methods like Sobel-Comp are unable to fully account for all the mediator-outcome confounders affected by the exposure (also known as co-mediators), thus reducing the power to detect mediating CpGs and potentially biasing our effect estimates.[42,103–105] While it is possible to jointly model multiple mediators using the Bayesian high-dimensional mediation method[106] and its use may have reduced the multiple testing burden and increased the power to detect independent effects, this method is computationally heavy and only a few thousand mediators would have been evaluated simultaneously at a time.[106–108] Evaluating our mediation analysis models to account for multiple correlated mediators are of interest for future analysis. Our results may indicate that methylation is not a critical component of the mediating pathway between neighborhood exposures and cognitive/WMH outcomes. Our observed associations should also be considered with caution due to the limited statistical power inherent in our sample. The small sample size may have restricted our ability to detect the total effects between neighborhood characteristics and

cognitive/WMH outcomes that could exist within the population. As a result, our findings may not be generalizable beyond our sample.

Our study also had other limitations. Our findings may be affected by residual confounding by unmeasured variables, increased exposure to factors including air pollution, potential for chance social interactions, crime, physical disability, discrimination and structural racism that may be due to increased walking in the neighborhood which influences cognitive function, or factors related to study design (e.g., cross-sectional nature). Moreover, we did not investigate the important ways in which air pollution, structural racism and stress are mediators on the pathways of specific neighborhood-cognitive function/WMH associations. Also, further longitudinal and life-course studies that explore mediation pathways between early-life, mid-life and late-life neighborhood, methylation, and cognitive function/WMH measures are needed. In this study, neighborhood characteristics were based on current home addresses, and we did not take into account that earlier or longer-term neighborhood exposures may be important for late-life cognitive function/WMH.

Our study also has notable strengths. To our knowledge, this study is the first to examine the role of DNA methylation in mediating the relationships between neighborhood characteristics and cognitive function/WMH in a cohort of older adults without diagnosed dementia. Our study was also conducted in AA, an understudied population with a higher prevalence of dementia[109,110] and higher conferred risk of cognitive decline and dementia from neighborhood environment compared to EA.[111] Additionally, with rich cognitive and WMH measures, we were able to assess associations with multiple cognitive domains, general cognitive function, and a risk factor for VaD. We were also able to adjust for neighborhood socioeconomic disadvantage to control for the influence of income, education, employment and other SES

indicators that might independently affect cognitive health. We also controlled for confounding by census tract population density because it could influence the availability of stores and cognitive outcomes. High-density urban areas may have greater access to stores and services and low-density rural areas may have lower access to these destinations. Both densities may affect cognitive health, so adjusting for population density ensures that our results are not skewed by these population differences and are more accurate. Also, we utilized a powerful high dimensional mediation method that reduced the likelihood of false positives. Lastly, our primary analysis used 1-mile density buffers around participants' homes, which provide more precise spatial representation of neighborhoods than administrative boundaries and may more accurately reflect nearby places and distances that an older adult would walk.

3.6 Conclusion

In the present study, we found that destination density had small but notable effects on several domains of cognitive function in AA without dementia. However, we detected no significant mediating effects of DNA methylation on these associations. Upon further examination of the potential pathways between the neighborhood environment and cognitive function, we may develop potential behavioral, infrastructural, and pharmaceutical interventions to allow aging in place and healthy brain aging in older adults, especially marginal populations that are most at risk.

3.7 References

1. Manly JJ, Jones RN, Langa KM, et al. Estimating the Prevalence of Dementia and Mild Cognitive Impairment in the US: The 2016 Health and Retirement Study Harmonized Cognitive Assessment Protocol Project. *JAMA Neurology*. Published online October 24, 2022. doi:10.1001/jamaneurol.2022.3543

2. Hurd MD, Martorell P, Delavande A, Mullen KJ, Langa KM. Monetary Costs of Dementia in the United States. *N Engl J Med*. 2013;368(14):1326-1334. doi:10.1056/NEJMsa1204629

3. Nelson L, Tabet N. Slowing the progression of Alzheimer's disease; what works? *Ageing Research Reviews*. 2015;23:193-209. doi:10.1016/j.arr.2015.07.002

4. Vemuri P, Lesnick TG, Przybelski SA, et al. Effect of intellectual enrichment on AD biomarker trajectories: Longitudinal imaging study. *Neurology*. 2016;86(12):1128-1135. doi:10.1212/WNL.0000000000002490

5. Aneshensel CS, Ko MJ, Chodosh J, Wight RG. The Urban Neighborhood and Cognitive Functioning in Late Middle Age. *J Health Soc Behav*. 2011;52(2):163-179. doi:10.1177/0022146510393974

6. Wight RG, Aneshensel CS, Miller-Martinez D, et al. Urban neighborhood context, educational attainment, and cognitive function among older adults. *Am J Epidemiol*. 2006;163(12):1071-1078. doi:10.1093/aje/kwj176

7. Nooyens ACJ, van Gelder BM, Verschuren WMM. Smoking and cognitive decline among middle-aged men and women: the Doetinchem Cohort Study. *Am J Public Health*. 2008;98(12):2244-2250. doi:10.2105/AJPH.2007.130294

8. Lautenschlager NT, Almeida OP. Physical activity and cognition in old age. *Curr Opin Psychiatry*. 2006;19(2):190-193. doi:10.1097/01.yco.0000214347.38787.37

9. Eckstrom E, Neukam S, Kalin L, Wright J. Physical Activity and Healthy Aging. *Clin Geriatr Med*. 2020;36(4):671-683. doi:10.1016/j.cger.2020.06.009

10. Sallis JF, Saelens BE, Frank LD, et al. Neighborhood built environment and income: examining multiple health outcomes. *Social science & medicine*. 2009;68(7):1285-1293.

11. Koohsari MJ, Badland H, Sugiyama T, Mavoa S, Christian H, Giles-Corti B. Mismatch between perceived and objectively measured land use mix and street connectivity: associations with neighborhood walking. *Journal of Urban Health*. 2015;92:242-252.

12. Besser LM, McDonald NC, Song Y, Kukull WA, Rodriguez DA. Neighborhood Environment and Cognition in Older Adults: A Systematic Review. *American Journal of Preventive Medicine*. 2017;53(2):241-251. doi:10.1016/j.amepre.2017.02.013

13. Rosso AL, Flatt JD, Carlson MC, et al. Neighborhood socioeconomic status and cognitive function in late life. *American journal of epidemiology*. 2016;183(12):1088-1097.

14. Lang IA, Llewellyn DJ, Langa KM, Wallace RB, Huppert FA, Melzer D. Neighborhood deprivation, individual socioeconomic status, and cognitive function in older people: analyses from the English Longitudinal Study of Ageing. *Journal of the American Geriatrics Society*. 2008;56(2):191-198.

15. Shih RA, Ghosh-Dastidar B, Margolis KL, et al. Neighborhood socioeconomic status and cognitive function in women. *American journal of public health*. 2011;101(9):1721-1728.

16. Lisabeth L, Diez Roux A, Escobar J, Smith M, Morgenstern L. Neighborhood Environment and Risk of Ischemic Stroke: The Brain Attack Surveillance in Corpus Christi (BASIC) Project. *American Journal of Epidemiology*. 2007;165(3):279-287. doi:10.1093/aje/kwk005

17. Brown AF, Liang LJ, Vassar SD, et al. Neighborhood disadvantage and ischemic stroke: the Cardiovascular Health Study (CHS). *Stroke*. 2011;42(12):3363-3368.

18. Kuller LH, Shemanski L, Manolio T, et al. Relationship between ApoE, MRI findings, and cognitive function in the Cardiovascular Health Study. *Stroke*. 1998;29(2):388-398. doi:10.1161/01.str.29.2.388

19. Markus HS, Hunt B, Palmer K, Enzinger C, Schmidt H, Schmidt R. Markers of Endothelial and Hemostatic Activation and Progression of Cerebral White Matter Hyperintensities. *Stroke*. 2005;36(7):1410-1414. doi:10.1161/01.STR.0000169924.60783.d4

20. Williams OA, Zeestraten EA, Benjamin P, et al. Predicting Dementia in Cerebral Small Vessel Disease Using an Automatic Diffusion Tensor Image Segmentation Technique. *Stroke*. 2019;50(10):2775-2782. doi:10.1161/STROKEAHA.119.025843

21. Prins ND, Scheltens P. White matter hyperintensities, cognitive impairment and dementia: an update. *Nat Rev Neurol*. 2015;11(3):157-165. doi:10.1038/nrneurol.2015.10

22. Hunt JFV, Vogt NM, Jonaitis EM, et al. Association of Neighborhood Context, Cognitive Decline, and Cortical Change in an Unimpaired Cohort. *Neurology*. 2021;96(20):e2500-e2512. doi:10.1212/WNL.0000000000011918

23. Pohl DJ, Seblova D, Avila JF, et al. Relationship between Residential Segregation, Later-Life Cognition, and Incident Dementia across Race/Ethnicity. *International Journal of Environmental Research and Public Health*. 2021;18(21):11233. doi:10.3390/ijerph182111233

24. Jang JB, Hicken MT, Mullins M, et al. Racial Segregation and Cognitive Function Among Older Adults in the United States: Findings From the Reasons for Geographic and Racial Differences in Stroke Study. *J Gerontol B Psychol Sci Soc Sci*. 2022;77(6):1132-1143. doi:10.1093/geronb/gbab107

25. Forrester SN, Gallo JJ, Whitfield KE, Thorpe RJ. A Framework of Minority Stress: From Physiological Manifestations to Cognitive Outcomes. *Gerontologist*. 2019;59(6):1017-1023. doi:10.1093/geront/gny104

26. Schulz A, Northridge ME. Social determinants of health: implications for environmental health promotion. *Health Educ Behav*. 2004;31(4):455-471. doi:10.1177/1090198104265598

27. Auchincloss AH, Moore KA, Moore LV, Roux AVD. Improving retrospective characterization of the food environment for a large region in the United States during a historic time period. *Health & place*. 2012;18(6):1341-1347.

28. Powell LM, Chaloupka FJ, Slater SJ, Johnston LD, O'Malley PM. The availability of local-area commercial physical activity-related facilities and physical activity among adolescents. *Am J Prev Med*. 2007;33(4 Suppl):S292-300. doi:10.1016/j.amepre.2007.07.002

29. Gordon-Larsen P, Nelson MC, Page P, Popkin BM. Inequality in the built environment underlies key health disparities in physical activity and obesity. *Pediatrics*. 2006;117(2):417-424. doi:10.1542/peds.2005-0058

30. Finlay J, Esposito M, Li M, et al. Can Neighborhood Social Infrastructure Modify Cognitive Function? A Mixed-Methods Study of Urban-Dwelling Aging Americans. *J Aging Health*. 2021;33(9):772-785. doi:10.1177/08982643211008673

31. Needham BL, Smith JA, Zhao W, et al. Life course socioeconomic status and DNA methylation in genes related to stress reactivity and inflammation: The multi-ethnic study of atherosclerosis. *Epigenetics*. 2015;10(10):958-969. doi:10.1080/15592294.2015.1085139

32. Stringhini S, Polidoro S, Sacerdote C, et al. Life-course socioeconomic status and DNA methylation of genes regulating inflammation. *Int J Epidemiol*. 2015;44(4):1320-1330. doi:10.1093/ije/dyv060

33. Giurgescu C, Nowak AL, Gillespie S, et al. Neighborhood Environment and DNA Methylation: Implications for Cardiovascular Disease Risk. *J Urban Health*. 2019;96(1):23-34. doi:10.1007/s11524-018-00341-1

34. Smith JA, Zhao W, Wang X, et al. Neighborhood characteristics influence DNA methylation of genes involved in stress response and inflammation: The Multi-Ethnic Study of Atherosclerosis. *Epigenetics*. 2017;12(8):662-673. doi:10.1080/15592294.2017.1341026

35. Lei MK, Beach SRH, Simons RL, Philibert RA. Neighborhood crime and depressive symptoms among African American women: Genetic moderation and epigenetic mediation of effects. *Soc Sci Med*. 2015;146:120-128. doi:10.1016/j.socscimed.2015.10.035

36. Marioni RE, McRae AF, Bressler J, et al. Meta-analysis of epigenome-wide association studies of cognitive abilities. *Mol Psychiatry*. 2018;23(11):2133-2144. doi:10.1038/s41380-017-0008-y

37. McCartney DL, Hillary RF, Conole ELS, et al. Blood-based epigenome-wide analyses of cognitive abilities. *Genome Biology*. 2022;23(1):26. doi:10.1186/s13059-021-02596-5

38. Ammous F, Zhao W, Ratliff SM, et al. Epigenome-wide association study identifies DNA methylation sites associated with target organ damage in older African Americans. *Epigenetics*. 2021;16(8):862-875. doi:10.1080/15592294.2020.1827717

39. Yang Y, Knol MJ, Wang R, et al. Epigenetic and integrative cross-omics analyses of cerebral white matter hyperintensities on MRI. *Brain*. Published online August 9, 2022:awac290. doi:10.1093/brain/awac290

40. Chu SH, Loucks EB, Kelsey KT, et al. Sex-specific epigenetic mediators between early life social disadvantage and adulthood BMI. *Epigenomics*. 2018;10(6):707-722.

41. Loucks EB, Huang YT, Agha G, et al. Epigenetic mediators between childhood socioeconomic disadvantage and mid-life body mass index: the New England family study. *Psychosomatic medicine*. 2016;78(9):1053.

42. Wang YZ, Zhao W, Ammous F, et al. DNA Methylation Mediates the Association Between Individual and Neighborhood Social Disadvantage and Cardiovascular Risk Factors. *Front Cardiovasc Med*. 2022;9:848768. doi:10.3389/fcvm.2022.848768

43. Mayeda ER, Glymour MM, Quesenberry CP, Whitmer RA. Inequalities in dementia incidence between six racial and ethnic groups over 14 years. *Alzheimers Dement*. 2016;12(3):216-224. doi:10.1016/j.jalz.2015.12.007

44. Barnes LL, Bennett DA. Alzheimer's disease in African Americans: risk factors and challenges for the future. *Health Aff (Millwood)*. 2014;33(4):580-586. doi:10.1377/hlthaff.2013.1353

45. Brewster P, Barnes L, Haan M, et al. Progress and future challenges in aging and diversity research in the United States. *Alzheimers Dement*. 2019;15(7):995-1003. doi:10.1016/j.jalz.2018.07.221

46. Rajan KB, Weuve J, Barnes LL, Wilson RS, Evans DA. Prevalence and incidence of clinically diagnosed Alzheimer's disease dementia from 1994 to 2012 in a population study. *Alzheimers Dement*. 2019;15(1):1-7. doi:10.1016/j.jalz.2018.07.216

47. Pickle LW, Mungiole M, Gillum RF. Geographic Variation in Stroke Mortality in Blacks and Whites in the United States. *Stroke*. 1997;28(8):1639-1647. doi:10.1161/01.STR.28.8.1639

48. Glymour MM, Manly JJ. Lifecourse social conditions and racial and ethnic patterns of cognitive aging. *Neuropsychol Rev*. 2008;18(3):223-254. doi:10.1007/s11065-008-9064-z

49. Bell ML, Ebisu K. Environmental Inequality in Exposures to Airborne Particulate Matter Components in the United States. *Environ Health Perspect*. 2012;120(12):1699-1704. doi:10.1289/ehp.1205201

50. Daniels PR, Kardia SLR, Hanis CL, et al. Familial aggregation of hypertension treatment and control in the Genetic Epidemiology Network of Arteriopathy (GENOA) study. *Am J Med*. 2004;116(10):676-681. doi:10.1016/j.amjmed.2003.12.032

51. Lezak P of NP and NMD, Lezak MD, Howieson AP of N and PDB, et al. *Neuropsychological Assessment*. Oxford University Press; 2004.

52. Smith JA, Mosley Jr TH, Turner ST, Kardia SL. Shared Genetic Effects among Measures of Cognitive Function and Leukoaraiosis. *Edited by Amit Agrawal*. Published online 2012:39.

53. Davies G, Armstrong N, Bis JC, 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). *Mol Psychiatry*. 2015;20(2):183-192. doi:10.1038/mp.2014.188

54. Jack CR, Twomey CK, Zinsmeister AR, Sharbrough FW, Petersen RC, Cascino GD. Anterior temporal lobes and hippocampal formations: normative volumetric measurements from MR images in young adults. *Radiology*. 1989;172(2):549-554. doi:10.1148/radiology.172.2.2748838

55. Smith JA, Turner ST, Sun YV, et al. Complexity in the genetic architecture of leukoaraiosis in hypertensive sibships from the GENOA Study. *BMC Med Genomics*. 2009;2:16. doi:10.1186/1755-8794-2-16

56. Jack CR, O'Brien PC, Rettman DW, et al. FLAIR histogram segmentation for measurement of leukoaraiosis volume. *J Magn Reson Imaging*. 2001;14(6):668-676. doi:10.1002/jmri.10011

57. Fortin JP, Fertig E, Hansen K. shinyMethyl: interactive quality control of Illumina 450k DNA methylation arrays in R. Published online September 19, 2014. doi:10.12688/f1000research.4680.2

58. Xu Z, Niu L, Li L, Taylor JA. ENmix: a novel background correction method for Illumina HumanMethylation450 BeadChip. *Nucleic Acids Research*. 2016;44(3):e20-e20. doi:10.1093/nar/gkv907

59. Fortin JP, Triche TJ Jr, Hansen KD. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. *Bioinformatics*. 2017;33(4):558-560. doi:10.1093/bioinformatics/btw691

60. Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*. 2014;30(10):1363-1369. doi:10.1093/bioinformatics/btu049

61. Niu L, Xu Z, Taylor JA. RCP: a novel probe design bias correction method for Illumina Methylation BeadChip. *Bioinformatics*. 2016;32(17):2659-2663. doi:10.1093/bioinformatics/btw285

62. Lehne B, Drong AW, Loh M, et al. A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. *Genome Biology*. 2015;16(1):37. doi:10.1186/s13059-015-0600-x

63. Pidsley R, Zotenko E, Peters TJ, et al. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biology*. 2016;17(1):208. doi:10.1186/s13059-016-1066-1

64. Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*. 2012;13(1):86. doi:10.1186/1471-2105-13-86

65. Du P, Zhang X, Huang CC, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics*. 2010;11(1):587. doi:10.1186/1471-2105-11-587

66. Weisenberger CDJ, Laird PW. Comprehensive DNA Methylation Analysis on the Illumina® Infinium® Assay Platform. :4.

67. Lockstone HE. Exon array data analysis using Affymetrix power tools and R statistical software. *Briefings in Bioinformatics*. 2011;12(6):634-644. doi:10.1093/bib/bbq086

68. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Research*. 2003;31(4):e15. doi:10.1093/nar/gng015

69. Dai M, Wang P, Boyd AD, et al. Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Research*. 2005;33(20):e175. doi:10.1093/nar/gni179

70. Saha A, Battle A. False positives in trans-eQTL and co-expression analyses arising from RNA-sequencing alignment errors. Published online November 28, 2018. doi:10.12688/f1000research.17145.1

71. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 2007;8(1):118-127. doi:10.1093/biostatistics/kxj037

72. Gao X, Jia M, Zhang Y, Breitling LP, Brenner H. DNA methylation changes of whole blood cells in response to active smoking exposure in adults: a systematic review of DNA methylation studies. *Clin Epigenet*. 2015;7(1):113. doi:10.1186/s13148-015-0148-3

73. Johnston K, Ver Hoef JM, Krivoruchko K, Lucas N. *Using ArcGIS Geostatistical Analyst*. Vol 380. Esri Redlands; 2001.

74. Desktop EA. Redlands. *CA: Environmental Systems Research Institute*. Published online 2011.

75. Silverman BW. *Density Estimation for Statistics and Data Analysis*. Routledge; 2018.

76. Hirsch JA, Moore KA, Clarke PJ, et al. Changes in the Built Environment and Changes in the Amount of Walking Over Time: Longitudinal Results From the Multi-Ethnic Study of Atherosclerosis. *American Journal of Epidemiology*. 2014;180(8):799-809. doi:10.1093/aje/kwu218

77. Hirsch JA, Moore KA, Barrientos-Gutierrez T, et al. Built environment change and change in BMI and waist circumference: Multi-ethnic Study of Atherosclerosis. *Obesity*. 2014;22(11):2450-2457. doi:10.1002/oby.20873

78. Census Tract Level State Maps of the Modified Retail Food Environment Index (mRFEI).

79. Bureau UC. Summary File 1 Dataset. Census.gov. Accessed November 7, 2022. https://www.census.gov/data/datasets/2000/dec/summary-file-1.html

80. Bureau UC. Summary File 3 Dataset. Census.gov. Accessed November 7, 2022. https://www.census.gov/data/datasets/2000/dec/summary-file-3.html

81. Bureau UC. American Community Survey (ACS). Census.gov. Accessed November 7, 2022. https://www.census.gov/programs-surveys/acs

82. Bureau UC. American Community Survey 3-Year Data (2007-2013). Census.gov. Accessed November 7, 2022. https://www.census.gov/data/developers/data-sets/acs-3year.html

83. Roux AVD, Merkin SS, Arnett D, et al. Neighborhood of Residence and Incidence of Coronary Heart Disease. *N Engl J Med*. 2001;345(2):99-106. doi:10.1056/NEJM200107123450205

84. Du J, Zhou X, Hao W, Liu Y, Smith JA, Mukherjee B. Methods for Large-scale Single Mediator Hypothesis Testing: Possible Choices and Comparisons. Published online March 24, 2022. doi:10.48550/arXiv.2203.13293

85. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1995;57(1):289-300. doi:10.1111/j.2517-6161.1995.tb02031.x

86. Cerin E. Building the evidence for an ecological model of cognitive health. *Health Place*. 2019;60:102206. doi:10.1016/j.healthplace.2019.102206

87. Wu YT, Prina AM, Jones A, Matthews FE, Brayne C. The Built Environment and Cognitive Disorders: Results From the Cognitive Function and Ageing Study II. *American Journal of Preventive Medicine*. 2017;53(1):25-32. doi:10.1016/j.amepre.2016.11.020

88. Cohn-Schwartz E. Pathways From Social Activities to Cognitive Functioning: The Role of Physical Activity and Mental Health. *Innov Aging*. 2020;4(3):igaa015. doi:10.1093/geroni/igaa015

89. Clarke PJ, Weuve J, Barnes L, Evans DA, Mendes de Leon CF. Cognitive decline and the neighborhood environment. *Annals of Epidemiology*. 2015;25(11):849-854. doi:10.1016/j.annepidem.2015.07.001

90. Besser LM, Rodriguez DA, McDonald N, et al. Neighborhood built environment and cognition in non-demented older adults: The Multi-Ethnic Study of Atherosclerosis. *Social Science & Medicine*. 2018;200:27-35. doi:10.1016/j.socscimed.2018.01.007

91. Magaziner J, Cadigan DA, Hebel JR, Parry RE. Health and living arrangements among older women: Does living alone increase the risk of illness? *Journal of Gerontology*. 1988;43:M127-M133. doi:10.1093/geronj/43.5.M127

92. Clarke PJ, Ailshire JA, House JS, et al. Cognitive function in the community setting: the neighbourhood as a source of ‘cognitive reserve’? *J Epidemiol Community Health*. 2012;66(8):730-736. doi:10.1136/jech.2010.128116

93. Besser LM. *Neighborhood Built Environment Characteristics and Cognition in Non-Demented Older Adults*. Ph.D. The University of North Carolina at Chapel Hill. Accessed April 15, 2023. https://www.proquest.com/docview/1917681682/abstract/12A28C3159BD447CPQ/1

94. Marottoli RA, de Leon CFM, Glass TA, Williams CS, Cooney LM Jr, Berkman LF. Consequences of Driving Cessation: Decreased Out-of-Home Activity Levels. *The Journals of Gerontology: Series B*. 2000;55(6):S334-S340. doi:10.1093/geronb/55.6.S334

95. Cassarino M, Setti A. Environment as ‘Brain Training’: A review of geographical and physical environmental influences on cognitive ageing. *Ageing Research Reviews*. 2015;23:167-182. doi:10.1016/j.arr.2015.06.003

96. Finlay J, Esposito M, Tang S, et al. Fast-food for thought: Retail food environments as resources for cognitive health and wellbeing among aging Americans? *Health & Place*. 2020;64:102379. doi:10.1016/j.healthplace.2020.102379

97. Finlay J, Esposito M, Kim MH, Gomez-Lopez I, Clarke P. Closure of ‘third places’? Exploring potential consequences for collective health and wellbeing. *Health & Place*. 2019;60:102225. doi:10.1016/j.healthplace.2019.102225

98. Buonocore JJ, Lee HJ, Levy JI. The Influence of Traffic on Air Quality in an Urban Neighborhood: A Community–University Partnership. *Am J Public Health*. 2009;99(Suppl 3):S629-S635. doi:10.2105/AJPH.2008.149138

99. Kwate NOA. Fried chicken and fresh apples: Racial segregation as a fundamental cause of fast food density in black neighborhoods. *Health & Place*. 2008;14(1):32-44. doi:10.1016/j.healthplace.2007.04.001

100. Perrin AJ, Caren N, Skinner AC, Odulana A, Perrin EM. The unbuilt environment: culture moderates the built environment for physical activity. *BMC Public Health*. 2016;16(1):1227. doi:10.1186/s12889-016-3866-3

101. Eliasoph N, Lichterman P. Culture in Interaction. *American Journal of Sociology*. 2003;108(4):735-794. doi:10.1086/367920

102. Huang YT. Genome-wide analyses of sparse mediation effects under composite null hypotheses. *The Annals of Applied Statistics*. 2019;13(1):60-84. doi:10.1214/18-AOAS1181

103. Zeng P, Shao Z, Zhou X. Statistical methods for mediation analysis in the era of high-throughput genomics: Current successes and future challenges. *Comput Struct Biotechnol J*. 2021;19:3209-3224. doi:10.1016/j.csbj.2021.05.042

104. Blum MGB, Valeri L, Fran çois O, et al. Challenges Raised by Mediation Analysis in a High-Dimension Setting. *Environmental Health Perspectives*. 2020;128(5):055001. doi:10.1289/EHP6240

105. Clark-Boucher D, Zhou X, Du J, et al. Methods for Mediation Analysis with High-Dimensional DNA Methylation Data: Possible Choices and Comparison. Published online February 14, 2023:2023.02.10.23285764. doi:10.1101/2023.02.10.23285764

106. Song Y, Zhou X, Zhang M, et al. Bayesian shrinkage estimation of high dimensional causal mediation effects in omics studies. *Biometrics*. 2020;76(3):700-710. doi:10.1111/biom.13189

107. Chén OY, Crainiceanu C, Ogburn EL, Caffo BS, Wager TD, Lindquist MA. High-dimensional multivariate mediation with application to neuroimaging data. *Biostatistics*. 2018;19(2):121-136. doi:10.1093/biostatistics/kxx027

108. Huang YT. Variance component tests of multivariate mediation effects under composite null hypotheses. *Biometrics*. 2019;75(4):1191-1204. doi:10.1111/biom.13073

109. 2021 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*. 2021;17(3):327-406. doi:10.1002/alz.12328

110. Tang MX, Cross P, Andrews H, et al. Incidence of AD in African-Americans, Caribbean Hispanics, and Caucasians in northern Manhattan. *Neurology*. 2001;56(1):49-56. doi:10.1212/wnl.56.1.49

111. Reitz C, Jun G, Naj A, Rajbhandary R. Variants in the ATP-Binding Cassette Transporter (ABCA7), Apolipoprotein E ϵ4,and the Risk of Late-Onset Alzheimer Disease in African Americans. *JAMA*. 2013;309(14):1483-1492. doi:10.1001/jama.2013.2973

3.8 Tables

Table 3-1. Sample characteristics of Genetic Epidemiology Network of Arteriopathy (GENOA) African Americans (N = 542)

	Mean (SD) or n%
Age at cognition measurement (years)	62.52 (7.69)
Sex	
Female	403 (74.35%)
Male	139 (25.65%)
Educational attainment	
Completed at least four years of college or technical/trade school	155 (28.60%)
Completed high school degree/GED	252 (46.49%)
Less than high school degree/GED	135 (24.91%)
Smoking status	
Current smoker	83 (15.31%)
Former smoker	125 (23.06%)
Never smoker	334 (61.62%)
General cognitive function	0.03 (0.99)
Delayed recall (RAVLT, number of words recalled)	6.95 (3.29)
Processing speed (DSST, number of symbols)	33.82 (13.04)
Word fluency (COWA-FAS, number of words)	29.40 (11.64)
Visual conceptual tracking (TMTA, seconds to test completion)	63.75 (35.22)
White matter hyperintensity (WMH, cm^3) [a]	9.42 (9.19)
Total intracranial volume (TIV, cm^3) [a]	1376.58 (129.81)
Neighborhood characteristics	
Neighborhood socioeconomic disadvantage	3.41 (3.46)
Fast food destination density [b]	0.75 (0.85)
Unfavorable food stores without alcohol density [b]	1.94 (1.75)
Unfavorable food stores with alcohol density [b]	1.24 (1.13)
Favorable food stores density [b]	0.22 (0.31)
Total physical activity destinations density [b]	0.34 (0.37)
Total social engagement destinations density [b]	14.37 (10.85)
Total popular walking destination density [b]	3.53 (3.13)
Alcoholic drinking places density [b]	0.36 (0.62)
Total stores density [b]	15.82 (12.80)
Total food stores density [b]	3.34 (3.08)
MRFEI with alcohol [c]	0.10 (0.13)
MRFEI without alcohol [c]	0.12 (0.14)

Abbreviations: RAVLT, Rey Auditory Verbal Learning Test; DSST, Digit Symbol Substitution Test; COWA-FAS, Controlled Oral Word Association Test; TMTA, Trail Making Test A; WMH, White Matter Hyperintensity; MRFEI, Modified Retail Food Environment Index

Abbreviations: RAVLT, Rey Auditory Verbal Learning Test; DSST, Digit Symbol Substitution Test; COWA-FAS, Controlled Oral Word Association Test; TMTA, Trail Making Test A; WMH, White Matter Hyperintensity

a. Sample size = 466.
b. Simple density measures per square mile for 1-mile buffer size.
c. Derived from simple density measures per square mile for 1-mile buffer size.

Table 3-2. Associations between neighborhood socioeconomic disadvantage and cognitive function/ White matter hyperintensity

	General cognitive function (N=477)				White matter hyperintensity (N=404)			
	Model 1a		Model 2a		Model 1b		Model 2b	
	β	P	β	P	β	P	β	P
Neighborhood socioeconomic disadvantage	-0.01	0.30	-0.01	0.36	2.0E-3	0.83	0.01	0.28

Model 1a: cognitive function = neighborhood socioeconomic disadvantage + age at measurement + sex + PC1-4 + education + smoking status + familial relatedness (random effect)

Model 2a: cognitive function = Model 1a + census tract population density + census tract (random effect)

Model 1b: WMH = Model 1a + total intracranial volume

Model 2b: WMH = Model 2a + total intracranial volume

*P<0.05

Table 3-3. Associations between neighborhood socioeconomic disadvantage and cognitive measures (Model 2a; N=477)

	DSST		COWA-FAS		RAVLT		TMTA	
	β	P	β	P	β	P	β	P
Neighborhood socioeconomic disadvantage	-0.01	0.95	0.02	0.92	-0.03	0.66	0.02	0.07

Abbreviations: DSST, Digit Symbol Substitution Test; COWA-FAS, Controlled Oral Word Association Test; RAVLT, Rey Auditory Verbal Learning Test; TMTA, Trail Making Test A.

Model 2a: neurocognitive measure = neighborhood socioeconomic disadvantage + age at measurement + sex + PC1-4 + education + smoking status + population density + familial relatedness (random effect) + census tract (random effect)

*P<0.05

Table 3-4. Associations between simple density of neighborhood destinations per square mile for 1-mile buffer size and cognitive function/WMH

Neighborhood characteristics	General cognitive function				White matter hyperintensity			
	Model 1a (N=542)		Model 2a (N=477)		Model 1b (N=466)		Model 2b (N=404)	
	β	P	β	P	β	P	β	P
Fast food destination density	-0.02	0.53	-0.03	0.39	0.03	0.23	0.04	0.25
Unfavorable food stores without alcohol density	-0.02	0.38	-0.02	0.37	0.01	0.40	0.02	0.24
Unfavorable food stores with alcohol density	-0.03	0.26	-0.05	0.14	0.02	0.26	0.03	0.25
Favorable food stores density	-0.08	0.45	-0.11	0.31	0.02	0.83	-0.01	0.84
Total physical activity destinations density	-0.07	0.36	-0.05	0.58	0.03	0.65	0.05	0.58
Total social engagement destinations density	-3.16E-03	0.29	-3.59E-03	0.35	1.59E-03	0.49	3.46E-03	0.24
Total popular walking destination density	-3.75E-03	0.71	-2.49E-03	0.84	0.01	0.38	0.01	0.25
Alcoholic drinking places density	-0.01	0.78	0.01	0.89	1.86E-03	0.99	0.03	0.52
Total stores density	-1.47E-03	0.49	-2.80E-03	0.36	7.55E-04	0.66	2.90E-03	0.21
Total food stores density	-5.15E-03	0.63	-3.80E-03	0.77	2.21E-03	0.78	8.61E-03	0.37
Modified Retail Food Environment Index with alcohol	-0.10	0.73	-0.13	0.69	0.17	0.41	0.08	0.74
Modified Retail Food Environment Index without alcohol	-0.02	0.93	-0.05	0.85	0.10	0.58	0.03	0.90

Model 1a: cognitive function = neighborhood characteristic + age at measurement + PC1-4 + sex + education + smoking status + familial relatedness (random effect)
Model 2a: cognitive function = Model 1a + neighborhood socioeconomic disadvantage + census tract population density (random effect) + census tract (random effect)

Model 1b: WMH = Model 1a + total intracranial volume

Model 2b: WMH = Model 2a + total intracranial volume

*P<0.05

Table 3-5. Associations between simple density of neighborhood destinations per square mile for 1-mile buffer size and cognitive measures (Model 2a; N=477)

Neighborhood characteristics	DSST		COWA-FAS		RAVLT		TMTA	
	β	P	β	P	β	P	β	P
Fast food destination density	-0.39	0.45	0.27	0.63	0.10	0.57	0.05	**0.04***
Unfavorable food stores without alcohol density	-0.17	0.55	-0.19	0.52	0.13	0.18	0.02	0.19
Unfavorable food stores with alcohol density	-0.45	0.28	-0.07	0.87	0.01	0.94	0.04	**0.04***
Favorable food stores density	-1.46	0.30	0.20	0.89	-0.31	0.52	0.12	0.08
Total physical activity destinations density	-1.07	0.39	-1.18	0.37	0.44	0.30	0.05	0.38
Total social engagement destinations density	-0.06	0.26	-0.03	0.61	0.02	0.25	2.14E-03	0.36
Total popular walking destination density	-0.05	0.77	0.02	0.88	0.09	0.09	0.01	0.20
Alcoholic drinking places density	0.16	0.85	-0.93	0.28	0.62	**0.03***	-3.11E-03	0.94
Total stores density	-0.05	0.19	-0.02	0.67	0.02	0.24	1.44E-03	0.44
Total food stores density	-0.01	0.95	-0.11	0.53	0.10	0.07	0.01	0.41
Modified Retail Food Environment Index with alcohol	-3.56	0.36	4.15	0.32	-0.64	0.65	0.20	0.28
Modified Retail Food Environment Index without alcohol	-3.29	0.36	4.43	0.25	0.55	0.66	0.20	0.21

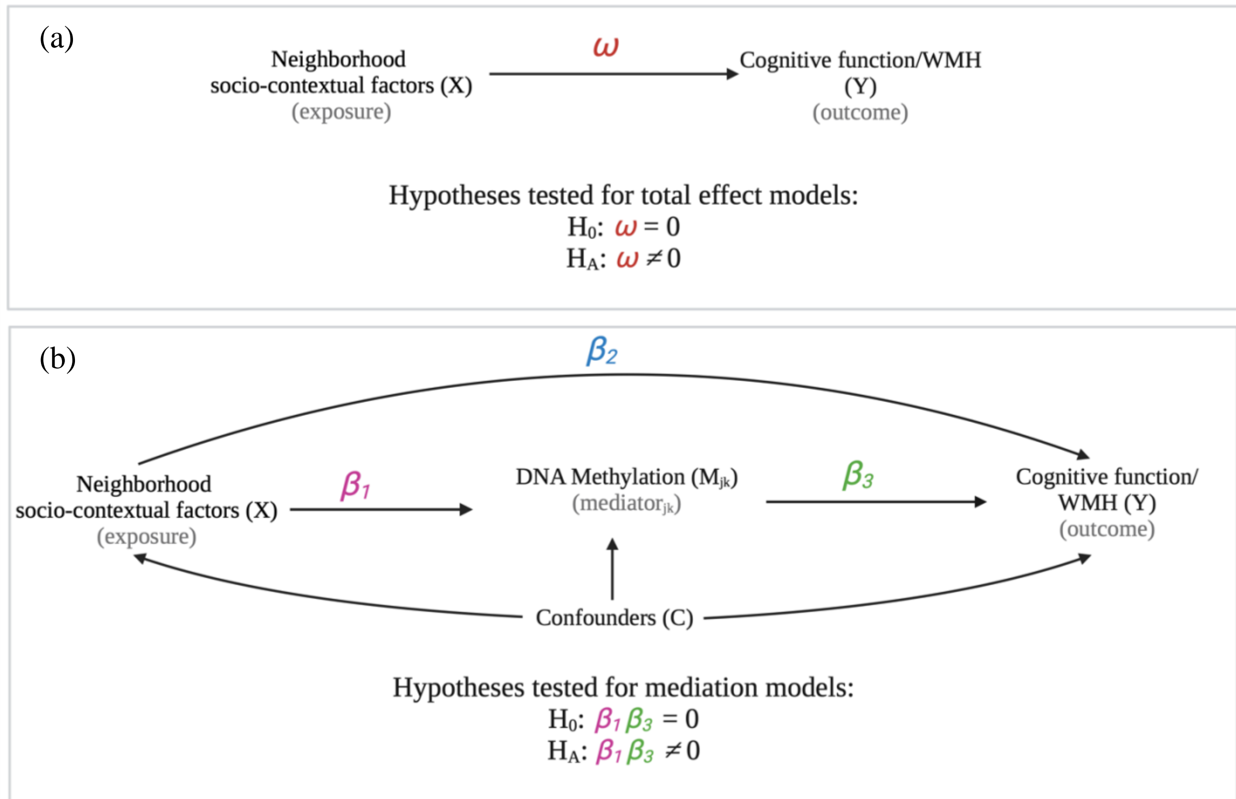
Abbreviations: DSST, Digit Symbol Substitution Test; COWA-FAS, Controlled Oral Word Association Test; RAVLT, Rey Auditory Verbal Learning Test; TMTA, Trail Making Test A

Model 2a: neurocognitive measure = neighborhood characteristic + age at measurement + PC1-4 + sex + education + smoking status + neighborhood socioeconomic disadvantage + census tract population density + familial relatedness (random effect) + census tract (random effect)

*P<0.05

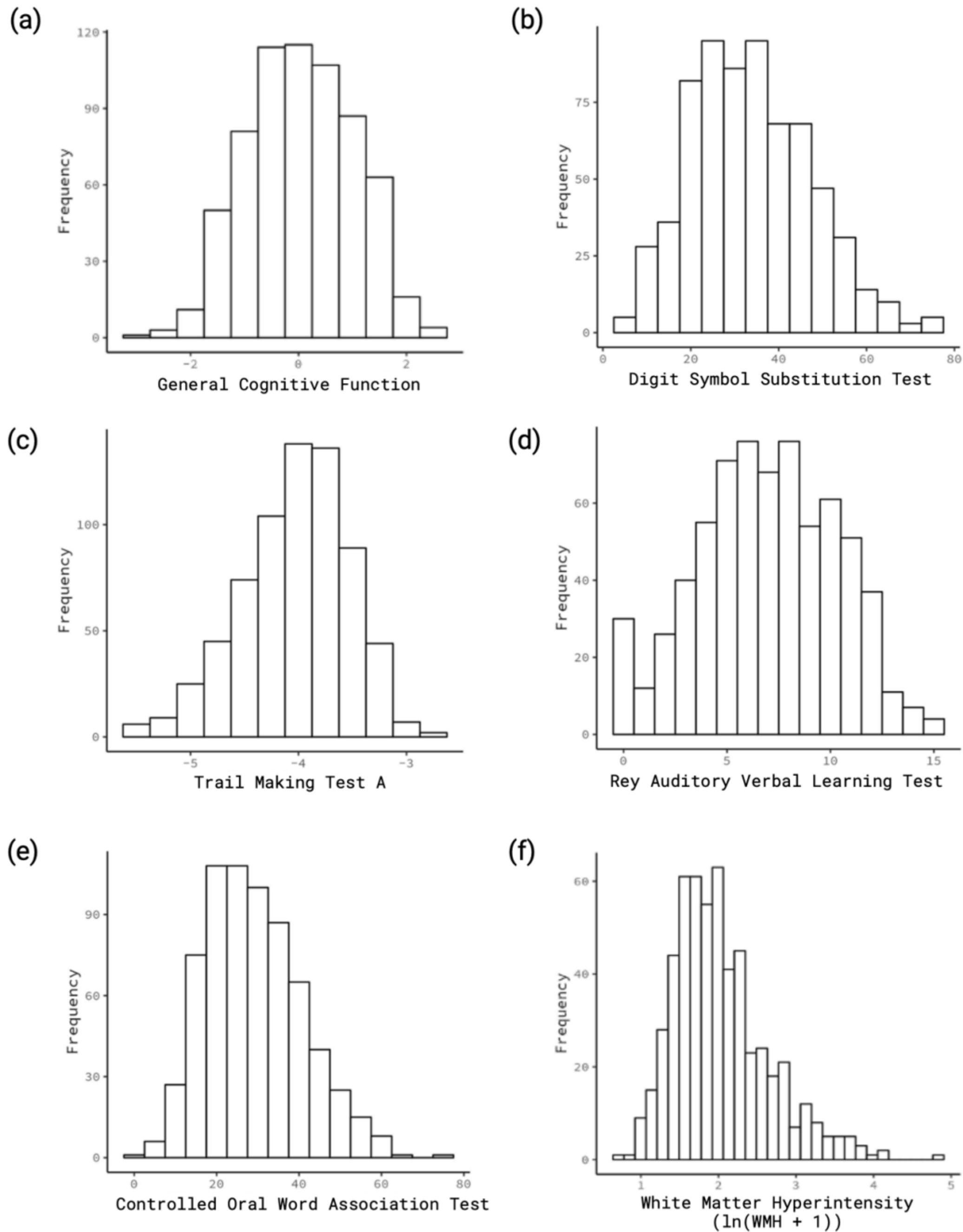
3.9 Figures

Figure 3-1. Directed acyclic graph (DAG) of the hypothesized associations for the epigenetic mediation between neighborhood characteristics (exposures) and cognitive/WMH outcomes.



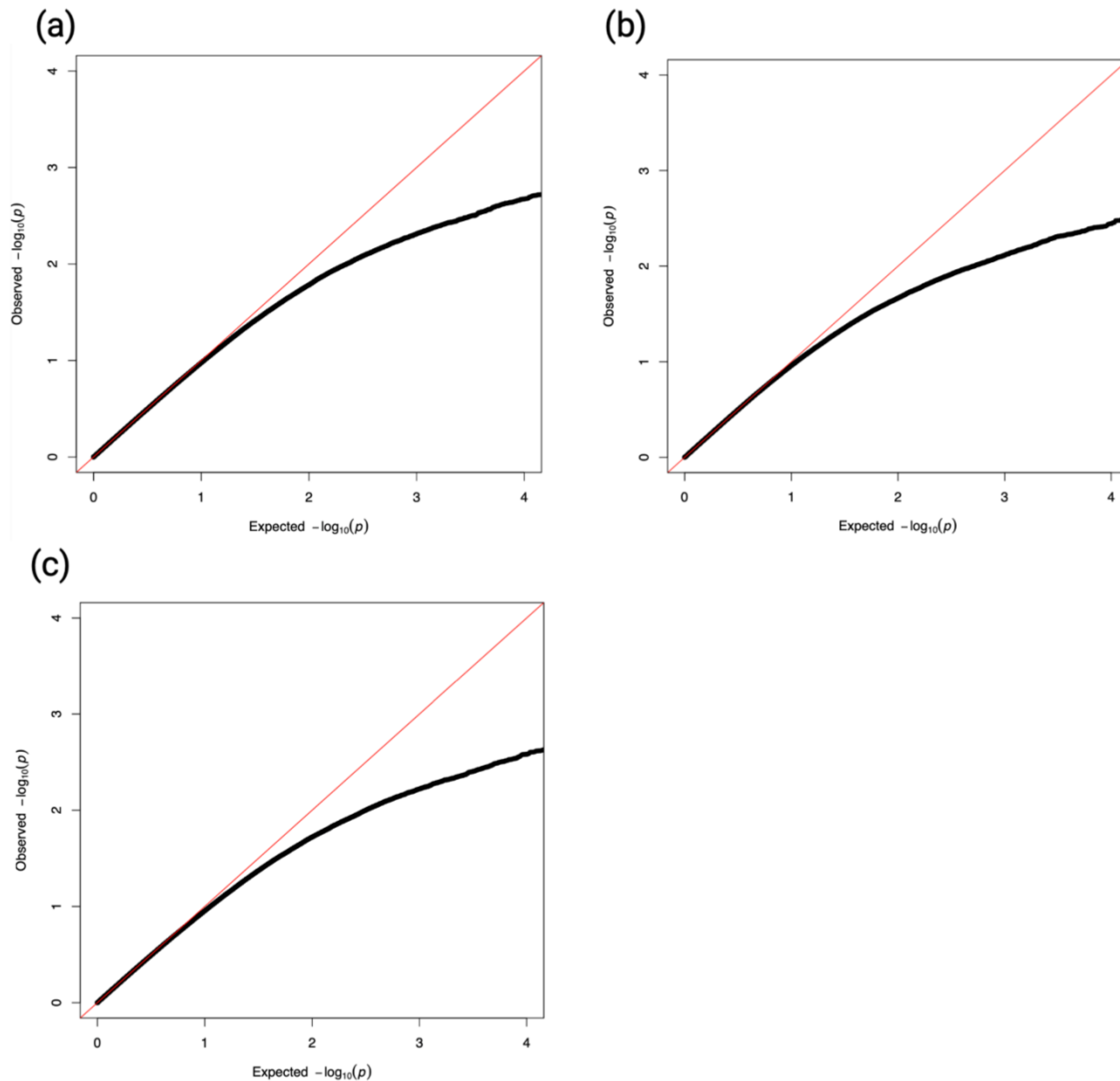
Directed acyclic graph (DAG) of the hypothesized associations for the epigenetic mediation between neighborhood characteristics (exposures) and cognitive/WMH outcomes. (a) The total effect associations between neighborhood characteristic (X) and cognitive function/WMH (Y). ω is the effect estimate of the neighborhood characteristic on cognitive function/WMH. (b) The mediation effect obtained through the cross-product-based mediation approach obtained by multiplying the exposure-mediator effect (β_1) and the mediator-outcome effect (β_3).

Figure 3-2. Distributions of cognitive and structural brain measures



Distributions of cognitive and structural brain measures. (a) General cognitive function, (b) Digit symbol substitution test, (c) Trail making test A, (d) Rey auditory verbal learning test, (e) Controlled oral word association test and (f) Log-transformed white matter hyperintensity (ln(WMH+1))

Figure 3-3. Quantile-quantile plots for the epigenetic mediation of the associations between neighborhood characteristics and cognitive function.



Quantile-quantile (QQ) plots for the epigenetic mediation of the associations between neighborhood characteristics and cognitive function. QQ plots for the Sobel-Comp mediation hypothesis testing method with N=477 observations. The exposures are simple densities per square mile for 1-mile buffer sizes, the outcomes are neurocognitive measures, and the mediators are 857,121 CpG sites. The exposure – outcome models tested are as follows: (a) alcohol drinking places density – RAVLT, (b) fast food destination density – TMTA, and (c) unfavorable food stores (with alcohol) density – TMTA. Mediation models are adjusted for age, sex, education, smoking status, first four principal components, neighborhood socioeconomic disadvantage, and census tract population density, with family and census tracts as random effects.

3.10 Supplementary Methods

General cognitive function

The following four cognitive domains were evaluated a year after Phase II, on average, as part of GMBI:[1,2]

1. The Weschler Adult Intelligence Scale-Revised: Digit Symbol Substitution Test (DSST) measured complex visual attention, sustained and focused concentration, response speed and visuomotor coordination. The DSST measures executive function of working memory.[3] In this test, participants matched symbols to numbers according to a key located at the top of the page. The DSST score comprised the number of symbols correctly matched within 90 seconds. Scores ranged from 3 symbols to 88 symbols correctly matched within 90 seconds.
2. The Controlled Oral Word Association Test (COWA-FAS) tested for verbal fluency (phonetic association) and language. This test requires participants to name as many animals as possible that start with the letters F, A, and S in 1 minute. The score consisted of the total number of admissible animal names generated.
3. The Rey Auditory Verbal Learning Test (RAVLT) measured delayed recall, relating to the cognitive functions of new learning, immediate memory span and vulnerability to interference in learning and recognition memory. Its score was determined by the number of words recalled after a 30-minute delay. Scores ranged from 0 to 15.
4. The Trail Making Test A (TMTA) evaluated visual conceptual tracking as participants need to connect a set of 25 circles quickly and accurately. TMTA provides information on the cognitive functions of visual search, scanning, processing speed and executive functions. The

natural logarithm of seconds to completion for the task was used and recoded so that higher scores indicate better cognitive function. The maximum was 240 seconds to complete.

Neighborhood environment exposures

1. **GIS-based measures**

Population densities of recreational, social, and healthy food environments were derived from GIS[5] data using Dun and Bradstreet data as compiled by Walls and Associates in the National Establishment Series (NETS) database[6] for 1996-2015. Addresses were geocoded using the TeleAtlas EZ-Locate web-based geocoding software (Tele Atlas North America, Inc., Lebanon, New Hampshire). NETS yearly datasets were categorized based on Standard Industrial Classification (SIC) codes. Densities per square mile were created for 0.5-,1-, and 3-mile buffers around the home addresses of GENOA participants at Phase I using ArcGIS V.9.3 (ESRI, Inc., Redlands, California).[7,8] Densities were calculated using two approaches: 1) simple densities per square mile within the buffer region and 2) kernel densities per square mile within the buffer region, with greater weighting towards resources located closer to the home of a participant. Total density scores by category were created by adding together densities from each type of establishment.

For each participant, we estimated the densities for the following destinations: fast-food restaurants (chain and non-chain), total physical activity facilities, total social engagement destinations, and alcohol outlets. Summary density measures were also created for densities of unfavorable food stores (with and without alcohol), healthy (favorable) food stores, popular walking destinations, total stores, and total food stores.

Fast food restaurants are places that specialize in low preparation time foods that are eaten cafeteria-style or take-away (SIC #581203, except for coffee shops (#58120304)). Physical activity facilities measure was created using 114 SIC codes consisting of the recreational and physical activity establishments such as indoor conditioning, dance, bowling, golf, team and racquet sports, and water activities derived from lists used in previous studies.[11,12] Healthy food availability was defined using healthy food stores such as fruit and vegetable markets (SIC #5431) and supermarkets (grocery stores (SIC #5411) with at least $2 million in annual sales or at least 25 employees or name being on standardized supermarket chain name lists as described in other studies).[13] Social engagement destinations, consisting of places which promote social interaction, were derived from 430 SIC codes based on previous work.[14,15] These SIC codes include locations such as beauty shops and barbers, sports entertainment, exercise facilities, amusements, libraries, museums and art galleries, religious organizations, eating and dining places. Alcohol outlets were identified as liquor stores and on-site drinking places (restaurants and nightclubs/bars).

Categories for favorable food stores consisted of supermarkets (chain and non-chain) and fruit and vegetable markets. Unfavorable food stores (without alcohol) included convenience stores, bakeries/nuts/candy/ice cream stores, and fast-food restaurants (chain and non-chain). Unfavorable food stores with alcohol included alcohol outlets. Popular walking destinations were created from six different categories including postal service, drug stores and pharmacy, banks and credit unions, food stores (non-beverage), eating and dining places (non-beverage) and drinking places (non-alcoholic). Total stores variable was created by summing food stores, recreational facilities, popular walking destinations (non-food- and food-based), and social

engagement (non-food- and food-based). Total food stores variable was calculated from the sum of favorable food stores, neutral food stores, unfavorable food stores and other eating places.

The modified retail food environment index (MRFEI) measured the number of healthy and less healthy food retailers within census tracts across states, based on typical food offerings in specific retail stores.[16] The MREI was a proportion calculated as the number of favorable food stores divided by the total of favorable and unfavorable stores (with and without alcohol outlets). The MRFEI represents the proportion of all food retailers in a given census tract that are healthy and ranges from 0 or "food desert" (e.g., no healthy food vendors) to 1 or "healthy" food vendors only. MRFEI variables were calculated for 0.5-, 1- and 3-mile buffer regions.

2. Census-based measures

Neighborhood socioeconomic disadvantage was assessed using data collected in the 2000 U.S. Census,[17,18] American Community Survey (ACS) 2005-2009,[19] and ACS 2007-2011[20] estimates. Data was linked to GENOA participant data (Phase I; 1995-2000) by Census tract using Census and ACS estimates for the closest time period. A composite index was previously developed using factor analysis to determine which socioeconomic indicator variables from the Census can be meaningfully combined into a summary score. Six variables representing the dimensions of wealth and income (log of the median household income; log of the median value of housing units; and percent of household with interest, dividend or net rental income), education (the percentage of adults 25 years of age or older who had completed high school and the percentage of adults 25 years of age or older who had completed college (i.e., Bachelor's degree)), and occupation (the percentage of employed persons 16 years of age or older in executive, managerial or professional specialty occupations) were used to characterize

neighborhood socioeconomic disadvantage for each census tract.[21] Z-scores for each census tract were estimated for each variable, and neighborhood socioeconomic disadvantage was defined as the sum of Z-scores from the six variables, with higher scores indicating more disadvantage.

References

1. Lezak, P. of N.P. and N.M.D., Lezak, M.D., Howieson, A.P. of N. and P.D.B., Howieson, D.B., Loring, P. of N.D.W., Loring, D.W., and Fischer, J.S. (2004). Neuropsychological Assessment (Oxford University Press).

2. Smith, J.A., Mosley Jr, T.H., Turner, S.T., and Kardia, S.L. (2012). Shared Genetic Effects among Measures of Cognitive Function and Leukoaraiosis. Edited by Amit Agrawal, 39.

3. Jaeger, J. (2018). Digit Symbol Substitution Test. J Clin Psychopharmacol *38*, 513–519. 10.1097/JCP.0000000000000941.

4. Davies, G., Armstrong, N., Bis, J.C., Bressler, J., Chouraki, V., Giddaluru, S., Hofer, E., Ibrahim-Verbaas, C.A., Kirin, M., Lahti, J., et al. (2015). Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N=53 949). Mol Psychiatry *20*, 183–192. 10.1038/mp.2014.188.

5. Johnston, K., Ver Hoef, J.M., Krivoruchko, K., and Lucas, N. (2001). Using ArcGIS geostatistical analyst (Esri Redlands).

6. Associates, W.& (2014). National establishment time-series (nets) database (Walls & Associates Oakland).

7. Desktop, E.A. (2011). Redlands. CA: Environmental Systems Research Institute.

8. Silverman, B.W. (2018). Density estimation for statistics and data analysis (Routledge).

9. Hirsch, J.A., Moore, K.A., Clarke, P.J., Rodriguez, D.A., Evenson, K.R., Brines, S.J., Zagorski, M.A., and Diez Roux, A.V. (2014). Changes in the Built Environment and Changes in the Amount of Walking Over Time: Longitudinal Results From the Multi-Ethnic Study of Atherosclerosis. American Journal of Epidemiology *180*, 799–809. 10.1093/aje/kwu218.

10. Hirsch, J.A., Moore, K.A., Barrientos-Gutierrez, T., Brines, S.J., Zagorski, M.A., Rodriguez, D.A., and Diez Roux, A.V. (2014). Built environment change and change in BMI and waist circumference: Multi-ethnic Study of Atherosclerosis. Obesity *22*, 2450–2457. 10.1002/oby.20873.

11. Powell, L.M., Chaloupka, F.J., Slater, S.J., Johnston, L.D., and O′Malley, P.M. (2007). The availability of local-area commercial physical activity-related facilities and physical activity among adolescents. Am J Prev Med *33*, S292-300. 10.1016/j.amepre.2007.07.002.

12. Gordon-Larsen, P., Nelson, M.C., Page, P., and Popkin, B.M. (2006). Inequality in the built environment underlies key health disparities in physical activity and obesity. Pediatrics *117*, 417–424. 10.1542/peds.2005-0058.

13. Auchincloss, A.H., Moore, K.A., Moore, L.V., and Roux, A.V.D. (2012). Improving retrospective characterization of the food environment for a large region in the United States during a historic time period. Health & place *18*, 1341−1347.

14. Hoehner, C.M., and Schootman, M. (2010). Concordance of Commercial Data Sources for Neighborhood-Effects Studies. J Urban Health *87*, 713–725. 10.1007/s11524-010-9458-0.

15. Hirsch, J.A., Grengs, J., Schulz, A., Adar, S.D., Rodriguez, D.A., Brines, S.J., and Diez Roux, A.V. (2016). How much are built environments changing, and where?: Patterns of change by neighborhood sociodemographic characteristics across seven U.S. metropolitan areas. Social Science & Medicine *169*, 97−105. 10.1016/j.socscimed.2016.09.032.

16. Census Tract Level State Maps of the Modified Retail Food Environment Index (mRFEI).

17. Bureau, U.C. Summary File 1 Dataset. Census.gov. https://www.census.gov/data/datasets/2000/dec/summary-file-1.html.

18. Bureau, U.C. Summary File 3 Dataset. Census.gov. https://www.census.gov/data/datasets/2000/dec/summary-file-3.html.

19. Bureau, U.C. American Community Survey (ACS). Census.gov. https://www.census.gov/programs-surveys/acs.

20. Bureau, U.C. American Community Survey 3-Year Data (2007-2013). Census.gov. https://www.census.gov/data/developers/data-sets/acs-3year.html.

21. Roux, A.V.D., Merkin, S.S., Arnett, D., Chambless, L., Massing, M., Nieto, F.J., Sorlie, P., Szklo, M., Tyroler, H.A., and Watson, R.L. (2001). Neighborhood of Residence and Incidence of Coronary Heart Disease. N Engl J Med *345*, 99–106. 10.1056/NEJM200107123450205.

3.11 Supplementary Material

Table S3-6. Pearson's correlations among the cognitive/WMH outcomes (N=466)

	General cognitive function	RAVLT	DSST	COWA-FAS	TMTA	WMH
General cognitive function	1					
RAVLT	0.597***	1				
DSST	0.897***	0.425***	1			
COWA-FAS	0.670***	0.214**	0.483***	1		
TMTA	0.792***	0.290***	0.681***	0.304***	1	
WMH	-0.335***	-0.276***	-0.322***	-0.119	-0.272***	1

Abbreviations: RAVLT: Rey Auditory Verbal Learning Test; DSST: Digit Symbol Substitution Task; COWA-FAS: Controlled Oral Word Association Test; TMTA: Trail Making Test A; WMH: White Matter Hyperintensity
* $p<0.05$, ** $p<0.01$, *** $p<0.001$

Table 3-7. Pearson's correlations among neighborhood socioeconomic disadvantage and neighborhood simple density measures per square mile for 1-mile buffer size (N=542)

Neighbor-hood characterist-ics	Neighborhood -Socio-economic Dis-advantage	Fast Food destin-ation den-sity	Unfavor-able food stores without alcohol density	Unfavor-able food stores with alcohol density	Favor-able food stores density	Total physical activity destinate-ions density	Total social engage-ment destinate-ions density	Total popular walking destinate-ion density	Alcohol-ic drinking places density	Total stores density	Total food stores density	MRFEI with alcohol	MR-FEI with out alco-hol
Neighborhood Socio-economic Disadvantage	1.00												
Fast Food destination density	-0.20 **	1.00											
Unfavorable food stores without alcohol density	0.33 ***	0.60 ***	1.00										
Unfavorable food stores with alcohol density	-0.21 ***	0.93 ***	0.56 ***	1.00									
Favorable food stores density	4.00E-3	0.37 ***	0.15*	0.37 ***	1.00								
Total physical activity destinations density	0.14*	0.26 ***	0.34 ***	0.18 **	0.44 ***	1.00							
Total social engagement destinations density	0.56 ***	0.38 ***	0.68 ***	0.34 ***	0.40 ***	0.56 ***	1.00						
Total popular walking destination density	0.19 **	0.59 ***	0.52 ***	0.65 ***	0.47 ***	0.49 ***	0.72 ***	1.00					
Alcoholic drinking places density	0.54 ***	-0.05	0.72 ***	-0.16*	-0.11	0.27 ***	0.52 ***	0.08	1.00				
Total stores density	-1.00E-3	0.23 ***	-0.05	0.24 ***	0.92 ***	0.38 ***	0.30 ***	0.36 ***	-0.24 ***	1.00			
Total food stores density	0.17 **	0.18 **	0.13*	0.15*	0.89 ***	0.41 ***	0.43 ***	0.40 ***	0.05	0.91 ***	1.00		
MRFEI with alcohol	0.55 ***	0.39 ***	0.68 ***	0.37 ***	0.43 ***	0.55 ***	0.99 ***	0.77 ***	0.49 ***	0.32 ***	0.45 ***	1.00	
MRFEI without alcohol	0.45 ***	0.43 ***	0.84 ***	0.43 ***	0.18 **	0.45 ***	0.78 ***	0.76 ***	0.64 ***	0.04	0.24 ***	0.78 ***	1.00

Abbreviations: MRFEI, Modified Retail Food Environment Index

* p<0.05, **p<0.01, ***p<0.001

Table S3-8. Associations among neighborhood socioeconomic disadvantage and neighborhood simple density measures per square mile for 1-mile buffer size after adjusting for census tract population density (N=542)

Neighborhood characteristics	Neighborhood Socioeconomic Disadvantage	Fast Food dest. density	Unfav. food stores without alcohol density	Unfav. food stores with alcohol density	Favorable food stores density	Total physical activity dest. density	Total social engagement dest. Density	Total popular walking dest. density	Alcoholic drinking places density	Total stores density	Total food stores density	MRFEI with alcohol	MRFEI without alcohol
Neighborhood Socioeconomic Disadvantage	1.00												
Fast Food dest. density	-0.02***	1.00											
Unfav. food stores without alcohol density	0.68***	0.29***	1.00										
Unfav. food stores with alcohol density	0.92***	-0.05	0.68***	1.00									
Favorable food stores density	0.38***	0.14**	0.14***	0.39***	1.00								
Total physical activity dest. density	0.36***	0.11*	0.36***	0.28***	0.43***	1.00							
Total social engagement dest. density	0.52***	0.39***	0.71***	0.52***	0.40***	0.59***	1.00						
Total popular walking dest. density	0.68***	0.06	0.58***	0.75***	0.49***	0.52***	0.74***	1.00					
Alcoholic drinking places density	0.11*	0.42***	0.74***	0.02	-0.15**	0.26***	0.51***	0.11*	1.00				
Total stores density	0.53***	0.36***	0.70***	0.55***	0.44***	0.59***	0.99***	0.79***	0.47***	1.00			
Total food stores density	0.60***	0.29***	0.89***	0.61***	0.19***	0.43***	0.78***	0.77***	0.65***	0.78***	1.00		
MRFEI with alcohol	0.17***	0.16***	-0.12*	0.15***	0.86***	0.33***	0.23***	0.28***	-0.28***	0.25***	-0.03	1.00	
MRFEI without alcohol	0.16***	0.25***	-0.01	0.12*	0.85***	0.32***	0.28***	0.31***	-0.12*	0.30***	0.09	0.94***	1.00

Abbreviations: unfav., unfavorable; dest., destinations; MRFEI, Modified Retail Food Environment Index
* $p<0.05$, ** $p<0.01$, *** $p<0.001$

Table S3-9. Pearson's correlations among neighborhood socioeconomic disadvantage and simple and kernel densities per square mile for 1-mile buffer size (N=542)[a]

		Kernel density measures										
	Neigh-borhood character-istics	Fast Food dest.	Unfav. food stores without alcohol	Unfav. food stores with alcohol	Favor-able food stores	Total physical activity dest.	Total social engagement dest.	Total popular walking dest.	Alcoholic drinking places	Total stores	Total food stores	MRFEI with alcohol	MRFEI without alcohol
Simple density measures	Fast Food dest.	0.78***	0.46***	0.77***	0.34***	0.23***	0.34***	0.48***	-0.04	0.21***	0.14*	0.35***	0.30***
	Unfav. food stores without alcohol)	0.59***	0.86***	0.56***	0.05	0.38***	0.72***	0.49***	0.68***	-0.10	-0.04	0.71***	0.73***
	Unfav. food stores with alcohol	0.69***	0.38***	0.78***	0.35***	0.10	0.25***	0.48***	-0.16*	0.25***	0.18**	0.27***	0.26***
	Favorable food stores	0.30***	0.09	0.31***	0.70***	0.27***	0.29***	0.33***	-0.14*	0.71***	0.68***	0.32***	0.09
	Total physical activity dest.	0.20**	0.27***	0.14*	0.26***	0.72***	0.47***	0.35***	0.28***	0.25***	0.24***	0.47***	0.32***
	Total social engagement dest.	0.30***	0.53***	0.24***	0.22***	0.42***	0.87***	0.59***	0.52***	0.19**	0.29***	0.88***	0.63***
	Total popular walking dest.	0.37***	0.33***	0.39***	0.37***	0.18**	0.50***	0.78***	0.09	0.32***	0.41***	0.56***	0.56***
	Alcoholic drinking places	0.14*	0.70***	0.03	-0.22***	0.38***	0.65***	0.18**	0.93***	-0.31***	-0.19**	0.62***	0.66***
	Total stores	0.16*	-0.09	0.14*	0.67***	0.18**	0.15*	0.23***	-0.24***	0.82***	0.77***	0.17**	-0.03
	Total food stores	0.21***	0.12*	0.16**	0.60***	0.26***	0.36***	0.34***	0.04	0.69***	0.74***	0.38***	0.21***
	MRFEI with alcohol	0.30***	0.516***	0.246***	0.248***	0.392***	0.853***	0.621***	0.48***	0.222***	0.32***	0.868***	0.629***
	MRFEI without alcohol	0.34***	0.673***	0.31***	0.064	0.292***	0.688***	0.660***	0.634***	-0.019	0.128*	0.701***	0.839***

a. Values with grey shading correspond to the correlations between kernel and simple densities per square mile for 1-mile buffer size

Abbreviations: unfav., unfavorable; dest., destinations; MRFEI, Modified Retail Food Environment Index

* p<0.05, **p<0.01, ***p<0.001

Table S3-10. Associations between simple density of neighborhood destinations per square mile for ½-, 1- and 3- mile buffer sizes and cognitive function/WMH

Neighborhood characteristics	Buffer Size	General cognitive function				White matter hyperintensity			
		Model 1a (N=542)		Model 2a (N=477)		Model 1b (N=466)		Model 2b (N=404)	
		β	P	β	P	β	P	β	P
Fast Food destination density	1/2-mile	-0.01	0.55	-0.01	0.61	9.82E-04	0.95	-0.01	0.75
	1-mile	-0.02	0.53	-0.03	0.39	0.03	0.23	0.04	0.23
	3-mile	0.03	0.70	9.56E-05	0.99	-0.10	0.10	-0.05	0.48
Unfavorable food stores without alcohol density	1/2-mile	-0.01	0.17	-0.01	0.22	2.37E-03	0.75	-3.67E-04	0.96
	1-mile	-0.02	0.38	-0.02	0.37	0.01	0.40	0.02	0.24
	3-mile	-0.02	0.69	-0.04	0.40	-0.03	0.32	-4.72E-03	0.91
Unfavorable food stores with alcohol density	1/2-mile	-0.02	0.28	-0.02	0.28	0.01	0.58	-8.83E-04	0.95
	1-mile	-0.03	0.26	-0.05	0.14	0.02	0.26	0.03	0.24
	3-mile	0.01	0.89	-0.01	0.86	-0.05	0.19	-0.02	0.72
Favorable food stores density	1/2-mile	0.03	0.57	0.02	0.69	-0.01	0.90	-0.01	0.79
	1-mile	-0.08	0.45	-0.11	0.31	0.02	0.83	-0.01	0.90
	3-mile	-0.04	0.85	-0.12	0.68	-0.13	0.46	9.32E-04	0.99
Total physical activity destinations density	1/2-mile	-0.02	0.59	-0.01	0.91	3.58E-03	0.92	-0.05	0.19
	1-mile	-0.07	0.36	-0.05	0.58	0.03	0.65	0.05	0.53
	3-mile	0.01	0.96	-4.34E-03	0.98	-0.15	0.23	-0.03	0.83
Total social engagement destinations density	1/2-mile	-2.02E-03	0.27	-1.91E-03	0.40	-4.65E-04	0.75	-5.88E-04	0.74
	1-mile	-3.16E-03	0.29	-3.59E-03	0.35	1.59E-03	0.49	3.46E-03	0.24
	3-mile	-4.13E-03	0.39	-0.01	0.25	-1.57E-03	0.67	1.14E-03	0.81
Total popular walking destination density	1/2-mile	-1.39E-03	0.82	6.43E-05	0.99	1.11E-03	0.81	-1.56E-03	0.78
	1-mile	-3.75E-03	0.71	-2.49E-03	0.84	0.01	0.38	0.01	0.25
	3-mile	-3.79E-03	0.47	-0.01	0.63	-0.01	0.38	-2.14E-03	0.88
Alcoholic drinking places density	1/2-mile	-0.02	0.50	-0.01	0.65	9.17E-04	0.96	3.21E-03	0.88
	1-mile	-0.01	0.78	0.01	0.89	1.86E-03	0.96	0.03	0.53
	3-mile	-0.40	0.12	-0.71	**0.03***	0.04	0.85	0.08	0.74
Total stores density	1/2-mile	-9.52E-04	0.48	-1.67E-03	0.38	-8.86E-04	0.41	-4.61E-04	0.75

	1-mile	-1.47E-03	0.49	-2.80E-03	0.36	7.55E-04	0.66	2.95E-03	0.21
	3-mile	-1.31E-03	0.68	-0.01	0.23	-1.08E-03	0.67	1.26E-03	0.72
Total food stores density	1/2-mile	-4.19E-03	0.47	-2.84E-03	0.67	-4.58E-04	0.92	-6.68E-04	0.90
	1-mile	-0.01	0.63	-3.80E-03	0.77	2.21E-03	0.66	0.01	0.38
	3-mile	-4.73E-03	0.82	-0.02	0.51	-0.02	0.33	-3.81E-03	0.85
Modified Retail Food Environment Index with alcohol	1/2-mile	0.20	0.54	0.10	0.78	0.14	0.51	0.20	0.39
	1-mile	-0.10	0.73	-0.13	0.69	0.17	0.41	0.08	0.71
	3-mile	-0.22	0.66	-0.17	0.82	3.28E-04	0.99	0.27	0.61
Modified Retail Food Environment Index without alcohol	1/2-mile	0.34	0.25	0.26	0.40	0.06	0.74	0.11	0.59
	1-mile	-0.02	0.93	-0.05	0.85	0.10	0.58	0.03	0.88
	3-mile	-0.07	0.88	-0.01	0.99	0.16	0.64	0.30	0.53

Model 1a: Cognitive function = age at measurement+ PC1-4+ sex+ education+ smoking status + family (random effect)
Model 1b: WMH = Model 1a + total intracranial volume
Model 2a: Cognitive function = Model 1a + neighborhood socioeconomic disadvantage + census tract population density + family (random effect) + census tracts (random effect)
Model 2b: WMH = Model 2a + total intracranial volume
*P<0.05

Table S3-11. Associations between simple density of neighborhood destinations per square mile for ½-, 1- and 3- mile buffer sizes and cognitive measures (N=477)

Neighborhood characteristics	Buffer Size	DSST		COWA-FAS		RAVLT		TMTA	
		β	P	β	P	β	P	β	P
Fast Food destination density	1/2- mile	-0.25	0.37	0.01	0.97	0.09	0.34	0.02	0.11
	1-mile	-0.39	0.45	0.27	0.63	0.10	0.57	0.05	**0.04***
	3-mile	0.17	0.90	-0.61	0.66	0.42	0.37	0.02	0.74
Unfavorable food stores without alcohol density	½- mile	-0.16	0.24	-0.17	0.25	0.05	0.29	0.01	0.07
	1-mile	-0.17	0.55	-0.19	0.52	0.13	0.18	0.02	0.19
	3-mile	-0.50	0.47	-0.4	0.58	0.11	0.63	0.03	0.32
Unfavorable food stores with alcohol density	½- mile	-0.25	0.27	-0.04	0.86	-4.54E-03	0.95	0.02	0.07
	1-mile	-0.45	0.28	-0.07	0.87	0.01	0.94	0.04	**0.04***
	3-mile	0.01	0.99	-0.43	0.64	0.27	0.37	0.02	0.63
Favorable food stores density	1/2- mile	0.65	0.41	0.92	0.28	-0.04	0.89	0.02	0.60
	1-mile	-1.46	0.30	0.2	0.89	-0.31	0.52	0.12	0.08
	3-mile	-0.62	0.87	-2.59	0.51	0.39	0.76	0.08	0.66
Total physical activity destinations density	1/2- mile	-0.49	0.47	-0.62	0.38	0.08	0.73	-0.02	0.48
	1-mile	-1.07	0.39	-1.18	0.37	0.44	0.30	0.05	0.38
	3-mile	0.54	0.84	-1.23	0.66	1.28	0.16	0.04	0.76
Total social engagement destinations density	1/2- mile	-0.03	0.24	-0.04	0.23	0.01	0.27	1.23E-03	0.35
	1-mile	-0.06	0.26	-0.03	0.61	0.02	0.25	0.00	0.36
	3-mile	-0.07	0.41	-0.07	0.42	0.01	0.78	3.87E-03	0.32
Total popular walking destination density	1/2- mile	4.97E-03	0.96	-0.09	0.41	0.05	0.13	3.92E-03	0.35
	1-mile	-0.05	0.77	0.02	0.88	0.09	0.09	0.01	0.20
	3-mile	-0.04	0.88	-0.14	0.59	0.06	0.49	0.01	0.50
Alcoholic drinking places density	1/2- mile	-0.27	0.46	-0.44	0.27	0.25	**4.75E-02***	0.01	0.56
	1-mile	0.16	0.85	-0.93	0.28	0.62	**0.03***	-3.11E-03	0.94
	3-mile	-9.57	**0.03***	-2.33	0.61	-0.80	0.59	0.42	**0.04***
Total stores density	1/2- mile	-0.03	0.21	-0.03	0.24	0.01	0.29	9.66E-04	0.38
	1-mile	-0.05	0.19	-0.02	0.67	0.02	0.24	1.44E-03	0.44
	3-mile	-0.07	0.28	-0.05	0.41	0.01	0.76	2.16E-03	0.46
Total food stores density	1/2- mile	-0.03	0.74	-0.14	0.14	0.06	**4.83E-02***	4.73E-03	0.22

	1-mile	-0.01	0.95	-0.11	0.53	0.10	0.07	0.01	0.41
	3-mile	-0.17	0.63	-0.2	0.60	0.07	0.56	0.01	0.42
Modified Retail Food Environment Index with alcohol	1/2- mile	4.21	0.35	5.87	0.21	-0.69	0.63	0.14	0.53
	1-mile	-3.56	0.36	4.15	0.32	-0.64	0.65	0.20	0.28
	3-mile	0.26	0.98	-6.93	0.45	1.33	0.64	0.12	0.77
Modified Retail Food Environment Index without alcohol	1/2- mile	5.31	0.18	6.2	0.14	0.07	0.96	0.07	0.72
	1-mile	-3.29	0.36	4.43	0.25	0.55	0.66	0.20	0.21
	3-mile	0.20	0.98	-0.91	0.91	0.62	0.81	0.06	0.86

Abbreviations: DSST, Digit Symbol Substitution Test; COWA-FAS, Controlled Oral Word Association Test; RAVLT, Rey Auditory Verbal Learning Test; TMTA, Trail Making Test A.
Model 2a: cognitive measure = age at measurement + PC1-4 + sex + education + smoking status + neighborhood socioeconomic disadvantage + census tract population density + family (random effect) + census tracts (random effect)
*P<0.05

Table S3-12. Associations between kernel density of neighborhood destinations per square mile for ½-, 1- and 3- mile buffer sizes and cognitive function/WMH

Neighborhood characteristics	Buffer Size	General cognitive function				White matter hyperintensity			
		Model 1a (N=542)		Model 2a (N=477)		Model 1b (N=466)		Model 2b (N=404)	
		β	P	β	P	β	P	β	P
Fast Food destination density	1/2-mile	0.01	0.68	4.35E-03	0.77	-0.01	0.54	-0.01	0.46
	1-mile	-0.03	0.31	-0.04	0.22	0.01	0.57	0.01	0.62
	3-mile	-0.05	0.44	-0.10	0.20	-1.83E-03	0.97	0.04	0.45
Unfavorable food stores without alcohol density	1/2-mile	-1.52E-03	0.84	-7.92E-04	0.92	5.75E-04	0.92	-1.38E-03	0.82
	1-mile	-0.01	0.25	-0.02	0.26	0.01	0.57	0.01	0.53
	3-mile	-0.04	0.26	-0.06	0.14	-2.68E-03	0.92	0.02	0.46
Unfavorable food stores with alcohol density	1/2-mile	-2.05E-03	0.86	-1.71E-03	0.89	2.22E-04	0.98	-2.70E-03	0.77
	1-mile	-0.03	0.22	-0.03	0.16	0.01	0.38	0.01	0.42
	3-mile	-0.04	0.36	-0.07	0.18	-2.24E-03	0.95	0.03	0.43
Favorable food stores density	1/2-mile	0.02	0.68	-3.47E-03	0.93	-0.01	0.75	-0.01	0.70
	1-mile	-0.04	0.58	-0.10	0.25	0.01	0.85	0.01	0.90
	3-mile	-0.18	0.32	-0.31	0.15	-3.02E-03	0.98	0.07	0.64
Total physical activity destinations density	1/2-mile	0.03	0.33	0.04	0.28	0.02	0.51	-0.02	0.48
	1-mile	-0.02	0.80	-4.32E-03	0.95	2.18E-03	0.96	-0.04	0.42
	3-mile	-0.17	0.22	-0.22	0.19	-0.03	0.76	0.06	0.63
Total social engagement destinations density	1/2-mile	5.47E-04	0.68	8.09E-04	0.59	3.76E-06	1.00	1.91E-05	0.99
	1-mile	-2.53E-03	0.28	-3.37E-03	0.25	-4.83E-05	0.98	4.20E-04	0.85
	3-mile	-4.79E-03	0.23	-0.01	0.10	-1.09E-04	0.97	2.09E-03	0.58
Total popular walking destination density	1/2-mile	1.30E-03	0.78	2.24E-04	0.97	-1.75E-03	0.65	-3.44E-03	0.42
	1-mile	-3.50E-03	0.66	-0.01	0.56	2.27E-03	0.71	2.75E-03	0.71
	3-mile	-0.01	0.38	-0.02	0.18	-2.08E-03	0.84	0.01	0.54
Alcoholic drinking places density	1/2-mile	0.01	0.76	0.01	0.70	-1.27E-03	0.93	-0.01	0.74
	1-mile	-0.02	0.68	-0.01	0.82	-0.01	0.78	-0.01	0.87
	3-mile	-0.12	0.33	-0.14	0.43	-0.03	0.80	-0.01	0.95
Total stores density	1/2-mile	4.61E-04	0.70	5.97E-04	0.66	2.48E-05	0.98	-3.57E-05	0.97
	1-mile	-2.06E-03	0.31	-2.80E-03	0.28	1.14E-04	0.94	4.61E-04	0.82
	3-mile	-3.95E-03	0.24	-0.01	0.10	-6.11E-05	0.98	1.87E-03	0.55
	1/2-mile	1.35E-04	0.98	1.04E-03	0.84	-1.86E-03	0.61	-2.57E-03	0.51

Total food stores density	1-mile	-0.01	0.48	-0.01	0.51	1.47E-04	0.98	2.36E-03	0.74
	3-mile	-0.02	0.34	-0.03	0.20	-4.96E-03	0.71	0.01	0.62
Modified Retail Food Environment Index with alcohol	1/2-mile	0.34	0.25	0.03	0.92	0.17	0.38	0.28	0.18
	1-mile	-0.03	0.92	-0.14	0.59	0.18	0.32	0.20	0.29
	3-mile	-0.31	0.45	-0.70	0.19	0.12	0.69	0.23	0.56
Modified Retail Food Environment Index without alcohol	1/2-mile	0.51	0.05	0.24	0.43	0.15	0.40	0.27	0.18
	1-mile	0.04	0.85	-0.08	0.75	0.09	0.59	0.12	0.50
	3-mile	-0.20	0.60	-0.42	0.37	0.19	0.49	0.21	0.54

Model 1a: Cognitive function = age at measurement+ PC1-4+ sex+ education+ smoking status + family (random effect)
Model 1b: WMH = Model 1a + total intracranial volume
Model 2a: Cognitive function = Model 1a + neighborhood socioeconomic disadvantage + census tract population density + family (random effect) + census tracts (random effect)
Model 2b: WMH = Model 2a + total intracranial volume
*P<0.05

Table 3-13. Associations between kernel density of neighborhood destinations per square mile for ½-, 1- and 3- mile buffer sizes and cognitive measures (N=477)

Neighbor-hood character-istics	Buffer Size	DSST		COWA-FAS		RAVLT		TMTA	
		β	P	β	P	β	P	β	P
Fast Food destin-ation density	1/2-mile	-0.13	0.49	0.20	0.33	0.11	0.08	0.01	0.14
	1-mile	-0.49	0.20	0.04	0.93	0.08	0.52	0.05	**0.01***
	3-mile	-1.16	0.23	-0.10	0.92	-0.04	0.90	0.08	0.09
Unfavor-able food stores (without alcohol) density	1/2-mile	-0.07	0.50	0.05	0.67	0.07	0.05	0.01	0.13
	1-mile	-0.15	0.44	-0.15	0.48	0.08	0.19	0.02	**0.03***
	3-mile	-0.71	0.19	-0.27	0.64	0.06	0.74	0.04	0.11
Unfavor-able food stores (with alcohol) density	1/2-mile	-0.15	0.32	0.16	0.33	0.05	0.34	0.01	0.14
	1-mile	-0.35	0.24	-0.03	0.92	0.03	0.78	0.04	**0.01***
	3-mile	-0.80	0.23	-0.26	0.72	0.02	0.92	0.05	0.12
Favorable food stores density	1/2-mile	0.13	0.80	0.29	0.62	-0.07	0.70	0.02	0.47
	1-mile	-0.76	0.49	-0.06	0.96	-0.43	0.25	0.08	0.11
	3-mile	-2.68	0.32	-1.35	0.64	-1.05	0.26	0.14	0.30
Total physical activity destin-ations density	1/2-mile	-0.09	0.85	0.24	0.65	0.24	0.13	-0.01	0.51
	1-mile	-0.63	0.52	-0.40	0.70	0.39	0.23	0.01	0.89
	3-mile	-3.30	0.13	-2.59	0.27	0.76	0.31	0.15	0.15
Total social engage-ment destin-ations density	1/2-mile	-0.01	0.54	0.01	0.68	0.02	**0.01***	9.26E-04	0.29
	1-mile	-0.05	0.23	-0.04	0.32	0.02	0.12	3.24E-03	0.07
	3-mile	-0.09	0.16	-0.07	0.31	1.44E-03	0.95	4.38E-03	0.18
Total popular walking destin-ation density	1/2-mile	-0.03	0.68	0.01	0.91	0.04	0.07	0.01	**0.04***
	1-mile	-0.03	0.78	-0.08	0.55	0.07	0.08	0.01	**0.03***
	3-mile	-0.21	0.31	-0.21	0.35	0.03	0.69	0.01	0.17
Alcoholic drinking places density	1/2-mile	-1.03E-04	1.00	0.05	0.85	0.24	**0.01***	0.01	0.46
	1-mile	-0.07	0.90	-0.56	0.34	0.37	0.06	0.01	0.61
	3-mile	-1.55	0.48	-1.36	0.56	0.28	0.73	0.06	0.60

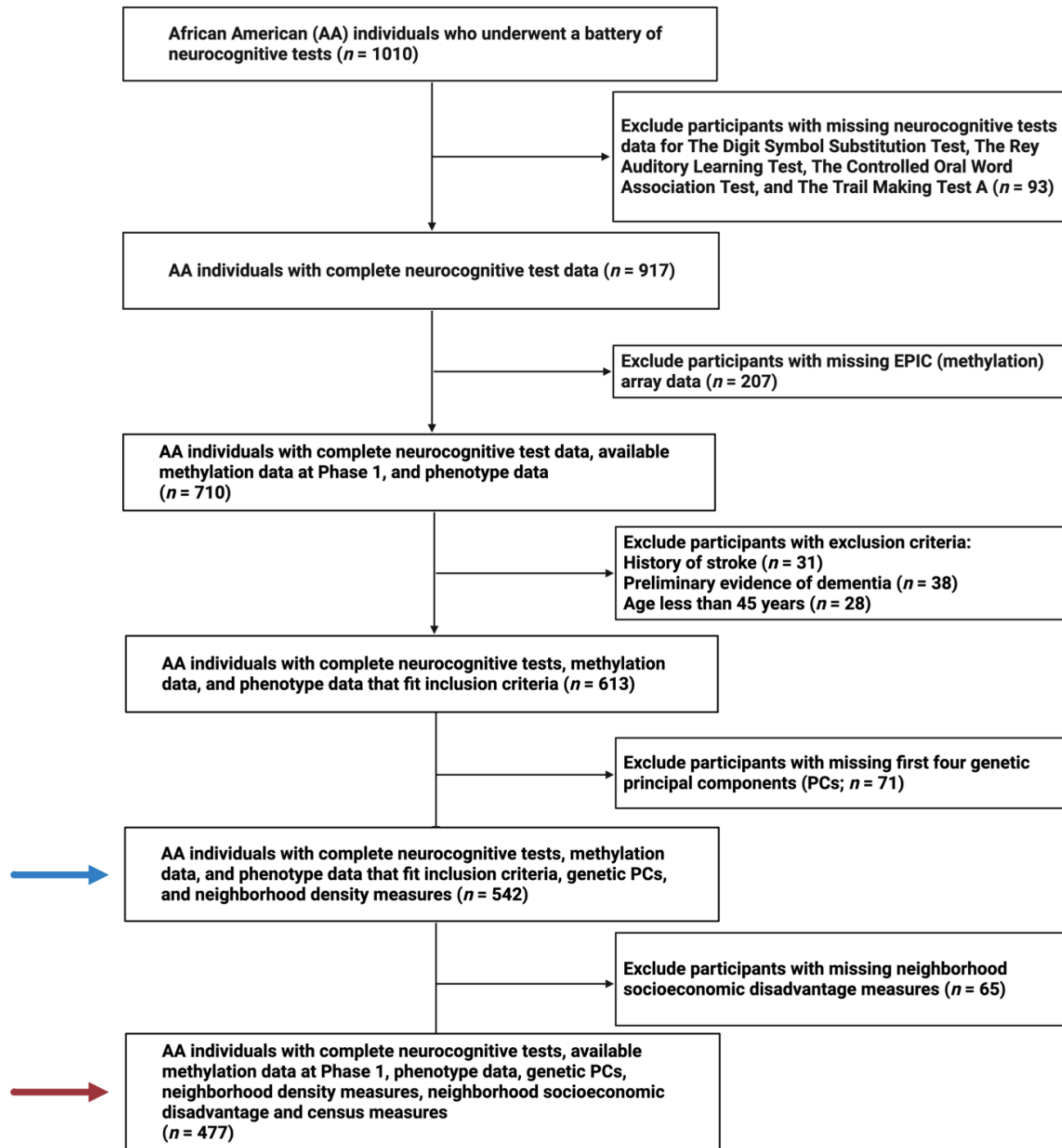
Total stores density	1/2-mile	-0.01	0.56	0.01	0.69	0.02	**0.01***	1.09E-03	0.17
	1-mile	-0.04	0.28	-0.03	0.35	0.02	0.10	2.95E-03	0.06
	3-mile	-0.07	0.17	-0.06	0.32	1.83E-03	0.92	3.69E-03	0.17
Total food stores density	1/2-mile	-0.01	0.85	0.01	0.94	0.05	**0.03***	4.61E-03	0.10
	1-mile	-0.02	0.86	-0.13	0.31	0.08	0.05	0.01	**0.04***
	3-mile	-0.28	0.33	-0.23	0.45	0.05	0.60	0.02	0.16
MRFEI with alcohol	1/2-mile	2.51	0.54	4.54	0.29	-1.08	0.42	0.04	0.83
	1-mile	-3.75	0.25	3.10	0.39	-0.97	0.41	0.08	0.61
	3-mile	-6.17	0.36	-7.32	0.30	-1.40	0.53	0.27	0.40
MRFEI without alcohol	1/2-mile	4.86	0.19	6.06	0.14	-0.59	0.63	-0.02	0.91
	1-mile	-3.20	0.30	3.77	0.26	-0.31	0.77	0.09	0.54
	3-mile	-3.41	0.57	-3.16	0.61	-0.99	0.62	0.18	0.53

Abbreviations: DSST, Digit Symbol Substitution Test; COWA-FAS, Controlled Oral Word Association Test; RAVLT, Rey Auditory Verbal Learning Test; TMTA, Trail Making Test A; MRFEI, Modified Retail Food Environment Index

Model 2a: cognitive measure = age at measurement + PC1-4 + sex + education + smoking status + neighborhood socioeconomic disadvantage + census tract population density + family (random effect) + census tracts (random effect)

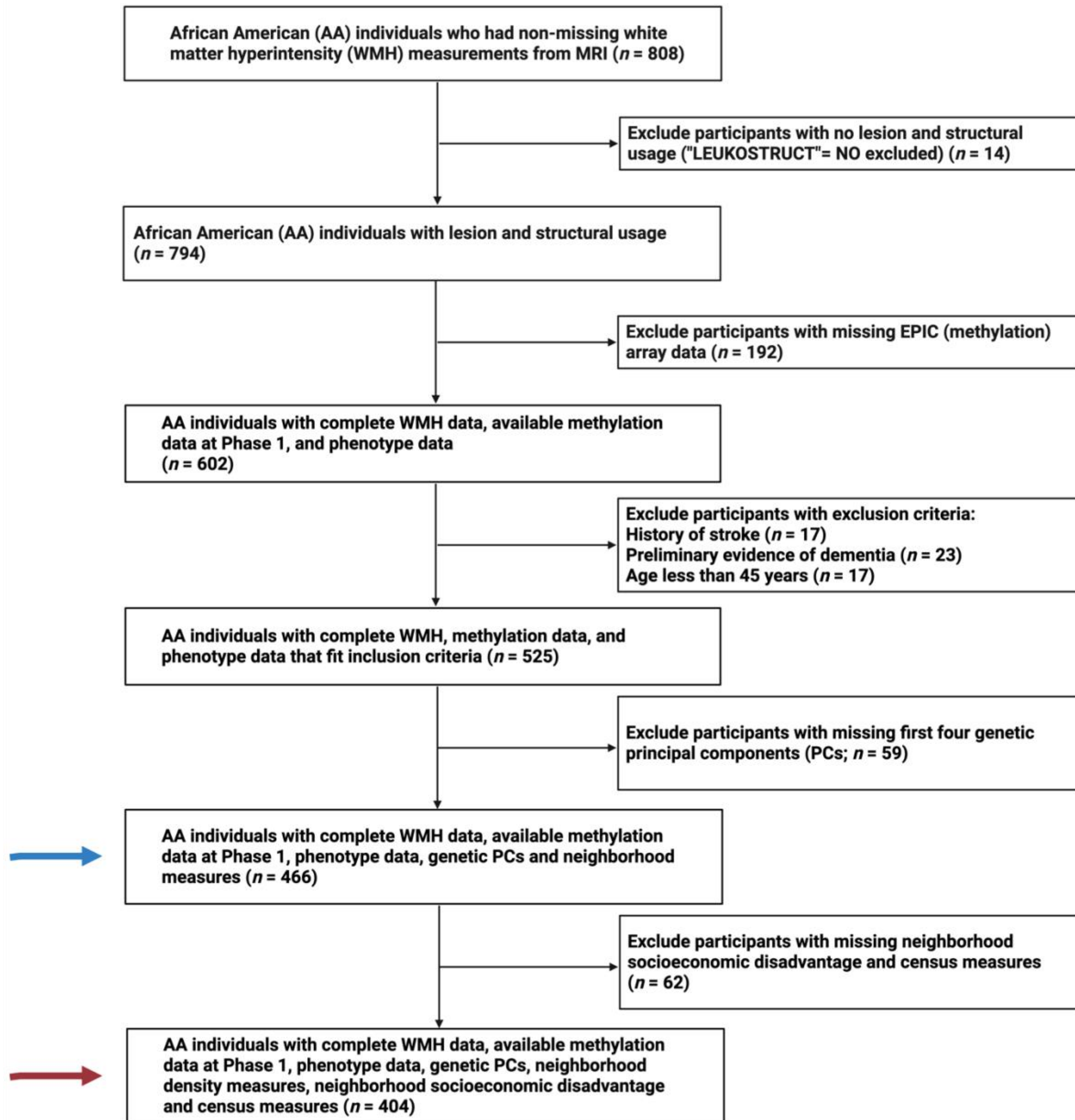
*P<0.05

Figure 3-4. Flow diagram illustrating sample sizes for neighborhood density and neighborhood socioeconomic disadvantage analyses for cognitive measures in GENOA African Americans.



Flow diagram illustrating sample sizes for neighborhood density and neighborhood socioeconomic disadvantage analyses for cognitive measures in GENOA AA.
Flow diagram illustrating sample sizes for neighborhood density analyses (Model 1a, n=542; blue arrow) and neighborhood socioeconomic disadvantage analyses (Model 2a, n=477; red arrow) for cognitive measures in GENOA AA.

Figure 3-5. Flow diagram illustrating sample sizes for neighborhood density and neighborhood socioeconomic disadvantage analyses for white matter hyperintensity in GENOA African Americans.



Flow diagram illustrating sample sizes for neighborhood density and neighborhood socioeconomic disadvantage analyses for white matter hyperintensity in GENOA AA. Flow diagram illustrating sample sizes for neighborhood density analyses, (Model 1b, n=466; blue arrow) and neighborhood socioeconomic disadvantage analyses (Model 2b, n=404; red arrow) for white matter hyperintensity in GENOA AA.

Chapter 4 . Multi-Ancestry Transcriptome-Wide Association Studies of Cognitive Function, White Matter Hyperintensity, and Alzheimer's Disease

4.1 Abstract

Genetic variants increase the risk of neurocognitive disorders in later life including Vascular Dementia (VaD) and Alzheimer's disease (AD), but the precise relationships between genetic risk factors and underlying disease etiology are not well understood. Transcriptome-wide association studies (TWAS) can be leveraged to better characterize the genes and biological pathways underlying genetic influences on disease. To date, almost all existing TWAS have been conducted using expression studies from individuals of a single genetic ancestry, primarily European. Using the joint likelihood-based inference framework in Multi-ancEstry TRanscriptOme-wide analysis (METRO), we leveraged gene expression data from European (EA) and African ancestries (AA) to identify genes associated with general cognitive function, white matter hyperintensity (WMH) and AD. Regions were fine-mapped using Fine-mapping Of CaUsal gene Sets (FOCUS). We identified 266, 23, and 69 genes associated with general cognitive function, WMH, and AD, respectively (Bonferroni-corrected alpha level $=P<2.9\text{x}10^{-6}$), some of which were previously identified. Enrichment analysis showed that many of the identified genes were in pathways related to innate immunity, vascular dysfunction, and neuroinflammation. Further, downregulation of *ICA1L* was associated with higher WMH and with AD, indicating its potential contribution to overlapping AD and VaD neuropathology. To our knowledge, our study is the first TWAS of cognitive function and neurocognitive disorders

that used expression mapping studies in multiple ancestries. This work may expand the benefits of TWAS studies beyond a single ancestry group and help to identify gene targets for pharmaceutical or preventative treatment for dementia.

4.2 Introduction

Adult-onset dementia is comprised of a group of aging-related neurocognitive disorders caused by the gradual degeneration of neurons and the loss of brain function. These changes lead to a decline in cognitive abilities and impairment of daily activities and independent function. In the United States, Alzheimer's disease (AD), the most common cause of dementia, affects 6.8 million adults age 65 and older.[1] The second most common form of dementia is vascular dementia (VaD), which often co-occurs with AD and is underdiagnosed.[1,2] VaD is often difficult to distinguish from AD because these diseases share cognitive symptoms including noticeable impairment in episodic and semantic memory. While AD and VaD often co-occur, each form of dementia has differing pathophysiology that may precede the illness decades prior.

AD is characterized by aggregation of amyloid-beta protein and neurofibrillary tangles in brain tissue,[3,4] while VaD may be caused by reduced blood flow to the brain as a result of small vessel disease (SVD) or stroke and is commonly seen in people with hypertension.[5] AD is diagnosed based on a battery of memory tests, brain-imaging tests for degeneration of brain cells and laboratory tests to assess the presence of amyloid and tau proteins in cerebrospinal fluid.[6] SVD is primarily detected on magnetic resonance imaging (MRI) as white matter hyperintensities (WMH). It has been hypothesized that vascular and neurodegenerative changes in the brain may interact in ways that increase the likelihood of cognitive impairment. A further

challenge in the field is distinguishing between individuals who are aging normally from those with dementia pathology.

A greater understanding of the pathological processes that influence cognitive function in older adults is critical for early intervention during the long preclinical or prodromal phase prior to dementia onset, especially in vulnerable populations.[7,8] For example, individuals of African ancestry (AA) have a greater burden of and risk for developing dementia compared to Non-Hispanic Whites.[9–12] Differences in gene expression, which are influenced by both genetic and non-genetic factors, likely play a role in shaping racial/ethnic health disparities in neurological outcomes. However, the underlying molecular and environmental mechanisms that influence gene expression are not fully understood, especially in populations with non-European ancestries. Given the multifactorial and complex nature of dementia, multi-omic data integration across ancestry groups may lend insight into these disparities, allowing the identification of targets for intervention and treatment in populations that are most at risk.[13]

Genome-wide association studies (GWAS) have identified genetic variants associated with cognitive function and dementia; however, most GWAS variants are located in non-coding regions so their functional consequences are difficult to characterize.[14] Transcriptome-wide association studies (TWAS) utilize gene expression and genetic data to increase power for identifying gene-trait associations and characterizing transcriptomic mechanisms underlying complex diseases. To date, however, few TWAS have been conducted on cognitive or structural brain measures. Further, previous TWAS have primarily been conducted in populations of European ancestry (EA), but these results cannot always be generalized to other genetic ancestries due to differences in allele frequencies, patterns of linkage disequilibrium (LD), and relationships between SNPs and gene expression between populations.[15–18] To better identify

gene-trait associations in non-EA ancestries, it is necessary to incorporate results from recent expression quantitative trait locus (eQTL) mapping studies, which identify genetic variants that explain variations in gene expression levels, conducted in different ancestry groups.[19]

Multi-ancEstry TRanscriptOme-wide analysis (METRO)[20] is a TWAS method that uses a joint likelihood-based inference framework to borrow complementary information across multiple ancestries to increase TWAS power. In this study, we used genotype and gene expression data from 1,032 AA and 801 EA from the Genetic Epidemiology Network of Arteriopathy (GENOA) and summary statistics from published GWAS[21–23] to identify genes associated with general cognitive function, white matter hyperintensity, and AD. We then examined the contribution of different ancestry-dependent transcriptomic profiles on the gene-trait associations. Greater knowledge of the underlying molecular mechanisms of dementia that are generalizable to both EA and AA is a critical step in evaluating potential causal variants and genes that could be targeted for pharmaceutical development.

4.3 Materials and Methods

4.3.1 Sample

The Genetic Epidemiology Network of Arteriopathy (GENOA)

The GENOA study is a community-based longitudinal study aimed at examining the genetic effects of hypertension and related target organ damage.[24] EA and AA hypertensive sibships were recruited if at least 2 siblings were clinically diagnosed with hypertension before age 60. All other siblings were invited to participate, regardless of their hypertension status. Exclusion criteria included secondary hypertension, alcoholism or drug abuse, pregnancy, insulin-dependent diabetes mellitus, active malignancy, or serum creatinine levels >2.5mg/dL. In

Phase I (1996-2001), 1,854 AA participants (Jackson, MS) and 1,583 EA participants (Rochester, MN) were recruited.[24] In Phase II (2000-2004), 1,482 AA and 1,239 EA participants were successfully followed up, and their potential target organ damage from hypertension was measured. Demographics, medical history, clinical characteristics, information on medication use, and blood samples were collected in each phase. After data cleaning and quality control, a total of 1,032 AA and 801 EA with genotype and gene expression data were available for analysis. Written informed consent was obtained from all participants, and approval was granted by participating institutional review boards (University of Michigan, University of Mississippi Medical Center, and Mayo Clinic).

4.3.2 Measures

A. Genetic data

AA and EA blood samples were genotyped using the Affymetrix® Genome-Wide Human SNP Array 6.0 or the Illumina 1M Duo. We followed the procedures outlined by Shang et al.[18] for data processing. For each platform, samples and SNPs with a call rate <95%, samples with mismatched sex, and duplicate samples were excluded. After removing outliers identified from genetic principal component analysis, there were 1,599 AA and 1,464 EA with available genotype data. Imputation was performed using the Segmented HAPlotype Estimation & Imputation Tool (SHAPEIT) v.2.r[25] and IMPUTE v.2[26] using the 1000 Genomes project phase I integrated variant set release (v.3) in NCBI build 37 (hg19) coordinates (released in March 2012). Imputation for each genotyping platform was performed separately and then combined. The final set of genotype data included 30,022,375 and 26,079,446 genetic variants for AA and EA, respectively. After removing genetic variants with MAF ≤ 0.01, imputation quality score

(INFO score) ≤ 0.4 in any platform-based imputation, and indels, a total of 13,793,193 SNPs in AA and 7,727,215 SNPs in EA were available for analysis. We used the GENESIS package[27] in R to infer population structure in the analytic sample, and the PC-AiR function was used to extract the first five genotype PCs which were subsequently used to adjust for population structure.

B. Gene expression data

Gene expression levels were measured from Epstein-Barr virus (EBV) transformed B-lymphoblastoid cell lines (LCLs) created from blood samples from a subset of GENOA AA (n=1,233) and EA (n=919). Gene expression levels of AA samples were measured using the Affymetrix Human Transcriptome Array 2.0, while gene expression levels of EA samples were measured using Affymetrix Human Exon 1.0 ST Array. We followed the procedures outlined by Shang et al.[18] In particular, the Affymetrix Expression Console was used for quality control and all array images passed visual inspection. In AA, 28 samples were removed due to either low signal-to-noise ratio (n=1), abnormal polyadenylated RNA spike-in controls (Lys < Phe < Thr < Dap; n=24), sample mislabeling (n=2), or low RNA integrity (n=1), resulting in a total of n=1,205 AA samples for analysis. In EA, duplicated samples (n=31), control samples (n=11) and sex mismatch samples (n=2) were removed, resulting in n=875 EA samples for analysis. We processed data in each population separately. Raw intensity data were processed using the Affymetrix Power Tool software.[28] AffymetrixCEL files were normalized using the Robust Multichip Average (RMA) algorithm which included background correction, quantile normalization, $\log_2$-transformation, and probe set summarization.[29] The algorithm also includes GC correction (GCCN), signal space transformation (SST), and gain lock (value=0.75) to

maintain linearity. The Brainarray custom CDF[30] v.19 was used to map the probes to genes. This custom CDF uses updated genomic annotations and multiple filtering steps to ensure that the probes used are specific for the intended gene cluster. Specifically, it removes probes with non-unique matching cDNA/EST sequences that can be assigned to more than one gene cluster. As a result, gene expression data processed using custom CDF are expected to be largely free of mappability issues. After mapping, ComBat[31] was used to remove batch effects. For each gene, we applied a linear regression model to adjust for age, sex, and first five genotype principal components (PCs). We then extracted the residuals and quantile normalized residuals across all samples. We analyzed a common set of 17,238 protein coding genes that were annotated in GENCODE (release 12).[32]

C. GWAS summary statistics

We used GWAS summary statistics for general cognitive function,[21] WMH,[22] and AD[23] as input for METRO. These GWASs were selected because they are the largest meta-analyses to date with publicly available summary statistics; however, we note that all three were conducted in primarily EA samples. Below, we describe each GWAS and also provide information about the corresponding TWAS analyses that were reported in two of the input GWAS (WMH and AD)[112] which use the same GWAS summary statistics as our analysis but different gene expression data.

i. General cognitive function

We obtained GWAS summary statistics for general cognitive function from a meta-analysis by Davies et al. (2018) that includes the Cohorts for Heart and Aging Research in

Genomic Epidemiology (CHARGE), the Cognitive Genomics Consortium (COGENT) consortia and the UK Biobank (UKB; Table 4-1).[15] This study included 300,486 EA individuals with ages between 16 and 102 years from 57 population-based cohorts. This is the largest available GWAS for general cognitive function, and there are currently no large-scale GWAS studies available in non-EA. General cognitive function was constructed from a number of cognitive tasks. Each cohort was required to have tasks that tested at least three different cognitive domains. Principal component analysis was performed on the cognitive tests scores within each cohort, and the first unrotated component was extracted as the measure of general cognitive function. Models performed within each cohort were adjusted for age, sex, and population stratification. Exclusion criteria included clinical stroke (including self-reported stroke) or prevalent dementia.

ii. White matter hyperintensity

We obtained the GWAS summary statistics for WMH from a meta-analysis conducted by Sargurupremraj et al. (2020) that included 48,454 EA and 2,516 AA with mean age of 66.0 (SD=7.5) years from 23 population-based studies from the CHARGE consortium and UKB (Table 4-1).[305] We obtained publicly available GWAS summary statistics from only EA individuals. Summary statistics for only EA are publicly available for this GWAS. WMH was measured from MRI scans obtained from scanners with field strengths ranging from 1.5 to 3.0 Tesla and interpreted using a standardized protocol blinded to clinical or demographic features. In addition to T1 and T2 weighted scans, some cohorts included fluid-attenuated inversion recovery (FLAIR) and/or proton density (PD) sequences to measure WMH from cerebrospinal fluid. WMH volume measures were inverse normal transformed, and models adjusted for sex,

age, genetic PCs and intracranial volume (ICV). Exclusion criteria included history of stroke or other pathologies that influence measurement of WMH at the time of MRI.

To functionally characterize and prioritize individual WMH genomic risk loci, Sargurupremraj et al[305] (2020) conducted TWAS using TWAS-Fusion[310] with summary statistics from the WMH SNP-main effects (EA only) analysis and weights from gene expression reference panels from blood (Netherlands Twin Registry; Young Finns Study), arterial (Genotype-Tissue Expression, GTEx), brain (GTEx, CommonMind Consortium) and peripheral nerve tissue (GTEx). This study did not perform fine-mapping following TWAS analysis.

iii. Alzheimer's disease

We obtained the GWAS summary statistics for Alzheimer's disease from stage I meta-analysis by Bellenguez et al. (2022) that included EA from the European Alzheimer and Dementia Biobank (EADB), GR@ACE, EADI, GERAD/PERADES, DemGene, Bonn, the Rotterdam study, CCHS study, NxC and the UKB (Table 4-1).[112] The meta-analysis was performed on 39,106 clinically diagnosed AD cases, 46,828 proxy-AD and related dementia (ADD) cases, and 401,577 controls. AD cases were clinically diagnosed in all cohorts except UKB, where individuals were identified as proxy-ADD cases if their parents had dementia. Participants without the clinical diagnosis of AD, or those without any family history of dementia, were used as controls. Models performed within each cohort were adjusted for PCs and genotyping centers, when necessary.

To examine the downstream effects of new AD-associated variants on molecular phenotypes in various AD-relevant tissues, Bellenguez et al. (2022) conducted a TWAS with stage I AD GWAS results. The TWAS was performed by training functional expression and

splicing reference panels based on the Accelerating Medicines Partnership (AMP)-AD bulk brain and EADB lymphoblastoid cell lines (LCL) cohorts, while leveraging pre-calculated reference panel weights[311] for the GTEx dataset[204] in tissues and cells of interest. TWAS associations were then fine-mapped using Fine-mapping Of CaUsal gene Sets (FOCUS).[312]

4.3.3 Statistical Methods

A. Multi-ancestry transcriptome-wide association study

Using the Multi-ancEstry TRanscriptOme-wide analysis (METRO),[44] we conducted high-powered TWAS with calibrated type I error control to identify the key gene-trait associations and transcriptomic mechanisms underlying AD, WMH and general cognitive function. Since gene expression prediction models constructed in different ancestries may contain complementary information, even when the input GWAS was conducted in a single ancestry,[44] we used METRO to model gene expression from EA and AA simultaneously. METRO uses a joint-likelihood framework that accounts for SNP effect size heterogeneity and LD differences across ancestries. The framework selectively upweights information from the ancestry that has greater certainty in the gene expression prediction model, increasing power and allowing characterization of the relative contribution of each ancestry to the TWAS results.

METRO is described in Li et al.[44] Briefly, each gene is examined separately using gene expression data from M different genetic ancestries. $\boldsymbol{Z_m}$ is the n_m-vector of gene expression measurements on n_m individuals in the mth ancestry with $m \in \{1,\ldots,M\}$. For the gene of interest, all *cis*-SNPs (p), which are in potential linkage disequilibrium (LD) with each other, were extracted as predictors for gene expression. $\boldsymbol{G_m}$ is denoted as the $n_m \: x \: p$ genotype matrix for these *cis*-SNPs. Besides the gene expression data, we also used GWAS summary statistics from n

individuals for an outcome trait of interest. $\boldsymbol{\gamma}$ is the n-vector of outcome measurements in the GWAS data and $\boldsymbol{G}$ is the corresponding $n \times p$ genotype matrix on the same set of p *cis*-SNPs. The expression vector $\boldsymbol{z_m}$, the outcome vector $\boldsymbol{\gamma}$, and each column of the genotype matrixes are centered and standardized. $\boldsymbol{G_m}$ and $\boldsymbol{G}$ have a mean of zero and variance of one. For each TWAS, we used GWAS summary statistics in the form of marginal z-scores and a SNP-SNP correlation (LD) matrix estimated with genotype data from our GENOA EA sample. The following equations describe the relationships between the SNPs, gene expression and the outcome:

$$\boldsymbol{z}_m = \boldsymbol{G}_m \boldsymbol{\beta}_m + \epsilon_m, \qquad \text{(Equation 1)}$$

$$\boldsymbol{y} = \alpha(\boldsymbol{G}\boldsymbol{\beta}) + \epsilon_y, \qquad \text{(Equation 2)}$$

Equation (1) describes the relationship between gene expression and the *cis*-SNP genotypes in the gene expression study in GENOA for the mth ancestry (EA or AA). β_m is a p vector of the *cis*-SNP effects on the gene expression in the mth ancestry and ε_m is an n_m-vector of residual errors with each element following an independent and normal distribution N(0, $\sigma^2{}_m$) with an ancestry specific variance $\sigma^2{}_m$. Equation (2) describes the relationship between the genetically regulated gene expression (GreX), calculated from estimated SNP prediction weights, and the outcome trait (general cognitive function, WMH or AD) from the GWAS. There, $G\beta$ denotes an n-vector of GreX constructed for the GWAS individuals, where $\beta = \Sigma_m \, w_m \beta_m$ is a p-vector of SNP effects on the gene expression in the GWAS data, where the weights $\Sigma^M{}_{m=1} \, w_m=1$ and $w_m \geq 0$. The alpha value (α) is the effect of GreX constructed for the GWAS individuals on the outcome trait, and ε_m is an n_m-vector of residual errors with each element following an independent and normal distribution N(0, $\sigma^2{}_m$). Both equations, specified based on separate studies, are connected through the predictive SNP effects on the gene expression (β_m and β). A

key assumption made is that the SNP effects on the gene expression in the GWAS, β, can be expressed as a weighted summation of the SNP effects on gene expression in the expression studies conducted across ancestries.

We derived the overall GreX effect α and the contribution weight of each ancestry (w_1 for AA and w_2 for EA) to infer the extent and contribution of the two genetic ancestries in informing the GreX-trait association. The joint model defined in Equations 1 and 2 allows us to borrow association strength across multiple ancestries to enable powerful inference of GreX-trait associations for general cognitive function, WMH and AD. We declared the gene to be significant if the p-value was below the corresponding Bonferroni corrected threshold for the number of tested genes ($P<0.05/17{,}238 = 2.90\times10^{-6}$). Manhattan plots and quantile-quantile (QQ) plots were generated using the *qqman*[313] R package.

B. Fine-mapping analysis

Since genes residing in the same genomic region may share eQTLs or contain eQTL SNPs in LD with each other, TWAS test statistics for genes in the same region can be highly correlated, making it difficult to identify the true biologically relevant genes among them. To prioritize the putatively causal genes identified by METRO for general cognitive function, WMH, and AD, we conducted TWAS fine-mapping using FOCUS (Fine-mapping Of CaUsal gene Sets).[312] To identify a genomic region with at least one significant gene detected by METRO, we obtained a set of independent, non-overlapping genomic regions, or LD blocks, using Ldetect.[314] In each analyzed genomic block, using a standard Bayesian approach, we assigned a posterior inclusion probability (PIP) for each gene to be causal, given the observed TWAS statistics. We used gene-level Z scores, created from p-values using the inverse

cumulative distribution function (CDF) of a standard normal distribution, as input into FOCUS. We then ranked the PIPs and computed the 90%-credible set that contains the causal gene with 90% probability. In the FOCUS analysis, a null model which assumes none of the genes in the region are causally associated with the trait is also considered as a possible outcome and may be included in the credible set. Through fine-mapping, we narrowed down significantly associated genes identified by METRO to a shorter list of putatively true associations.

C. Characterization of identified genes

To interpret our TWAS findings, both before and after fine-mapping, we further examined whether the genes identified by METRO overlapped with those previously identified by their corresponding input GWAS. We created a set of Venn diagrams of overlapping genes identified using METRO with those from the SNP-based GWAS association results[15,112,305] mapped to the nearest gene using the *VennDiagram* R package.[315] We then constructed a second set of Venn diagrams showing overlapping genes identified using METRO with genes identified by gene-based association analyses in each of the input GWAS studies. The gene-based analyses were conducted using MAGMA[316] (general cognitive function[15] and WMH[305]) or gene prioritization tests (AD[112]). Finally, we created a set of Venn diagrams comparing genes identified using METRO with those identified in the TWAS that were conducted as part of the WMH[305] and AD[112] input GWAS studies. We used the *geneSynonym* R package[317] to ensure that genes named differently across studies were captured.

D. Functional enrichment analysis

To characterize the biological function of the identified genes by METRO for general cognitive function, WMH and AD, we performed gene set enrichment analysis. Specifically, we used the g:GOSt[318] tool on the web software g:Profiler and mapped the genes to known functional informational sources, including Gene Ontology (GO): molecular function (MF), GO: biological process (BP), GO: cellular component (CC), Kyoto Encyclopedia of Genes and Genomes (KEGG), Reactome (REAC), WikiPathways (WP), Transfac (TF), MiRTarBase (MIRNA), Human Protein Atlas (HPA), CORUM protein complexes, and Human Phenotype Ontology (HP). In this analysis, we used the default option g:SCS method (Set Counts and Sizes) in g:Profiler for multiple testing correction and presented pathways identified with an adjusted p-value < 0.05. Driver terms in GO are highlighted using a two-stage algorithm for filtering GO enrichment results, providing a more efficient and reliable approach compared to traditional clustering methods. This feature groups significant terms into sub-ontologies based on their relations, and the second stage identifies leading gene sets that give rise to other significant functions in the same group of terms. This method uses a greedy search strategy that recalculates hypergeometric p-values and results in the consideration of multiple leading terms in a component, rather than selection of terms with the highest significance level.

4.4 Results

In Table 4-1, we provide descriptive statistics for the samples used in the eQTL mapping study (e.g., 1,032 AA and 801 EA from GENOA) and the three input GWAS.[15,112,305] The GENOA eQTL study included participants with a mean age of 56.9 (SD=10.0) years. More than half of participants were female (65.6%). Mean age of participants was 56.9 (SD=7.8) years in

the general cognitive function GWAS[15], and 64.2 years in the WMH GWAS.[305] In the AD GWAS,[112] mean age was 73.6 (SD=8.1) years for cases and 67.9 (SD=8.6) years for controls.

Using METRO, we identified 602 genes associated with general cognitive function, 45 genes associated with WMH, and 231 genes associated with AD that were significant at the Bonferroni corrected alpha level ($P<2.90x10^{-6}$; Figure 4-1, Tables S4-4 – S4-6). Genomic inflation factors for the TWAS p-values ranged from 1.45 to 2.55 (Figure 4-2). Among the three neurocognitive outcomes, prior to fine-mapping, METRO TWAS identified the *ICA1L* gene overlapping between WMH and AD; the *FMNL1* gene overlapping between WMH and general cognitive function; and 22 genes enriched in AD-related pathways and functions overlapping between general cognitive function and AD (Figure 4-3a; Figure S4-12). After fine-mapping, the only overlapping gene that remained was *ICA1L* between WMH and AD (Figure 4-3b).

For all identified genes, we also examined the contribution weights of expression prediction models for the EA and AA ancestries, prior to fine-mapping ($P<2.90x10^{-6}$; Figure 4-4). For the WMH TWAS, we found that the EA weights, on average, had a substantially higher contribution than AA weights for the identified genes (65.7%), and the proportion of genes with higher EA than AA weights was also large (65.2%). This is consistent with Li et al. (2022)[44] who found that the gene expression prediction models constructed in the same ancestry as the input GWAS, in this case EA, often have larger contribution weights than those constructed in other ancestries. However, for both general cognitive function and AD, the contributions from EA and AA weights were similar, which likely increased power to identify genes relevant to AA.

After fine-mapping, there were 266 genes in the 90%-credible set across 172 different genomic regions for general cognitive function. This gene set included 82 genes that were not previously identified in the SNP-based GWAS results (mapped to the nearest gene) or the gene-

based analysis results from Davies et al. (2018)[15] (Figure 4-5, Table S4-4); however, it is likely that some of these genes are in broader genomic regions tagged by the GWAS-identified SNPs. Specifically, there were 126 and 168 overlapping genes between METRO and the SNP-based and gene-based associations from Davies et al. (2018),[15] respectively (Figure 4-5). The 266 METRO-identified genes were enriched in regulatory pathways involved in protein binding ($p_{adj} = 1.17 \times 10^{-5}$), developmental cell growth ($p_{adj} = 3.33 \times 10^{-5}$), and protein metabolic process ($p_{adj} = 7.18 \times 10^{-4}$), as well as neurodevelopmental processes such as neuron to neuron synapse ($p_{adj} = 1.22 \times 10^{-3}$) and neuron projection ($p_{adj} = 7.14 \times 10^{-3}$; Figure 4-6). The 82 genes that were not previously identified in Davies et al. (2018)[15] were enriched for positive regulation of biological process ($p_{adj} = 1.77 \times 10^{-2}$), proteasome activator complex ($p_{adj} = 1.00 \times 10^{-2}$), nucleoplasm ($p_{adj} = 1.29 \times 10^{-2}$) and chromatin ($p_{adj} = 4.71 \times 10^{-5}$; Figure S4-13).

After fine-mapping, there were 23 genes in the 90%-credible set across 15 genomic regions for WMH, including 12 genes that were not previously identified in the SNP-based GWAS results mapped to the nearest gene or the gene-based analysis results from Sargurupremraj et al. (2020)[305] (Figure 4-7, Table S4-5). Specifically, there were 7 and 12 overlapping genes between METRO and the SNP-based and gene-based associations from Sargurupremraj et al. (2020),[305] respectively (Figure 4-7). The 23 METRO-identified genes were enriched for zinc finger motif ($p_{adj} = 1.27 \times 10^{-2}$), miRNA has-212-5p ($p_{adj} = 1.94 \times 10^{-2}$) and retinal inner plexiform layer ($p_{adj} = 3.86 \times 10^{-2}$; Figure 4-8). The 12 genes associated with WMH that were previously not identified by Sargurupremraj et al. (2020)[305] were enriched for DNA binding domain Zinc Finger Protein 690 (ZNF690; $p_{adj} = 2.52 \times 10^{-3}$) and ClpX protein degradation complex ($p_{adj} = 4.97 \times 10^{-2}$; Figure S4-14).

After fine-mapping, there were 69 genes in the 90%-credible set across 56 genomic regions associated with AD, including 45 genes that were not previously identified in the SNP-based GWAS results mapped to the nearest gene or the gene prioritization analysis results from Bellenguez et al. (2022)[112] (Figure 4-9, Table S4-6). Specifically, there were 16 and 14 overlapping genes between METRO and the SNP-based and gene prioritization test results from Bellenguez et al. (2022),[112] respectively (Figure 4-9). The 69 METRO-identified genes were enriched for AD-associated processes including regulation of amyloid fibril formation (p_{adj} = 1.87 x 10^{-3}), amyloid-beta clearance (p_{adj} = 1.90 x 10^{-3}), microglial cell activation (p_{adj} = 5.79 x 10^{-3}), amyloid-beta metabolic process (p_{adj} = 1.07 x 10^{-2}), and neurofibrillary tangle (p_{adj} = 2.80 x 10^{-4}; Figure 4-10). The 45 genes associated with AD that were previously not identified by Bellenguez et al. (2022)[112] were enriched for hematopoietic cell lineage (p_{adj} = 1.73 x 10^{-3}) and neurofibrillary tangle (p_{adj} = 9.13 x 10^{-3}; Figure S4-15).

We compared the genes identified by METRO before and after fine-mapping with those identified by TWAS studies in Sargurupremraj et al. (2020)[305] and Bellenguez et al. (2022)[112] which used TWAS-Fusion (Figure 4-11). For WMH, there were 16 and 10 genes identified both by METRO before and after fine-mapping and by the TWAS-Fusion analysis conducted by Sargurupremraj et al. (2020)[305], respectively (Table 4-2). For AD, there were 24 and 10 genes identified both by METRO before and after fine-mapping and by the TWAS-Fusion followed by FOCUS fine-mapping analysis conducted by Bellenguez et al. (2022)[112] (Table 4-3). *ICA1L* was the only gene overlapping between all four AD and WMH TWAS association results.

4.5 Discussion

While previous studies have identified genes associated with cognitive function, WMH, and AD, there are few TWAS that utilize genetic and gene expression data from multiple ancestries to elucidate gene-trait associations and molecular mechanisms underlying the etiologies of cognitive function and neurocognitive disorders. Using the METRO method followed by FOCUS fine-mapping, we identified 266, 23, and 69 genes associated with general cognitive function, WMH, and AD, respectively, with 82, 12 and of them not previously identified in the original GWAS. Studying the transcriptomic mechanisms underlying cognitive function, WMH and dementia using both EA and AA expression data may enhance our understanding of cognitive health prior to and following the onset of dementia and further allow us to generalize findings from large scale EA GWAS to other ancestries.

AD and SVD have overlapping features that contribute to dementia neuropathology including breakdown of the blood-brain barrier[43] and the presence of small cortical and subcortical infarcts, microbleeds, perivascular spacing, and WMH in brain tissue.[44] After fine-mapping, Islet Cell Autoantigen 1 Like (*ICA1L*) was identified in both the WMH and AD TWAS. This is as a highly plausible prioritized gene that is likely to modulate the metabolism of amyloid precursor protein (APP)[23] and increase risk of AD. *ICA1L* encodes a protein whose expression is activated by type IV collagen and plays a crucial role in myelination.[45] Increased *ICA1L* expression is also associated with lower risk of AD[46–48] and small vessel strokes (SVS), the acute outcomes of cerebral SVD, which may lead to VaD.[49] Consistent with these studies, our TWAS found that decreased expression of *ICA1L* is associated with increased risk of AD and WMH, a subclinical indicator of SVD. Single-cell RNA-sequencing has shown *ICA1L* expression to be enriched in cortical glutamatergic excitatory neurons, which are crucial

components in neural development and neuropathology through their role in cell proliferation, differentiation, survival, neural network formation and cell death.[50,51] *ICA1L* has been examined as a possible drug target for SVD, AD, and other neurodegenerative diseases;[49,52] however, it is not recommended as a prioritized drug at this time due to potential side effects including increased risk of coronary artery disease and myocardial infarction as well as lower diastolic blood pressure.[52] Nevertheless, *ICA1L* may contribute to overlapping AD and VaD neuropathology, and it could be a potential target for therapeutics and/or preventative treatments for AD and VaD in the future if adverse events can be reduced.

Our TWAS of AD identified 45 genes that were not identified in the SNP-based GWAS results mapped to the nearest gene or the gene-based analysis reported in Bellenguez et al. (2022).[23] The 45 genes were enriched for hematopoietic cell lineage, which are progenitors of red and white blood cells including those related to immunity (e.g., natural killer cells, T- and B-lymphocytes and other types of leukocytes).[53–59] Our TWAS identified genes that have been previously associated with AD, including *APOE, TOMM40, APOC4, CLU, PICALM* and *CR2,* among others.[23,60,61] While our TWAS identified *APOE,* the largest genetic risk factor for AD in AA and EA, we did not identify *ABCA7* which confers the greatest genetic risk for AD in AA.[62–64] This finding is perhaps not surprising considering that our TWAS was conducted using an EA GWAS. The strength of association between *ABCA7* and AD has been shown to be comparatively weaker in EA than in AA.[64] To identify genes associated with AD risk in AA populations, it would be beneficial to perform a TWAS utilizing a well-powered AD GWAS in AA. This approach may reveal the involvement of *ABCA7* and other genes contributing to AD risk in AA populations.

In our AD TWAS, we also identified genes associated with other neurological and autoimmune diseases including Parkinson's disease (*CYB561*[65] and *SLC25A39*[66]), Crohn's disease (*ATG16L1*[67]), Amyotrophic lateral sclerosis (*SIGLEC9*[68]), and Riboflavin Transport Deficiency (*SLC52A1*[69]). These diseases have in common the progressive peripheral and cranial degeneration of neurons that impact processes such as voluntary muscle movement, vision, hearing and sensation. Although not explicitly identified in Bellenguez et al. (2022),[23] we also identified genes that were associated with AD in other studies including *RIN3* that is implicated in tau-mediated pathology, the *MS4A* (*4A* and *6A*) locus associated with mast cell activation, *TP53INP1* and *ZYX* that have been linked to myeloid enhancer activity,[70] and *APOC4*, which is located proximal to *APOE.*[71] We also identified additional genes involved in B cell autoimmunity (*HLA-DQA2,*[72,73] *CSTF1*[74]), neurodegenerative processes (*SUPT4H1,*[75] *C6orf10,*[76] *IKZF1,*[77] *DEDD*[78]), and neuronal growth (*IKZF1,*[77] *STYX*[79]). Our findings support the hypothesis that chronic activation of immune cells resident in the brain and peripheral nervous system appear to play a critical role in neuroinflammatory responses that drive the progression of neurodegeneration in AD.[80] Further, consistent with findings that AD and VaD often co-exist, our AD TWAS identified genes that were associated with lacunar and ischemic strokes as well as cerebral small vessel disease in other studies, including *SLC39A13,*[81] *RAPSN,*[81] *MAF1*,[82] and *MME.*[83,84]

Although our WMH TWAS identified 12 genes that were not included in the SNP-based GWAS results mapped to the nearest gene or the gene-based analysis reported in Sargurupremraj et al. (2020),[22] other studies found associations between *MAP1LC3B,*[85] *ARMS2*[86,87] and *HTRA1*[81] with ischemic stroke, lacunar stroke, and cerebral SVD. The WMH TWAS also identified genes associated with AD (*ARMS2*),[88] atrial fibrillation (*NEURL*[89] and *GJC1*[90]), innate immunity

(*EFTUD2*[91]) and apoptosis and neurodevelopment (*PDCD7,*[92] *FBXO31,*[93] and *ClpX*[94]). The 12 unique genes identified for WMH were enriched for DNA binding domain Zinc Finger Protein 690,[95] which plays an essential role in gene regulation, transcription and various cellular processes, and ClpX protein degradation complex,[96] which maintains protein homeostasis. Our findings were consistent with studies that showed neuroinflammation to be an immunological cascade reaction by glial cells of the central nervous system where innate immunity resides.

While our TWAS for general cognitive function did not show overlapping genes between the TWAS for AD and VaD, we identified genes associated with general cognitive function that were not explicitly identified by Davies et al. (2018)[21] which were associated with pre-clinical AD and VaD risk factors including cardiovascular diseases, immunity and Alzheimer's neuropathology. Our TWAS also identified genes previously associated with cognitive domains, neuropathology, and psychiatric illness including reading-related skills and neural structures (*SEMA6D*[97] and *SETBP1*[98]), working memory tasks (*CDH13*[99]) and Schizophrenia (*HP,*[100,101] *C18orf1*[102] and *TMEM180*[103]). There are likely also distinct transcriptomic mechanisms that differentiate cognitive function and normal age-related brain changes from pathways related to dementia. Individuals who never develop dementia or significant cognitive decline still experience brain deterioration in normal aging that includes gray and white matter loss and ventricular enlargement which is accompanied by memory decline.[104] Further, previous GWAS for general cognitive function and AD have shown few overlapping loci.[21,105] In addition, studies of older individuals who are cognitively "resilient" with intact cognitive function, despite the presence of AD neuropathology, have found the genetic architecture of cognitive resilience to be distinct from that of AD.[106] As such, relatively little is known about the pathways underlying cognitive aging in those without dementia. Thus, studying transcriptomic mechanisms that affect

general cognitive function before development of dementia may shed light on cognitive aging without dementia.

We also compared genes identified by METRO after fine-mapping with those identified by TWAS-Fusion in Sargurupremraj et al. (2020)[22] and Bellenguez et al. (2022).[23] Among the 92 genes associated with WMH in Sargurupremraj et al. (2020)[22] and 23 genes identified by METRO, 10 genes overlapped. We note that the Sargurupremraj et al. (2020)[22] did not perform fine-mapping of their TWAS results, which is likely why we identified substantially fewer genes. There were also 10 overlapping genes among the 66 genes associated with AD in Bellenguez et al. (2022)[23] and 69 genes identified by METRO. For both TWAS comparisons, a relatively small number of genes overlap likely due to differences in eQTL prediction modeling. Sargurupremraj et al. (2020)[22] and Bellenguez et al. (2022)[23] used eQTL data from brain tissue, while we used eQTL data from transformed beta lymphocytes in blood tissue. While brain tissue is more relevant to WMH and AD phenotypes, blood cells do touch every cell bed that affects the brain, and are related to chronic inflammation, immunity, and oxidative stress, which are linked to cognitive performance and dementia. TWAS results from blood tissue in multiple ancestries provide complementary information to those reported in the GWAS.

Several limitations in the present study should be noted. First, our gene expression levels were measured using transformed B-lymphocytes from immortalized cell lines in GENOA. While transformed B-lymphocytes are a convenient source of DNA from blood tissue, we lack eQTL data for tissues that may be most relevant for AD and WMH (e.g., brain tissue, small brain vessels, and microglia). However, B-lymphocytes provide a unique and efficient way to examine the functional effects of genetic variations on gene expression that minimizes environmental influences.[107] Second, METRO follows the standard TWAS approach of analyzing one gene at a

time. Since genes residing in the same genomic region may share eQTLs or contain eQTL SNPs that are in LD with each other, the TWAS test statistics of genes in the same region may be highly correlated. To that end, it may be challenging to identify the truly biologically relevant genes among them.[36,108] As such, we paired METRO with FOCUS to allow us to narrow down the list of potential causal genes for AD, VaD, and cognitive decline.[36,109] Lastly, we utilized EA GWAS that were publicly available for general cognitive function, WMH, and AD. As expected, the gene expression prediction models constructed in the same ancestry as the GWAS (EA) tended to have larger contribution weights than AA. As such, a future direction would be to conduct TWAS of these traits using summary statistics from GWAS with AA ancestry or multiple ancestries as they become available.

Our study also has notable strengths. To our knowledge, our study is the first TWAS using expression mapping studies in multiple ancestries (EA and AA) to identify genes associated with cognitive function and neurocognitive disorders. By leveraging the complementary information in gene expression prediction models constructed in EA and AA, as well as the uncertainty in SNP prediction weights, we were able to conduct a highly powered TWAS to identify important gene-trait associations and transcriptomic mechanisms related to innate immunity, vascular dysfunction and neuroinflammation underlying AD, VaD, and general cognitive function. Using METRO, we were also able to estimate the ancestry contribution weights for specific genes and identify the extent to which a gene in EA or AA may contribute to the trait. However, it is noteworthy that the larger the contribution of the expression prediction models in the same ancestry as the GWAS (EA, in this study) may allow for better predictive performance in the same ancestry. We also conducted FOCUS fine-mapping to narrow in on a list of putatively causal genes among multiple significant genes in a region. Our results suggest

that there are similar pathways that contribute to healthy cognitive aging and progression of dementia, as well as distinct pathways that are unique to each neuropathology. By understanding overlapping and unique genes and transcriptomic mechanisms associated with each outcome, we may identify possible targets for prevention and/or treatments for cognitive aging and dementia.

4.6 Conclusion

In the present study, we conducted a multi-ancestry TWAS in EA and AA to identify genes associated with general cognitive function, WMH and AD. We identified genes associated with innate immunity, vascular dysfunction, and neuroinflammation. The WMH and AD TWAS also indicated that downregulation of *ICA1L* may contribute to overlapping AD and VaD neuropathology. To our knowledge, this study is the first TWAS analysis using expression mapping studies in multiple ancestries to identify genes associated with cognitive function and neurocognitive disorders, which may help to identify gene targets for pharmaceutical or preventative treatment for dementia.

4.7 References

1. 2023 Alzheimer's disease facts and figures (2023). Alzheimer's & Dementia *19*, 1598–1695. 10.1002/alz.13016.

2. De Reuck, J., Maurage, C.-A., Deramecourt, V., Pasquier, F., Cordonnier, C., Leys, D., and Bordet, R. (2018). Aging and cerebrovascular lesions in pure and in mixed neurodegenerative and vascular dementia brains: a neuropathological study. Folia Neuropathol *56*, 81–87. 10.5114/fn.2018.76610.

3. Hof, P.R., Glannakopoulos, P., and Bouras, C. (1996). The neuropathological changes associated with normal brain aging. Histol Histopathol *11*, 1075–1088.

4. Perl, D.P. (2010). Neuropathology of Alzheimer's Disease. Mt Sinai J Med *77*, 32–42. 10.1002/msj.20157.

5. Cohuet, G., and Struijker-Boudier, H. (2006). Mechanisms of target organ damage caused by hypertension: Therapeutic potential. Pharmacology & Therapeutics *111*, 81–98. 10.1016/j.pharmthera.2005.09.002.

6. Weller, J., and Budson, A. (2018). Current understanding of Alzheimer's disease diagnosis and treatment. F1000Res *7*, F1000 Faculty Rev-1161. 10.12688/f1000research.14506.1.

7. Hubbard, B.M., Fenton, G.W., and Anderson, J.M. (1990). A quantitative histological study of early clinical and preclinical Alzheimer's disease. Neuropathol Appl Neurobiol *16*, 111–121. 10.1111/j.1365-2990.1990.tb00940.x.

8. Sperling, R.A., Aisen, P.S., Beckett, L.A., Bennett, D.A., Craft, S., Fagan, A.M., Iwatsubo, T., Jack, C.R., Kaye, J., Montine, T.J., et al. (2011). Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement *7*, 280–292. 10.1016/j.jalz.2011.03.003.

9. Mayeda, E.R., Glymour, M.M., Quesenberry, C.P., and Whitmer, R.A. (2016). Inequalities in dementia incidence between six racial and ethnic groups over 14 years. Alzheimers Dement *12*, 216–224. 10.1016/j.jalz.2015.12.007.

10. Barnes, L.L., and Bennett, D.A. (2014). Alzheimer's disease in African Americans: risk factors and challenges for the future. Health Aff (Millwood) *33*, 580–586. 10.1377/hlthaff.2013.1353.

11. Brewster, P., Barnes, L., Haan, M., Johnson, J.K., Manly, J.J., Nápoles, A.M., Whitmer, R.A., Carvajal-Carmona, L., Early, D., Farias, S., et al. (2019). Progress and future challenges in aging and diversity research in the United States. Alzheimers Dement *15*, 995–1003. 10.1016/j.jalz.2018.07.221.

12. Rajan, K.B., Weuve, J., Barnes, L.L., Wilson, R.S., and Evans, D.A. (2019). Prevalence and incidence of clinically diagnosed Alzheimer's disease dementia from 1994 to 2012 in a population study. Alzheimers Dement *15*, 1–7. 10.1016/j.jalz.2018.07.216.

13. Cacabelos, R., and Torrellas, C. (2015). Epigenetics of Aging and Alzheimer's Disease: Implications for Pharmacogenomics and Drug Response. International Journal of Molecular Sciences *16*, 30483–30543. 10.3390/ijms161226236.

14. Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., and Meyre, D. (2019). Benefits and limitations of genome-wide association studies. Nat Rev Genet *20*, 467–484. 10.1038/s41576-019-0127-1.

15. Lappalainen, T., Sammeth, M., Friedländer, M.R., 't Hoen, P.A., Monlong, J., Rivas, M.A., Gonzàlez-Porta, M., Kurbatova, N., Griebel, T., Ferreira, P.G., et al. (2013). Transcriptome and genome sequencing uncovers functional variation in humans. Nature *501*, 506–511. 10.1038/nature12531.

16. Mogil, L.S., Andaleon, A., Badalamenti, A., Dickinson, S.P., Guo, X., Rotter, J.I., Johnson, W.C., Im, H.K., Liu, Y., and Wheeler, H.E. (2018). Genetic architecture of gene expression traits across diverse populations. PLOS Genetics *14*, e1007586. 10.1371/journal.pgen.1007586.

17. Wojcik, G., Graff, M., Nishimura, K., Tao, R., Haessler, J., Gignoux, C., Highland, H., Patel, Y., Sorokin, E., Avery, C., et al. (2019). Genetic analyses of diverse populations improves discovery for complex traits. Nature *570*, 514–518. 10.1038/s41586-019-1310-4.

18. Shang, L., Smith, J.A., Zhao, W., Kho, M., Turner, S.T., Mosley, T.H., Kardia, S.L.R., and Zhou, X. (2020). Genetic Architecture of Gene Expression in European and African Americans: An eQTL Mapping Study in GENOA. Am J Hum Genet *106*, 496–512. 10.1016/j.ajhg.2020.03.002.

19. Nica, A.C., and Dermitzakis, E.T. (2013). Expression quantitative trait loci: present and future. Philos Trans R Soc Lond B Biol Sci *368*, 20120362. 10.1098/rstb.2012.0362.

20. Li, Z., Zhao, W., Shang, L., Mosley, T.H., Kardia, S.L.R., Smith, J.A., and Zhou, X. (2022). METRO: Multi-ancestry transcriptome-wide association studies for powerful gene-trait association detection. The American Journal of Human Genetics *109*, 783–801. 10.1016/j.ajhg.2022.03.003.

21. Davies, G., Lam, M., Harris, S.E., Trampush, J.W., Luciano, M., Hill, W.D., Hagenaars, S.P., Ritchie, S.J., Marioni, R.E., Fawns-Ritchie, C., et al. (2018). Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. Nat Commun *9*, 2098. 10.1038/s41467-018-04362-x.

22. Sargurupremraj, M., Suzuki, H., Jian, X., Sarnowski, C., Evans, T.E., Bis, J.C., Eiriksdottir, G., Sakaue, S., Terzikhan, N., Habes, M., et al. (2020). Cerebral small vessel disease genomics and its implications across the lifespan. Nature Communications *11*, 6285. 10.1038/s41467-020-19111-2.

23. Bellenguez, C., Küçükali, F., Jansen, I.E., Kleineidam, L., Moreno-Grau, S., Amin, N., Naj, A.C., Campos-Martin, R., Grenier-Boley, B., Andrade, V., et al. (2022). New insights into the genetic etiology of Alzheimer's disease and related dementias. Nat Genet, 1–25. 10.1038/s41588-022-01024-z.

24. Daniels, P.R., Kardia, S.L.R., Hanis, C.L., Brown, C.A., Hutchinson, R., Boerwinkle, E., Turner, S.T., and Genetic Epidemiology Network of Arteriopathy study (2004). Familial aggregation of hypertension treatment and control in the Genetic Epidemiology Network of Arteriopathy (GENOA) study. Am. J. Med. *116*, 676–681. 10.1016/j.amjmed.2003.12.032.

25. Delaneau, O., Zagury, J.-F., and Marchini, J. (2013). Improved whole-chromosome phasing for disease and population genetic studies. Nat Methods *10*, 5–6. 10.1038/nmeth.2307.

26. Howie, B.N., Donnelly, P., and Marchini, J. (2009). A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. PLOS Genetics *5*, e1000529. 10.1371/journal.pgen.1000529.

27. Gogarten, S.M., Sofer, T., Chen, H., Yu, C., Brody, J.A., Thornton, T.A., Rice, K.M., and Conomos, M.P. (2019). Genetic association testing using the GENESIS R/Bioconductor package. Bioinformatics *35*, 5346–5348.

28. Lockstone, H.E. (2011). Exon array data analysis using Affymetrix power tools and R statistical software. Briefings in Bioinformatics *12*, 634–644. 10.1093/bib/bbq086.

29. Irizarry, R.A., Bolstad, B.M., Collin, F., Cope, L.M., Hobbs, B., and Speed, T.P. (2003). Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Research *31*, e15. 10.1093/nar/gng015.

30. Dai, M., Wang, P., Boyd, A.D., Kostov, G., Athey, B., Jones, E.G., Bunney, W.E., Myers, R.M., Speed, T.P., Akil, H., et al. (2005). Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. Nucleic Acids Research *33*, e175. 10.1093/nar/gni179.

31. Johnson, W.E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics *8*, 118–127. 10.1093/biostatistics/kxj037.

32. Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., and Searle, S. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. Genome research *22*, 1760–1774.

33. Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B.W.J.H., Jansen, R., de Geus, E.J.C., Boomsma, D.I., Wright, F.A., et al. (2016). Integrative approaches for large-scale transcriptome-wide association studies. Nat Genet *48*, 245–252. 10.1038/ng.3506.

34. Barbeira, A.N., Bonazzola, R., Gamazon, E.R., Liang, Y., Park, Y., Kim-Hellmuth, S., Wang, G., Jiang, Z., Zhou, D., Hormozdiari, F., et al. (2021). Exploiting the GTEx

resources to decipher the mechanisms at GWAS loci. Genome Biology *22*, 49. 10.1186/s13059-020-02252-4.

35. THE GTEX CONSORTIUM (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science *369*, 1318–1330. 10.1126/science.aaz1776.

36. Mancuso, N., Freund, M.K., Johnson, R., Shi, H., Kichaev, G., Gusev, A., and Pasaniuc, B. (2019). Probabilistic fine-mapping of transcriptome-wide association studies. Nature genetics *51*, 675–682.

37. Turner, S.D. (2014). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. Preprint at bioRxiv, 10.1101/005165 10.1101/005165.

38. Berisa, T., and Pickrell, J.K. (2016). Approximately independent linkage disequilibrium blocks in human populations. Bioinformatics *32*, 283.

39. Chen, H., and Boutros, P.C. (2011). VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics *12*, 35. 10.1186/1471-2105-12-35.

40. de Leeuw, C.A., Stringer, S., Dekkers, I.A., Heskes, T., and Posthuma, D. (2018). Conditional and interaction gene-set analysis reveals novel functional pathways for blood pressure. Nat Commun *9*, 3768. 10.1038/s41467-018-06022-6.

41. Mancarcı, B.O. (2022). Gene Synonym.

42. Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., and Vilo, J. (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Research *47*, W191–W198. 10.1093/nar/gkz369.

43. Chagnot, A., Barnes, S.R., and Montagne, A. (2021). Magnetic Resonance Imaging of Blood–Brain Barrier permeability in Dementia. Neuroscience *474*, 14–29. 10.1016/j.neuroscience.2021.08.003.

44. Kalaria, R.N., and Sepulveda-Falla, D. (2021). Cerebral Small Vessel Disease in Sporadic and Familial Alzheimer Disease. The American Journal of Pathology *191*, 1888–1905. 10.1016/j.ajpath.2021.07.004.

45. Zhang, C., Qin, F., Li, X., Du, X., and Li, T. (2022). Identification of novel proteins for lacunar stroke by integrating genome-wide association data and human brain proteomes. BMC Medicine *20*, 211. 10.1186/s12916-022-02408-y.

46. Bai, B., Vanderwall, D., Li, Y., Wang, X., Poudel, S., Wang, H., Dey, K.K., Chen, P.-C., Yang, K., and Peng, J. (2021). Proteomic landscape of Alzheimer's Disease: novel insights into pathogenesis and biomarker discovery. Mol Neurodegener *16*, 55. 10.1186/s13024-021-00474-z.

47. Wingo, A.P., Liu, Y., Gerasimov, E.S., Gockley, J., Logsdon, B.A., Duong, D.M., Dammer, E.B., Robins, C., Beach, T.G., Reiman, E.M., et al. (2021). Integrating human brain proteomes with genome-wide association data implicates new proteins in Alzheimer's disease pathogenesis. Nat Genet *53*, 143–146. 10.1038/s41588-020-00773-z.

48. Ou, Y.-N., Yang, Y.-X., Deng, Y.-T., Zhang, C., Hu, H., Wu, B.-S., Liu, Y., Wang, Y.-J., Zhu, Y., Suckling, J., et al. (2021). Identification of novel drug targets for Alzheimer's disease by integrating genetics and proteomes from brain and blood. Mol Psychiatry *26*, 6065–6073. 10.1038/s41380-021-01251-6.

49. Cullell, N., Gallego-Fábrega, C., Cárcel-Márquez, J., Muiño, E., Llucià-Carol, L., Lledós, M., Martín-Campos, J.M., Molina, J., Casas, L., Almeria, M., et al. (2022). ICA1L Is Associated with Small Vessel Disease: A Proteome-Wide Association Study in Small Vessel Stroke and Intracerebral Haemorrhage. Int J Mol Sci *23*, 3161. 10.3390/ijms23063161.

50. Liu, D.-Z., and Sharp, F.R. (2012). Excitatory and Mitogenic Signaling in Cell Death, Blood–brain Barrier Breakdown, and BBB Repair after Intracerebral Hemorrhage. Transl. Stroke Res. *3*, 62–69. 10.1007/s12975-012-0147-z.

51. Lai, T.W., Zhang, S., and Wang, Y.T. (2014). Excitotoxicity and stroke: Identifying novel targets for neuroprotection. Progress in Neurobiology *115*, 157–188. 10.1016/j.pneurobio.2013.11.006.

52. Ge, Y.-J., Ou, Y.-N., Deng, Y.-T., Wu, B.-S., Yang, L., Zhang, Y.-R., Chen, S.-D., Huang, Y.-Y., Dong, Q., Tan, L., et al. (2023). Prioritization of Drug Targets for Neurodegenerative Diseases by Integrating Genetic and Proteomic Data From Brain and Blood. Biological Psychiatry *93*, 770–779. 10.1016/j.biopsych.2022.11.002.

53. Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., and Tanabe, M. (2021). KEGG: integrating viruses and cellular organisms. Nucleic Acids Research *49*, D545–D551. 10.1093/nar/gkaa970.

54. De Bruijn, M.F., and Speck, N.A. (2004). Core-binding factors in hematopoiesis and immune function. Oncogene *23*, 4238–4248. 10.1038/sj.onc.1207763.

55. Maillard, I., Adler, S.H., and Pear, W.S. (2003). Notch and the Immune System. Immunity *19*, 781–791. 10.1016/S1074-7613(03)00325-X.

56. Radtke, F., Wilson, A., Mancini, S.J.C., and MacDonald, H.R. (2004). Notch regulation of lymphocyte development and function. Nat Immunol *5*, 247–253. 10.1038/ni1045.

57. Rothenberg, E.V., and Taghon, T. (2005). MOLECULAR GENETICS OF T CELL DEVELOPMENT. Annu. Rev. Immunol. *23*, 601–649. 10.1146/annurev.immunol.23.021704.115737.

58. Shapiro-Shelef, M., and Calame, K. (2005). Regulation of plasma-cell development. Nat Rev Immunol *5*, 230–242. 10.1038/nri1572.

59. Spits, H. (2002). Development of αβ T cells in the human thymus. Nat Rev Immunol *2*, 760–772. 10.1038/nri913.

60. Lambert, J.-C., Ibrahim-Verbaas, C.A., Harold, D., Naj, A.C., Sims, R., Bellenguez, C., Jun, G., DeStefano, A.L., Bis, J.C., Beecham, G.W., et al. (2013). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet *45*, 1452–1458. 10.1038/ng.2802.

61. Hollingworth, P., Harold, D., Sims, R., Gerrish, A., Lambert, J.-C., Carrasquillo, M.M., Abraham, R., Hamshere, M.L., Pahwa, J.S., Moskvina, V., et al. (2011). Common variants in ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. Nat Genet *43*, 429–435. 10.1038/ng.803.

62. Reitz, C., and Mayeux, R. (2014). Genetics of Alzheimer's Disease in Caribbean Hispanic and African American Populations. Biological Psychiatry *75*, 534–541. 10.1016/j.biopsych.2013.06.003.

63. Reitz, C., Jun, G., Naj, A., and Rajbhandary, R. (2013). Variants in the ATP-Binding Cassette Transporter (ABCA7), Apolipoprotein E ϵ4,and the Risk of Late-Onset Alzheimer Disease in African Americans. JAMA *309*, 1483–1492. 10.1001/jama.2013.2973.

64. De Roeck, A., Van Broeckhoven, C., and Sleegers, K. (2019). The role of ABCA7 in Alzheimer's disease: evidence from genomics, transcriptomics and methylomics. Acta Neuropathol *138*, 201–220. 10.1007/s00401-019-01994-1.

65. Chatterjee, P., Roy, D., Bhattacharyya, M., and Bandyopadhyay, S. (2017). Biological networks in Parkinson's disease: an insight into the epigenetic mechanisms associated with this disease. BMC Genomics *18*, 721. 10.1186/s12864-017-4098-3.

66. Gialluisi, A., Reccia, M.G., Modugno, N., Nutile, T., Lombardi, A., Di Giovannantonio, L.G., Pietracupa, S., Ruggiero, D., Scala, S., Gambardella, S., et al. (2021). Identification of sixteen novel candidate genes for late onset Parkinson's disease. Mol Neurodegener *16*, 35. 10.1186/s13024-021-00455-2.

67. Salem, M., Ammitzboell, M., Nys, K., Seidelin, J.B., and Nielsen, O.H. (2015). ATG16L1: A multifunctional susceptibility factor in Crohn disease. Autophagy *11*, 585–594. 10.1080/15548627.2015.1017187.

68. Hansson, O., Kumar, A., Janelidze, S., Stomrud, E., Insel, P.S., Blennow, K., Zetterberg, H., Fauman, E., Hedman, Å.K., Nagle, M.W., et al. (2023). The genetic regulation of protein expression in cerebrospinal fluid. EMBO Mol Med *15*, e16359. 10.15252/emmm.202216359.

69. Cali, E., Dominik, N., Manole, A., and Houlden, H. (1993). Riboflavin Transporter Deficiency. In GeneReviews®, M. P. Adam, G. M. Mirzaa, R. A. Pagon, S. E. Wallace, L. J. Bean, K. W. Gripp, and A. Amemiya, eds. (University of Washington, Seattle).

70. Novikova, G., Kapoor, M., Tcw, J., Abud, E.M., Efthymiou, A.G., Chen, S.X., Cheng, H., Fullard, J.F., Bendl, J., Liu, Y., et al. (2021). Integration of Alzheimer's disease genetics and myeloid genomics identifies disease risk regulatory elements and genes. Nat Commun *12*, 1610. 10.1038/s41467-021-21823-y.

71. Allan, C.M., Walker, D., Segrest, J.P., and Taylor, J.M. (1995). Identification and characterization of a new human gene (APOC4) in the apolipoprotein E, C-I, and C-II gene locus. Genomics *28*, 291–300. 10.1006/geno.1995.1144.

72. Wang, P., Luo, M., Zhou, W., Jin, X., Xu, Z., Yan, S., Li, Y., Xu, C., Cheng, R., Huang, Y., et al. (2022). Global Characterization of Peripheral B Cells in Parkinson's Disease by Single-Cell RNA and BCR Sequencing. Front Immunol *13*, 814239. 10.3389/fimmu.2022.814239.

73. Lp, Y., and Mj, S. (1991). The cryptic HLA-DQA2 ("DX alpha") gene is expressed in human B cell lines. Journal of immunology (Baltimore, Md. : 1950) *147*.

74. Li, G.-J., Yang, Q.-H., Yang, G.-K., Yang, G., Hou, Y., Hou, L.-J., Li, Z.-X., and Du, L.-J. (2023). MiR-125b and SATB1-AS1 might be shear stress-mediated therapeutic targets. Gene *857*, 147181. 10.1016/j.gene.2023.147181.

75. Naguib, A., Sandmann, T., Yi, F., Watts, R.J., Lewcock, J.W., and Dowdle, W.E. (2019). SUPT4H1 Depletion Leads to a Global Reduction in RNA. Cell Rep *26*, 45-53.e4. 10.1016/j.celrep.2018.12.004.

76. Greaves, C.V., and Rohrer, J.D. (2019). An update on genetic frontotemporal dementia. J Neurol *266*, 2075–2086. 10.1007/s00415-019-09363-4.

77. Ballasch, I., García-García, E., Vila, C., Pérez-González, A., Sancho-Balsells, A., Fernández, J., Soto, D., Puigdellívol, M., Gasull, X., Alberch, J., et al. (2023). Ikzf1 as a novel regulator of microglial homeostasis in inflammation and neurodegeneration. Brain Behav Immun *109*, 144–161. 10.1016/j.bbi.2023.01.016.

78. Schutte, B., Henfling, M., and Ramaekers, F.C.S. (2006). DEDD association with cytokeratin filaments correlates with sensitivity to apoptosis. Apoptosis *11*, 1561–1572. 10.1007/s10495-006-9113-0.

79. Dahal, A., and Hinton, S.D. (2017). Antagonistic roles for STYX pseudophosphatases in neurite outgrowth. Biochem Soc Trans *45*, 381–387. 10.1042/BST20160273.

80. Burgaletto, C., Munafò, A., Di Benedetto, G., De Francisci, C., Caraci, F., Di Mauro, R., Bucolo, C., Bernardini, R., and Cantarella, G. (2020). The immune system on the TRAIL of Alzheimer's disease. Journal of Neuroinflammation *17*, 298. 10.1186/s12974-020-01968-1.

81. Traylor, M., Persyn, E., Tomppo, L., Klasson, S., Abedi, V., Bakker, M.K., Torres, N., Li, L., Bell, S., Rutten-Jacobs, L., et al. (2021). Genetic basis of lacunar stroke: a pooled analysis of individual patient data and genome-wide association studies. Lancet Neurol *20*, 351–361. 10.1016/S1474-4422(21)00031-4.

82. Chen, K., Zhu, L., Guo, L., Pan, Y.-B., and Feng, D.-F. (2020). Maf1 regulates dendritic morphogenesis and influences learning and memory. Cell Death Dis *11*, 606. 10.1038/s41419-020-02809-y.

83. Kumar, A.A., Yeo, N., Whittaker, M., Attra, P., Barrick, T.R., Bridges, L.R., Dickson, D.W., Esiri, M.M., Farris, C.W., Graham, D., et al. (2022). Vascular Collagen Type-IV in Hypertension and Cerebral Small Vessel Disease. Stroke *53*, 3696–3705. 10.1161/STROKEAHA.122.037761.

84. Miners, S., van Helmond, Z., Barker, R., Passmore, P.A., Johnston, J.A., Todd, S., McGuinness, B.M., Panza, F., Seripa, D., Solfrizzi, V., et al. (2012). Genetic variation in MME in relation to neprilysin protein and enzyme activity, Aβ levels, and Alzheimer's disease risk. Int J Mol Epidemiol Genet *3*, 30–38.

85. Chen, G., Li, L., and Tao, H. (2021). Bioinformatics Identification of Ferroptosis-Related Biomarkers and Therapeutic Compounds in Ischemic Stroke. Front Neurol *12*, 745240. 10.3389/fneur.2021.745240.

86. Pan, Y., Fu, Y., Baird, P.N., Guymer, R.H., Das, T., and Iwata, T. (2022). Exploring the contribution of ARMS2 and HTRA1 genetic risk factors in age-related macular degeneration. Prog Retin Eye Res, 101159. 10.1016/j.preteyeres.2022.101159.

87. Thee, E.F., Colijn, J.M., Cougnard-Grégoire, A., Meester-Smoor, M.A., Verzijden, T., Hoyng, C.B., Fauser, S., Hense, H.-W., Silva, R., Creuzot-Garcher, C., et al. (2022). The Phenotypic Course of Age-Related Macular Degeneration for ARMS2/HTRA1: The EYE-RISK Consortium. Ophthalmology *129*, 752–764. 10.1016/j.ophtha.2022.02.026.

88. Gatta, L.B., Vitali, M., Zanola, A., Venturelli, E., Fenoglio, C., Galimberti, D., Scarpini, E., and Finazzi, D. (2008). Polymorphisms in the LOC387715/ARMS2 putative gene and the risk for Alzheimer's disease. Dement Geriatr Cogn Disord *26*, 169–174. 10.1159/000151050.

89. Wang, P., Qin, W., Wang, P., Huang, Y., Liu, Y., Zhang, R., Li, S., Yang, Q., Wang, X., Chen, F., et al. (2018). Genomic Variants in NEURL, GJA1 and CUX2 Significantly Increase Genetic Susceptibility to Atrial Fibrillation. Sci Rep *8*, 3297. 10.1038/s41598-018-21611-7.

90. Li, R.-G., Xu, Y.-J., Ye, W.G., Li, Y.-J., Chen, H., Qiu, X.-B., Yang, Y.-Q., and Bai, D. (2021). Connexin45 (GJC1) loss-of-function mutation contributes to familial atrial fibrillation and conduction disease. Heart Rhythm *18*, 684–693. 10.1016/j.hrthm.2020.12.033.

91. Zhu, C., Xiao, F., and Lin, W. (2015). EFTUD2 on innate immunity. Oncotarget *6*, 32313–32314. 10.18632/oncotarget.5863.

92. Abo Elwafa, R., Gamaleldin, M., and Ghallab, O. (2019). The clinical and prognostic significance of FIS1, SPI1, PDCD7 and Ang2 expression levels in acute myeloid leukemia. Cancer Genet *233–234*, 84–95. 10.1016/j.cancergen.2018.12.001.

93. Vadhvani, M., Schwedhelm-Domeyer, N., Mukherjee, C., and Stegmüller, J. (2013). The centrosomal E3 ubiquitin ligase FBXO31-SCF regulates neuronal morphogenesis and migration. PLoS One *8*, e57530. 10.1371/journal.pone.0057530.

94. Serpeloni, F., Nätt, D., Assis, S.G. de, Wieling, E., and Elbert, T. (2020). Experiencing community and domestic violence is associated with epigenetic changes in DNA methylation of BDNF and CLPX in adolescents. Psychophysiology *57*, e13382. 10.1111/psyp.13382.

95. ZSCAN29 zinc finger and SCAN domain containing 29 [Homo sapiens (human)] - Gene - NCBI https://www.ncbi.nlm.nih.gov/gene/146050#summary.

96. Baker, T.A., and Sauer, R.T. (2012). ClpXP, an ATP-powered unfolding and protein-degradation machine. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research *1823*, 15–28. 10.1016/j.bbamcr.2011.06.007.

97. Thomas, T., Perdue, M.V., Khalaf, S., Landi, N., Hoeft, F., Pugh, K., and Grigorenko, E.L. (2021). Neuroimaging genetic associations between SEMA6D, brain structure, and reading skills. J Clin Exp Neuropsychol *43*, 276–289. 10.1080/13803395.2021.1912300.

98. Perdue, M.V., Mascheretti, S., Kornilov, S.A., Jasińska, K.K., Ryherd, K., Einar Mencl, W., Frost, S.J., Grigorenko, E.L., Pugh, K.R., and Landi, N. (2019). Common variation within the SETBP1 gene is associated with reading-related skills and patterns of functional neural activation. Neuropsychologia *130*, 44–51. 10.1016/j.neuropsychologia.2018.07.015.

99. Ziegler, G.C., Ehlis, A.-C., Weber, H., Vitale, M.R., Zöller, J.E.M., Ku, H.-P., Schiele, M.A., Kürbitz, L.I., Romanos, M., Pauli, P., et al. (2021). A Common CDH13 Variant Is Associated with Low Agreeableness and Neural Responses to Working Memory Tasks in ADHD. Genes *12*, 1356. 10.3390/genes12091356.

100. Trubetskoy, V., Pardiñas, A.F., Qi, T., Panagiotaropoulou, G., Awasthi, S., Bigdeli, T.B., Bryois, J., Chen, C.-Y., Dennison, C.A., Hall, L.S., et al. (2022). Mapping genomic loci implicates genes and synaptic biology in schizophrenia. Nature *604*, 502–508. 10.1038/s41586-022-04434-5.

101. Beeri, M.S., Lin, H.-M., Sano, M., Ravona-Springer, R., Liu, X., Bendlin, B.B., Gleason, C.E., Guerrero-Berroa, E., Soleimani, L., Launer, L.J., et al. (2018). Association of the Haptoglobin Gene Polymorphism With Cognitive Function and Decline in Elderly African American Adults With Type 2 Diabetes: Findings From the Action to Control Cardiovascular Risk in Diabetes-Memory in Diabetes (ACCORD-MIND) Study. JAMA Netw Open *1*, e184458. 10.1001/jamanetworkopen.2018.4458.

102. Kikuchi, M., Yamada, K., Toyota, T., and Yoshikawa, T. (2003). C18orf1 located on chromosome 18p11.2 may confer susceptibility to schizophrenia. J Med Dent Sci *50*, 225–229.

103. Zhang, Z., Shang, J., Dai, Z., Yao, Y., Shi, Y., Zhong, D., Liang, Y., Lai, C., Yang, Q., Feng, T., et al. (2022). Transmembrane Protein 170B is a Prognostic Biomarker and

Associated With Immune Infiltrates in Pancreatic Adenocarcinoma. Front Genet *13*, 848391. 10.3389/fgene.2022.848391.

104. Ge, Y., Grossman, R.I., Babb, J.S., Rabin, M.L., Mannon, L.J., and Kolson, D.L. (2002). Age-Related Total Gray Matter and White Matter Changes in Normal Adult Brain. Part I: Volumetric MR Imaging Analysis. American Journal of Neuroradiology *23*, 1327–1333.

105. Davies, G., Armstrong, N., Bis, J.C., Bressler, J., Chouraki, V., Giddaluru, S., Hofer, E., Ibrahim-Verbaas, C.A., Kirin, M., Lahti, J., et al. (2015). Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N=53 949). Mol Psychiatry *20*, 183–192. 10.1038/mp.2014.188.

106. Dumitrescu, L., Mahoney, E.R., Mukherjee, S., Lee, M.L., Bush, W.S., Engelman, C.D., Lu, Q., Fardo, D.W., Trittschuh, E.H., Mez, J., et al. (2020). Genetic variants and functional pathways associated with resilience to Alzheimer’s disease. Brain *143*, 2561–2575. 10.1093/brain/awaa209.

107. Fridman, A.L., and Tainsky, M.A. (2008). Critical pathways in cellular senescence and immortalization revealed by gene expression profiling. Oncogene *27*, 5975–5987. 10.1038/onc.2008.213.

108. Wainberg, M., Sinnott-Armstrong, N., Mancuso, N., Barbeira, A.N., Knowles, D.A., Golan, D., Ermel, R., Ruusalepp, A., Quertermous, T., Hao, K., et al. (2019). Opportunities and challenges for transcriptome-wide association studies. Nat Genet *51*, 592–599. 10.1038/s41588-019-0385-z.

109. Wu, C., and Pan, W. (2020). A powerful fine-mapping method for transcriptome-wide association studies. Hum Genet *139*, 199–213. 10.1007/s00439-019-02098-2.

4.8 Tables

Table 4-1 Sample characteristics of expression quantitative trait locus (eQTL) mapping study and genome-wide association studies (GWAS) participants.

eQTL mapping study: Genetic Epidemiology Network of Arteriopathy (GENOA)	
	Mean (SD) or N (%) or N
N	N=1833
Age (years)	56.85 (10.0)
Female	1202 (65.6%)
Race/Ethnicity	
African Americans	1032 (56.3%)
European Americans	801 (43.7%)

General cognitive function GWAS: CHARGE, COGENT, UKB	
	Mean (SD) or N (%) or N
N	300,486
Age (years)	56.91 (7.8)
Female	52.20%
Excluded for dementia and/or stroke diagnosis	N=4919

White matter hyperintensity (WMH) GWAS: CHARGE and UKB	
	Mean (SD) or N (%) or N
N	48,454
Age (years)	64.17
Female	29215 (57.6%)
WMH volume (cm^3)	7.06 (8.8)
Excluded for stroke or pathologies	N=1572

Alzheimer's Disease (AD) GWAS: EADB, GR@ACE, EADI, GERAD/PERADES, DemGene, Bonn, the Rotterdam study, the CCHS study, NxC and the UKB	
	Mean (SD) or N (%) or N
Discovery sample	
AD cases	N=39,106
AD proxy cases	N=46,828
Controls	N=401,577
Age (years)	
AD cases or proxy cases	73.55 (8.1)
controls	67.86 (8.6)
Female	
AD cases or proxy cases	62.90%
controls	56.10%

All GWAS[15,112,305] include only European ancestry participants.

Table 4-2. Genes for WMH identified both by METRO followed by fine-mapping with FOCUS and by TWAS-Fusion conducted by Sargurupremraj et al. (2020)

Gene	ENSG	Chr	Start	End	Gene Name
CALCRL	ENSG00000064989	2	188206691	188313187	calcitonin receptor like receptor
DCAKD	ENSG00000172992	17	43100706	43138499	dephospho-CoA kinase domain containing
EFEMP1	ENSG00000115380	2	56093102	56151274	EGF containing fibulin extracellular matrix protein 1
GJC1	ENSG00000182963	17	42875816	42908184	gap junction protein gamma 1
ICA1L	ENSG00000163596	2	203637873	203736489	islet cell autoantigen 1 like
KLHL24	ENSG00000114796	3	183353398	183402307	kelch like family member 24
NBEAL1	ENSG00000144426	2	203879331	204091101	neurobeachin like 1
NEURL	ENSG00000107954	10	105253462	105352303	neuralized E3 ubiquitin protein ligase 1
NMT1	ENSG00000136448	17	43035360	43186384	N-myristoyltransferase 1
WBP2	ENSG00000132471	17	73841780	73852588	WW domain binding protein 2

Abbreviations: HGNC, Human Genome Organisation Gene Nomenclature Committee

Table 4-3. Genes for AD identified both by METRO followed by fine-mapping with FOCUS and by TWAS-Fusion followed by fine-mapping with FOCUS conducted by Bellenguez et al. (2022)

Gene	ENSG	Chr	Start	End	Gene Name
BLNK	ENSG00000095585	10	97948927	98031344	B cell linker
CPSF3	ENSG00000119203	2	9563780	9613230	cleavage and polyadenylation specific factor 3
DDX54	ENSG00000123064	12	113594978	113623284	DEAD-box helicase 54
GRN	ENSG00000030582	17	42422614	42430474	granulin precursor
ICA1L	ENSG00000163596	2	203637873	203736489	islet cell autoantigen 1 like
KLF16	ENSG00000129911	19	1852398	1863578	KLF transcription factor 16
LACTB	ENSG00000103642	15	63414032	63434260	lactamase beta
PPP4C	ENSG00000149923	16	30087299	30096697	protein phosphatase 4 catalytic subunit
SHARPIN	ENSG00000179526	8	145153536	145163027	SHANK associated RH domain interactor
TBX6	ENSG00000149922	16	30097114	30103245	T-box transcription factor 6

Abbreviations: HGNC, Human Genome Organisation Gene Nomenclature Committee

References

1. Davies, G., Lam, M., Harris, S.E., Trampush, J.W., Luciano, M., Hill, W.D., Hagenaars, S.P., Ritchie, S.J., Marioni, R.E., Fawns-Ritchie, C., et al. (2018). Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. Nat Commun *9*, 2098. 10.1038/s41467-018-04362-x.

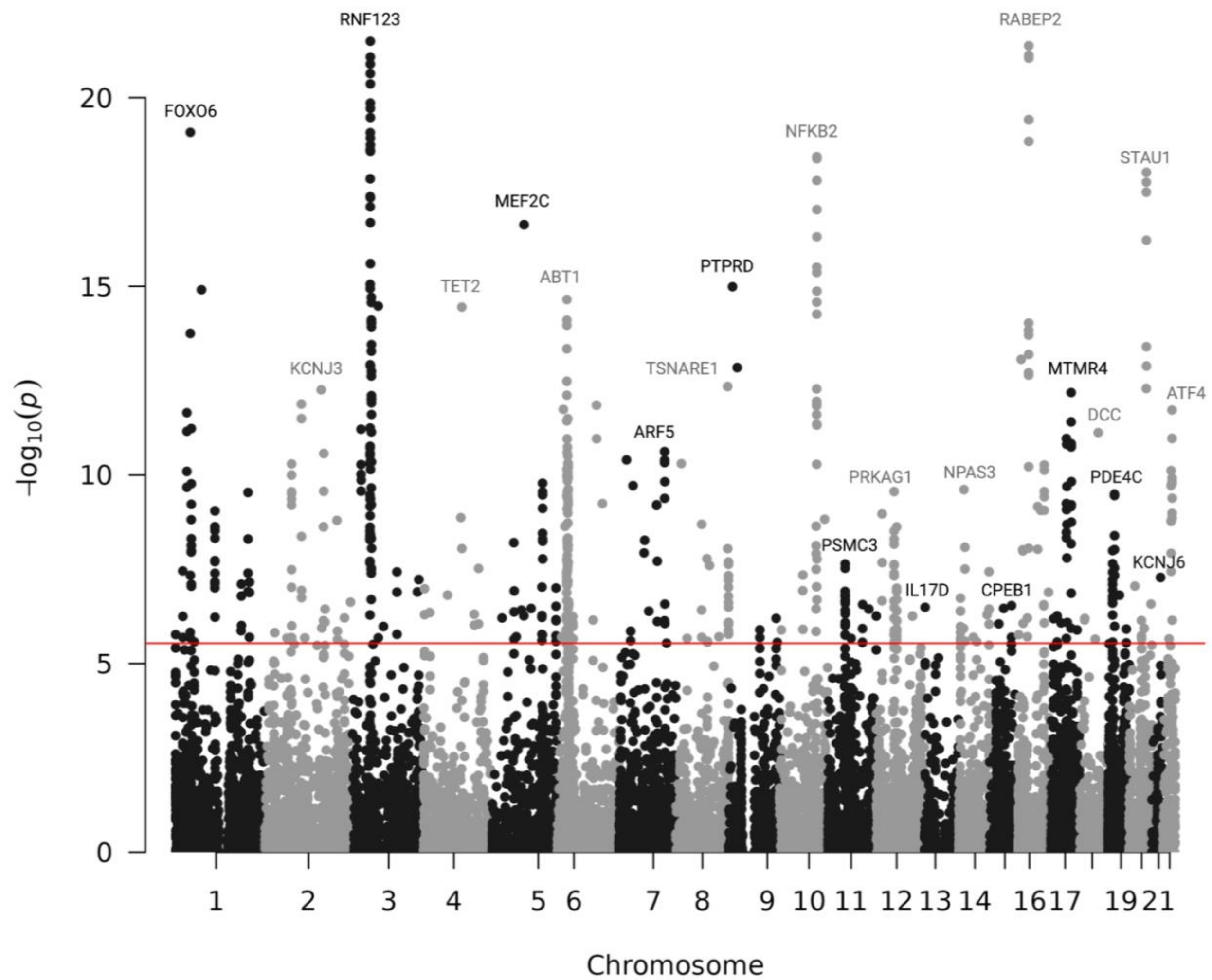
2. Sargurupremraj, M., Suzuki, H., Jian, X., Sarnowski, C., Evans, T.E., Bis, J.C., Eiriksdottir, G., Sakaue, S., Terzikhan, N., Habes, M., et al. (2020). Cerebral small vessel disease genomics and its implications across the lifespan. Nature Communications *11*, 6285. 10.1038/s41467-020-19111-2.

3. Bellenguez, C., Küçükali, F., Jansen, I.E., Kleineidam, L., Moreno-Grau, S., Amin, N., Naj, A.C., Campos-Martin, R., Grenier-Boley, B., Andrade, V., et al. (2022). New insights into the genetic etiology of Alzheimer′s disease and related dementias. Nat Genet, 1–25. 10.1038/s41588-022-01024-z.

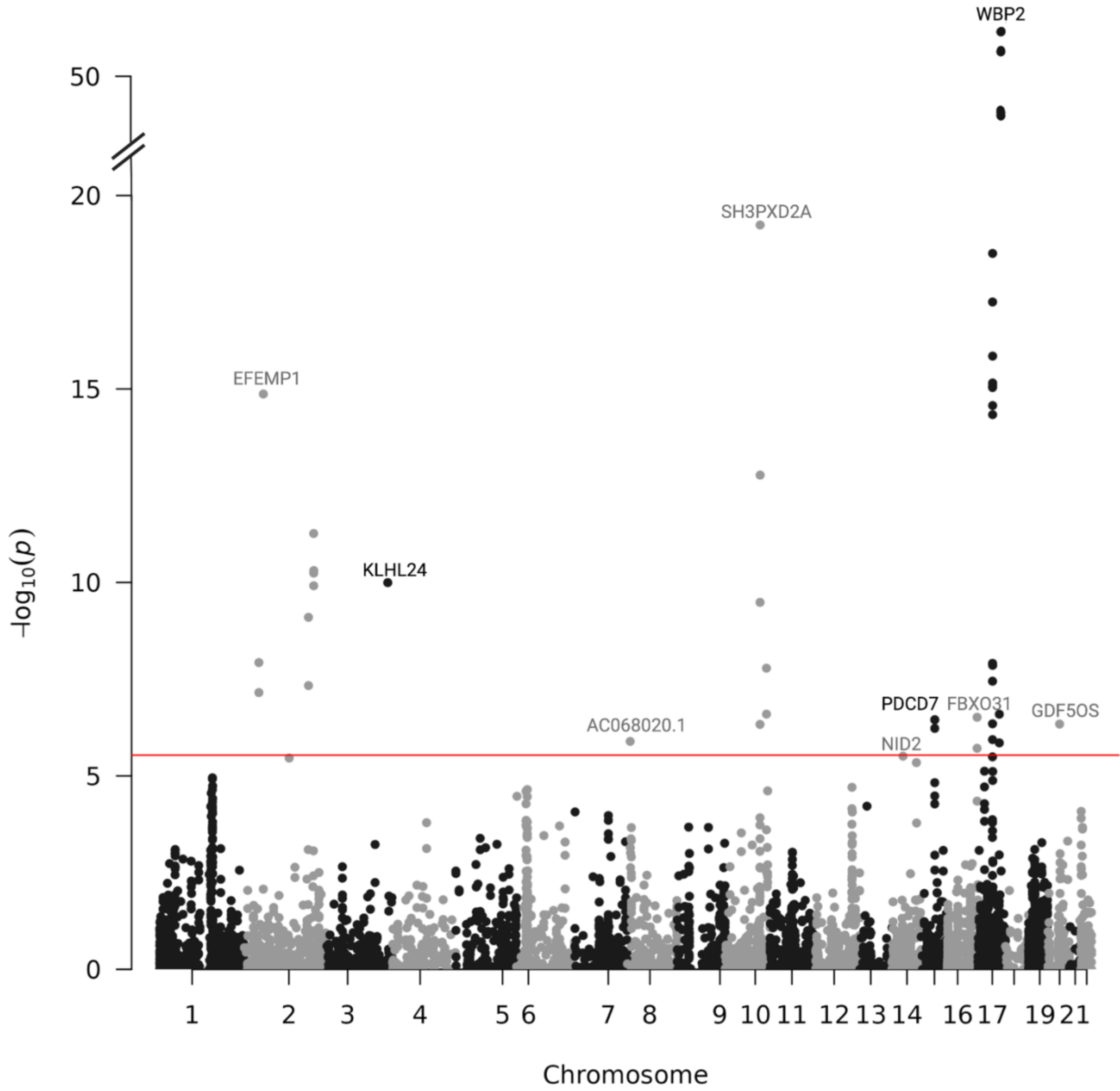
4.9 Figures

Figure 4-1. Manhattan plots of -log$_{10}$ p-values for gene-trait associations in METRO.

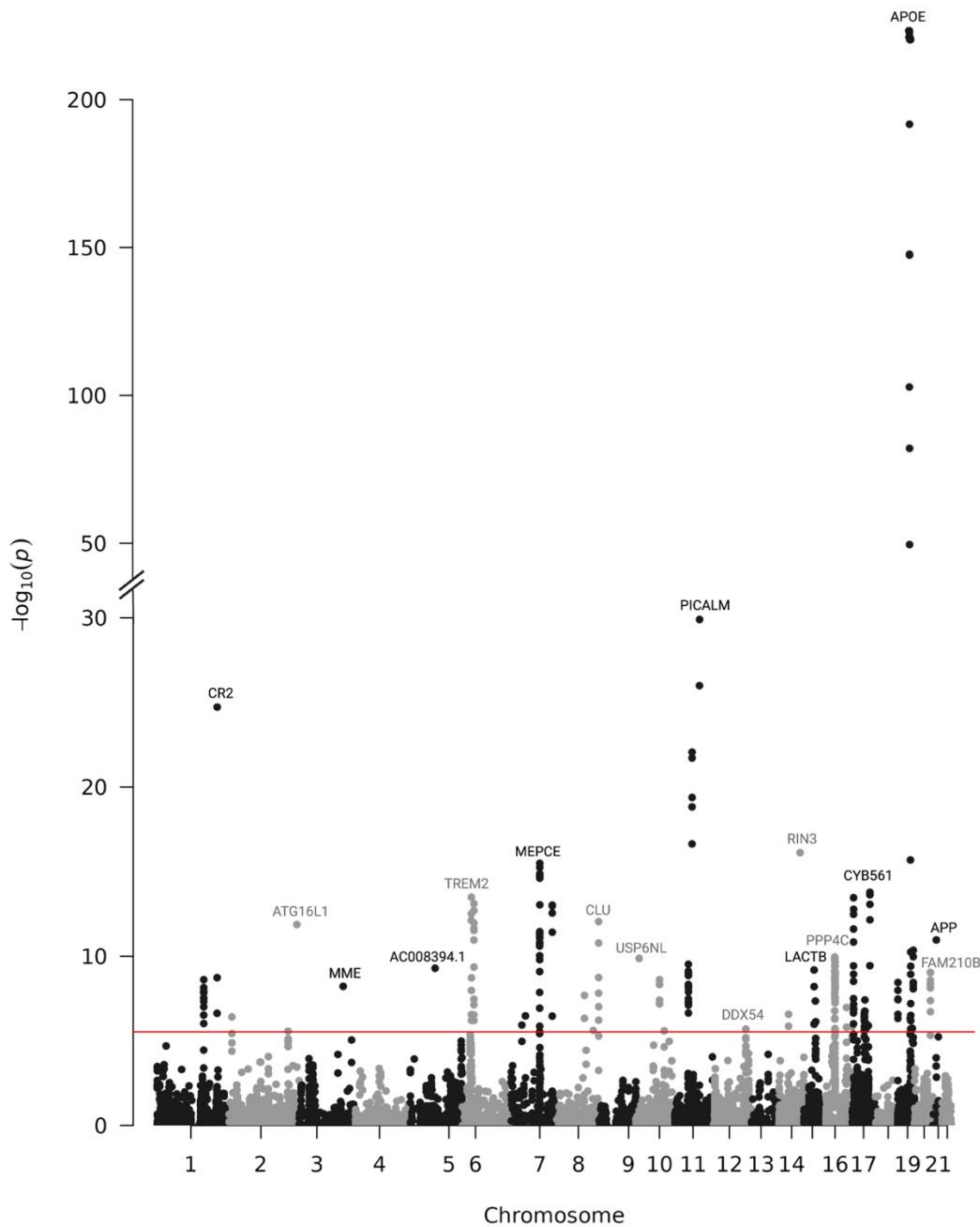
(a) General cognitive function



(b) White matter hyperintensity

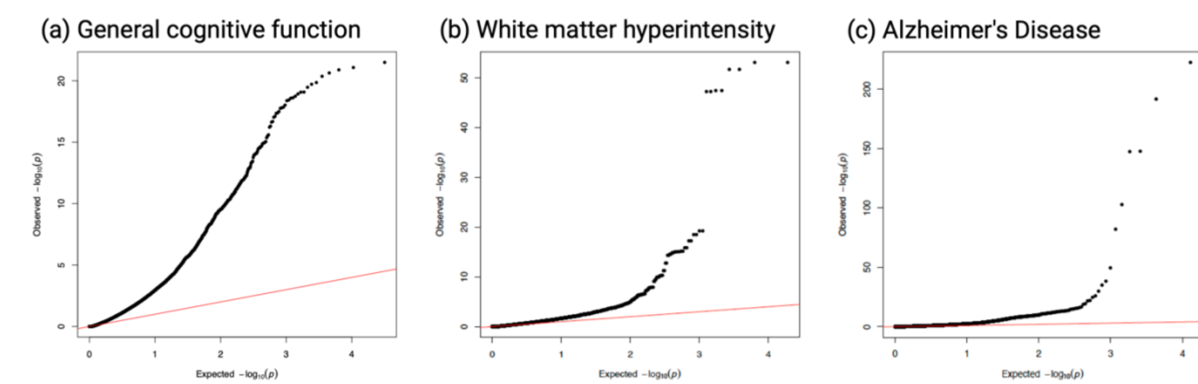


(c) Alzheimer's Disease



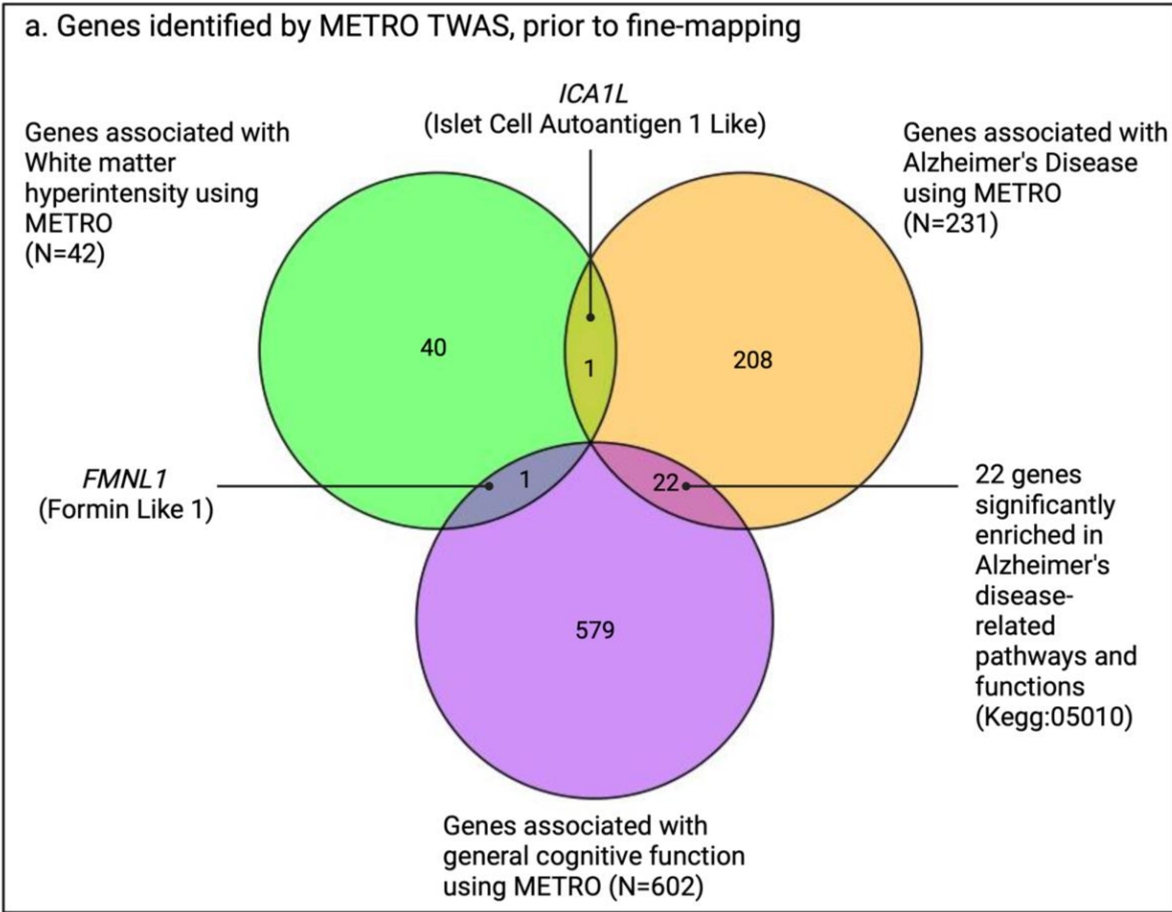
Manhattan plots of $-\log_{10}$ p-values in METRO for the associations between genes and (a) general cognitive function using summary statistics from Davies et al. (2018),[15] (b) White matter hyperintensity from Sargurupremraj et al. (2020)[375] and (c) Alzheimer's disease from Bellenguez et al. (2022),[112] using GENOA gene expression data. The red line indicates significance after Bonferroni correction ($P<2.90\times10^{-6}$).

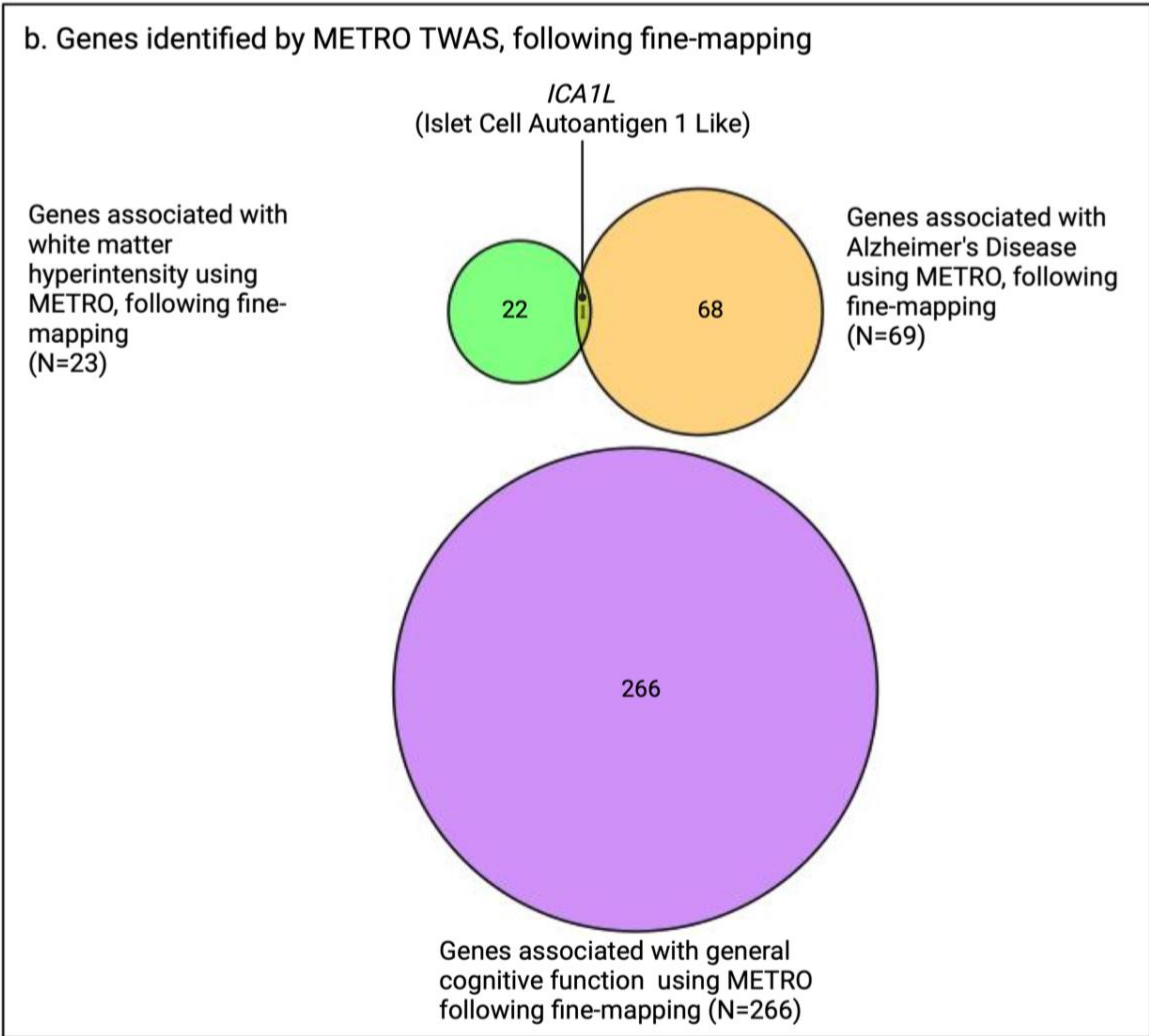
Figure 4-2. Quantile-quantile plots of -$\log_{10}$ p-values for gene-trait associations in METRO.



Q-Q plots of the associations between genes and (a) general cognitive function (λ= 2.55) using summary statistics from Davies et al. (2018),[15] (b) white matter hyperintensity (λ= 1.45) from Sargurupremraj et al. (2020)[375] and (c) Alzheimer's disease (λ= 2.09) from Bellenguez et al. (2022)[112] using GENOA gene expression data.

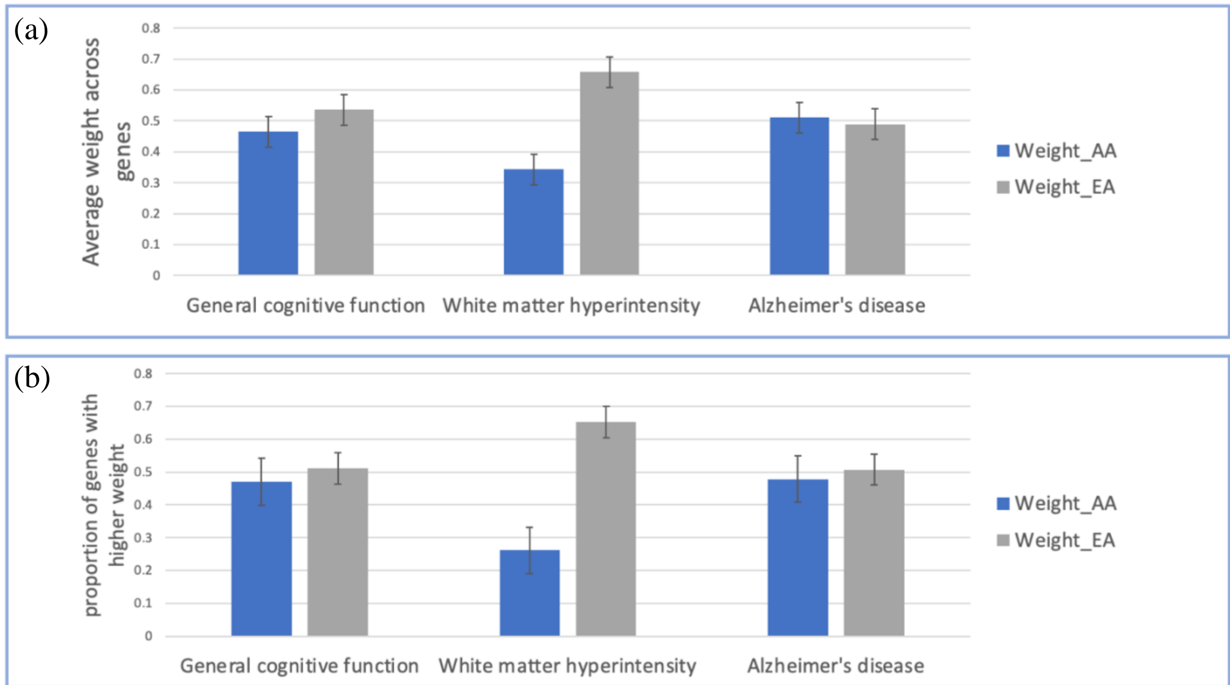
Figure 4-3. Venn diagrams comparing number of genes associated with general cognitive function, white matter hyperintensity and Alzheimer's disease using METRO, prior to and following FOCUS fine-mapping.





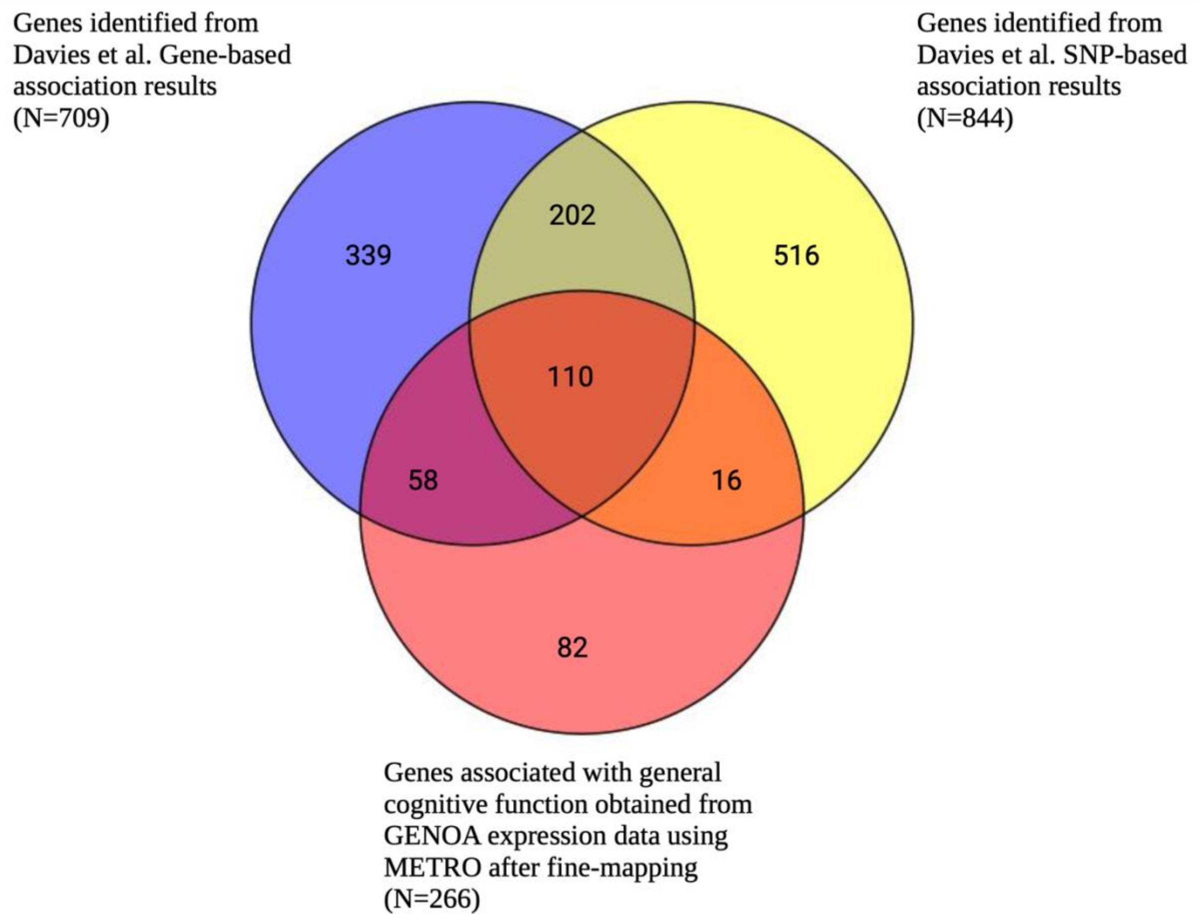
Venn diagrams comparing the number of genes associated with general cognitive function (purple; N=266 genes), white matter hyperintensity (WMH; green; N=23 genes) and Alzheimer's disease (AD; yellow; N=69 genes) (a) prior to fine-mapping and (b) following FOCUS[312] fine-mapping using METRO and GENOA expression data after Bonferroni correction ($P<2.90 \times 10^{-6}$), with GWAS summary statistics obtained from the Davies et al. (2018),[15] Sargurupremraj et al. (2020)[375] and Bellenguez et al. (2022).[112]

Figure 4-4. Contribution weights of expression prediction models across all significant genes identified by METRO.



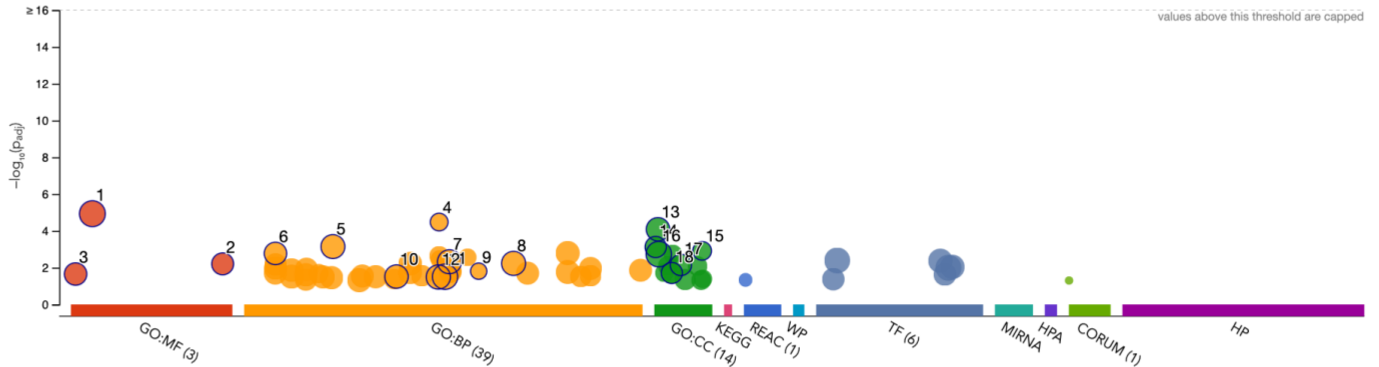
Bar plots of general cognitive function, white matter hyperintensity and Alzheimer's disease comparing (a) the average contribution weights of expression prediction models from African ancestry (AA) and European ancestry (EA) and (b) the proportion of significant genes with higher contribution weights of expression prediction models across all significant genes ($P<2.90\times10^{-6}$). Black bars are the standard errors for the estimated proportions.

Figure 4-5. Venn diagram comparing number of METRO-identified genes associated with general cognitive function following FOCUS fine-mapping and genes identified by Davies et al. (2018) gene-based and SNP-based analyses.



Venn diagram comparing the number of genes associated with general cognitive function obtained from METRO using GENOA gene expression data after Bonferroni correction ($P<2.90x10^{-6}$) and fine-mapping (red) and Davies et al. (2018).[15] Davies et al. results included SNP-based association results that were mapped to the nearest gene ($P<5x10^{-8}$; yellow), and gene-based association results ($P<2.75x10^{-6}$; blue).

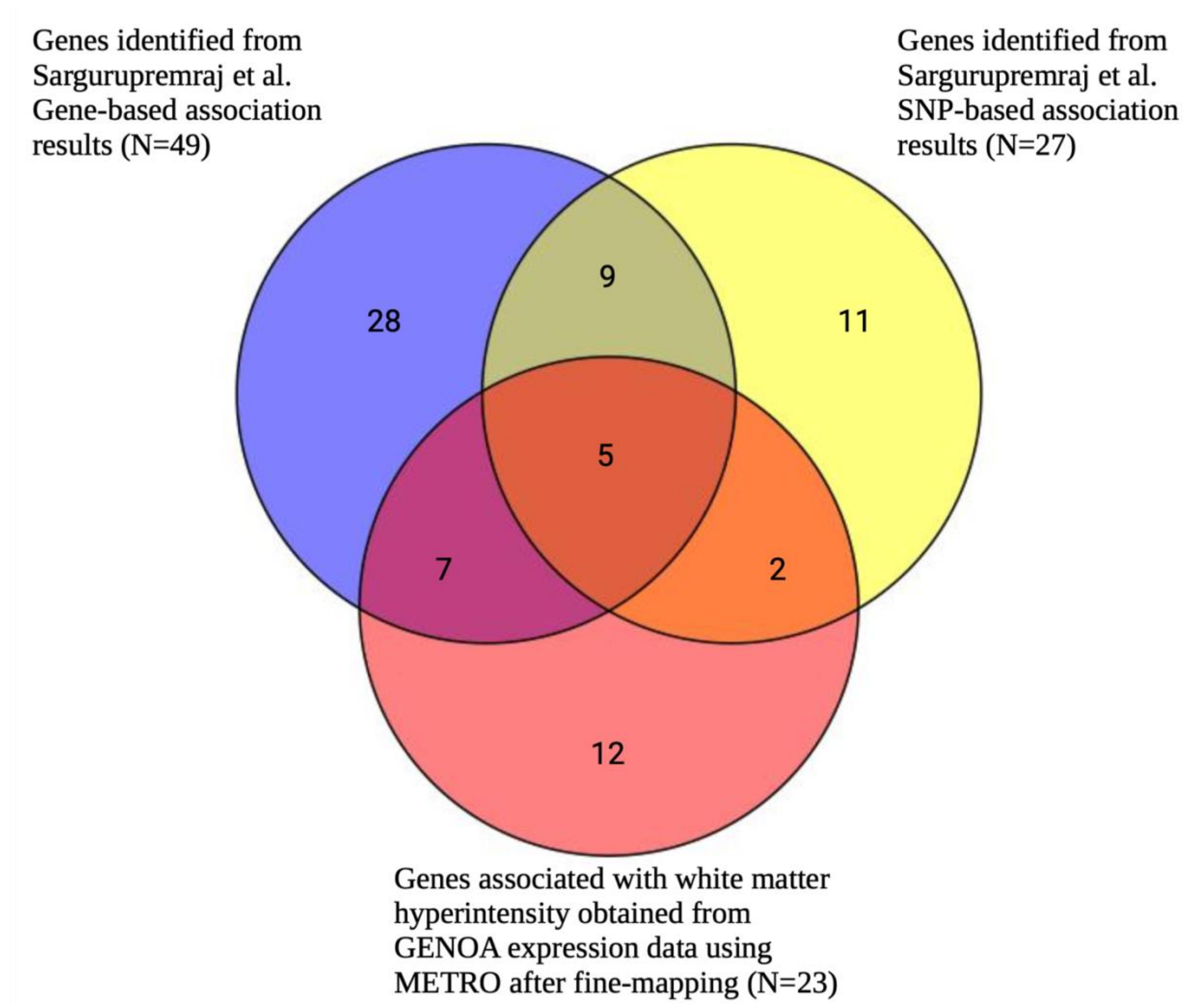
Figure 4-6. Functional enrichment analysis on the fine-mapped gene set identified for general cognitive function using METRO TWAS (N=266 genes).



ID	Term ID	Term Name	Adjusted P-value
		GO:MF	
1	GO:0005515	protein binding	1.17E-05
2	GO:0140110	Transcription regulator activity	6.23E-03
3	GO:0003677	DNA binding	2.23E-02
		GO:BP	
4	GO:0048588	developmental cell growth	3.33E-05
5	GO:0019538	Protein metabolic process	7.18E-04
6	GO:0006357	Regulation of transcription by RNA polymerase II	4.79E-03
7	GO:0051171	Regulation of nitrogen compound metabolic processes	4.79E-03
8	GO:0080090	Regulation of primary metabolic process	5.86E-03
9	GO:0061387	Regulation of extent of cell growth	1.56E-02
10	GO:0042221	Response to chemical	3.06E-02
11	GO:0050794	Regulation of cellular process	3.13E-02
12	GO:0048518	Positive regulation of biological process	3.24E-02
		GO:CC	
13	GO:0005654	nucleoplasm	8.23E-05
14	GO:0000785	chromatin	7.56E-04
15	GO:0098984	Neuron to neuron synapse	1.22E-03
16	GO:0005737	cytoplasm	1.84E-03
17	GO:0043005	neuron projection	7.14E-03
18	GO:0031967	organelle envelope	1.96E-02

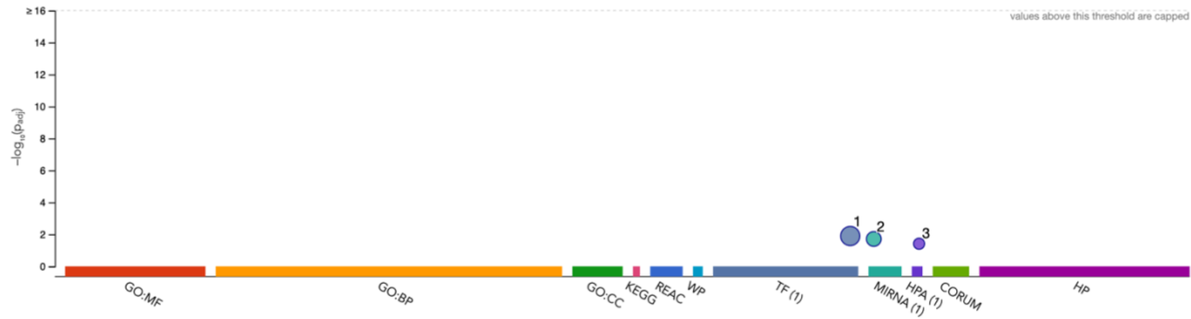
The top panel consists of a Manhattan plot that illustrates the enrichment analysis results. The x-axis represents functional terms that are grouped and color-coded by data sources, including Gene Ontology (GO): molecular function (MF; red), GO: biological process (BP; orange), GO: cellular component (CC; dark green), Kyoto Encyclopedia of Genes and Genomes (KEGG; pink), Reactome (REAC; dark blue), WikiPathways (WP; turquoise), Transfac (TF; light blue), MiRTarBase (MIRNA; emerald green), Human Protein Atlas (HPA; dark purple), CORUM protein complexes (light green), and Human Phenotype Ontology (HP; violet), in order from left to right. The y-axis shows the adjusted enriched $-\log_{10}$ p-values <0.05. Multiple testing correction was performed using g:SCS method (Set Counts and Sizes) that takes into account overlapping terms. The top panel highlights driver GO terms identified using the greedy filtering algorithm in g:Profiler. The light circles represent terms that were not significant after filtering. The circle sizes are in accordance with the corresponding term size (i.e., larger terms have larger circles). The number in parentheses following the source name in the x-axis shows how many significantly enriched terms were from this source.

Figure 4-7. Venn diagram comparing number of METRO-identified genes associated with white matter hyperintensity following FOCUS fine-mapping and genes identified by Sargurupremraj et al. (2020) gene-based and SNP-based analyses.



Venn diagram comparing the number of significantly associated genes associated with white matter hyperintensity (WMH) obtained from METRO using GENOA expression data after Bonferroni correction ($P<2.90x10^{-6}$), and fine-mapping (red) and Sargurupremraj et al. (2020).[375] Sargurupremraj et al. results included SNP-based association results that were mapped to the nearest gene ($P<5x10^{-8}$; yellow), and gene-based association results ($P<2.77x10^{-6}$; blue).

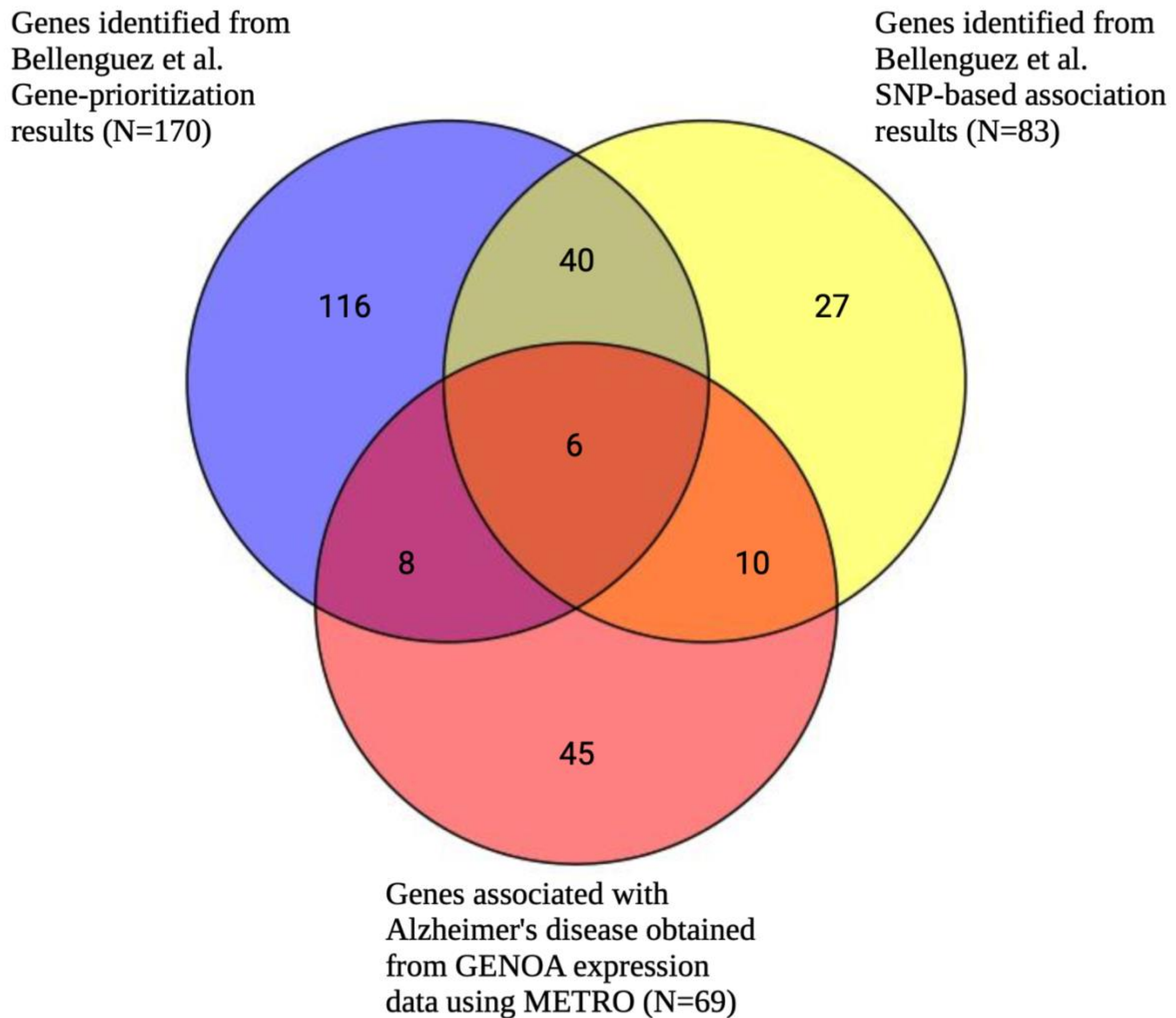
Figure 4-8. Functional enrichment analysis on the fine-mapped gene set identified for white matter hyperintensity using METRO TWAS (N=23 genes).



ID	Term ID	Term Name	Adjusted P-value
		TF	
1	TF:M12713	Factor: ZNF690; motif: GTCTACRCNG	1.27E-02
		MIRNA	
2	MIRNA: has-miR-212-5p	hsa-miR-212-5p	1.94E-02
		HPA	
3	HPA:0411242	Retina; inner plexiform layer [≥Medium]	3.86E-02

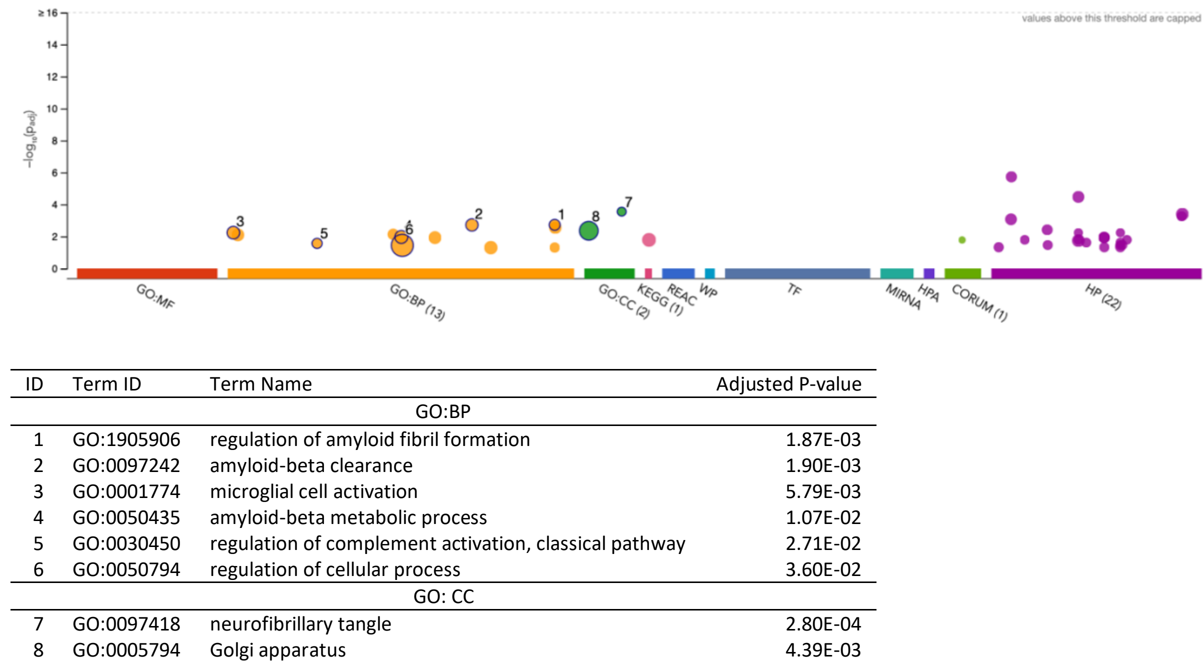
The top panel consists of a Manhattan plot that illustrates the enrichment analysis results. The x-axis represents functional terms that are grouped and color-coded by data sources, including Gene Ontology (GO): molecular function (MF; red), GO: biological process (BP; orange), GO: cellular component (CC; dark green), Kyoto Encyclopedia of Genes and Genomes (KEGG; pink), Reactome (REAC; dark blue), WikiPathways (WP; turquoise), Transfac (TF; light blue), MiRTarBase (MIRNA; emerald green), Human Protein Atlas (HPA; dark purple), CORUM protein complexes (light green), and Human Phenotype Ontology (HP; violet), in order from left to right. The y-axis shows the adjusted enriched -$\log_{10}$ p-values < 0.05. Multiple testing correction was performed using g:SCS method (Set Counts and Sizes) that takes into account overlapping terms. The top panel highlights driver GO terms identified using the greedy filtering algorithm in g:Profiler. The light circles represent terms that were not significant after filtering. The circle sizes are in accordance with the corresponding term size (i.e., larger terms have larger circles). The number in parentheses following the source name in the x-axis shows how many significantly enriched terms were from this source.

Figure 4-9. Venn diagram comparing number of METRO-identified genes associated with Alzheimer's disease following FOCUS fine-mapping and genes identified by Bellenguez et al. (2020) gene prioritization and SNP-based analyses.



Venn diagram comparing the number of significantly associated genes associated with Alzheimer's disease in European ancestry obtained from METRO using GENOA expression data after Bonferroni correction ($P<2.90x10^{-6}$) and fine-mapping (red) and Bellenguez et al. (2022).[112] Bellenguez et al. results included SNP-based association results that were mapped to the nearest gene ($P<5x10^{-8}$; yellow), and gene prioritization results for the genes in the novel AD risk loci (blue). In the gene prioritization analysis, Bellenguez et al. analyzed the downstream effects of new AD-associated loci on molecular phenotypes (i.e., expression, splicing, protein expression, methylation and histone acetylation) in various *cis*-quantitative trait loci (*cis*-QTL) catalogues from AD-relevant tissues, cell types and brain regions.

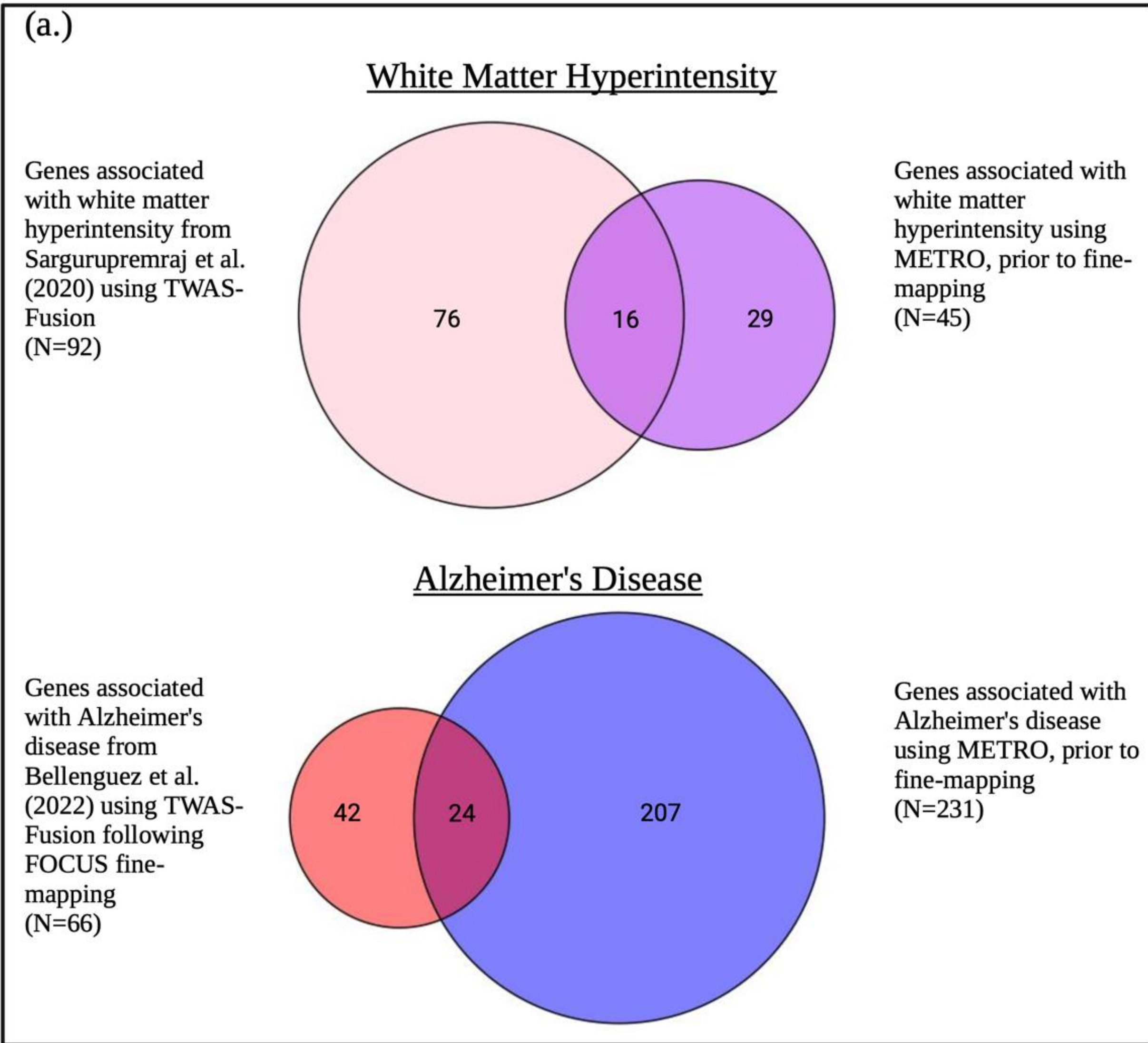
Figure 4-10. Functional enrichment analysis on the fine-mapped gene set identified for Alzheimer's disease using METRO TWAS (N=69 genes).

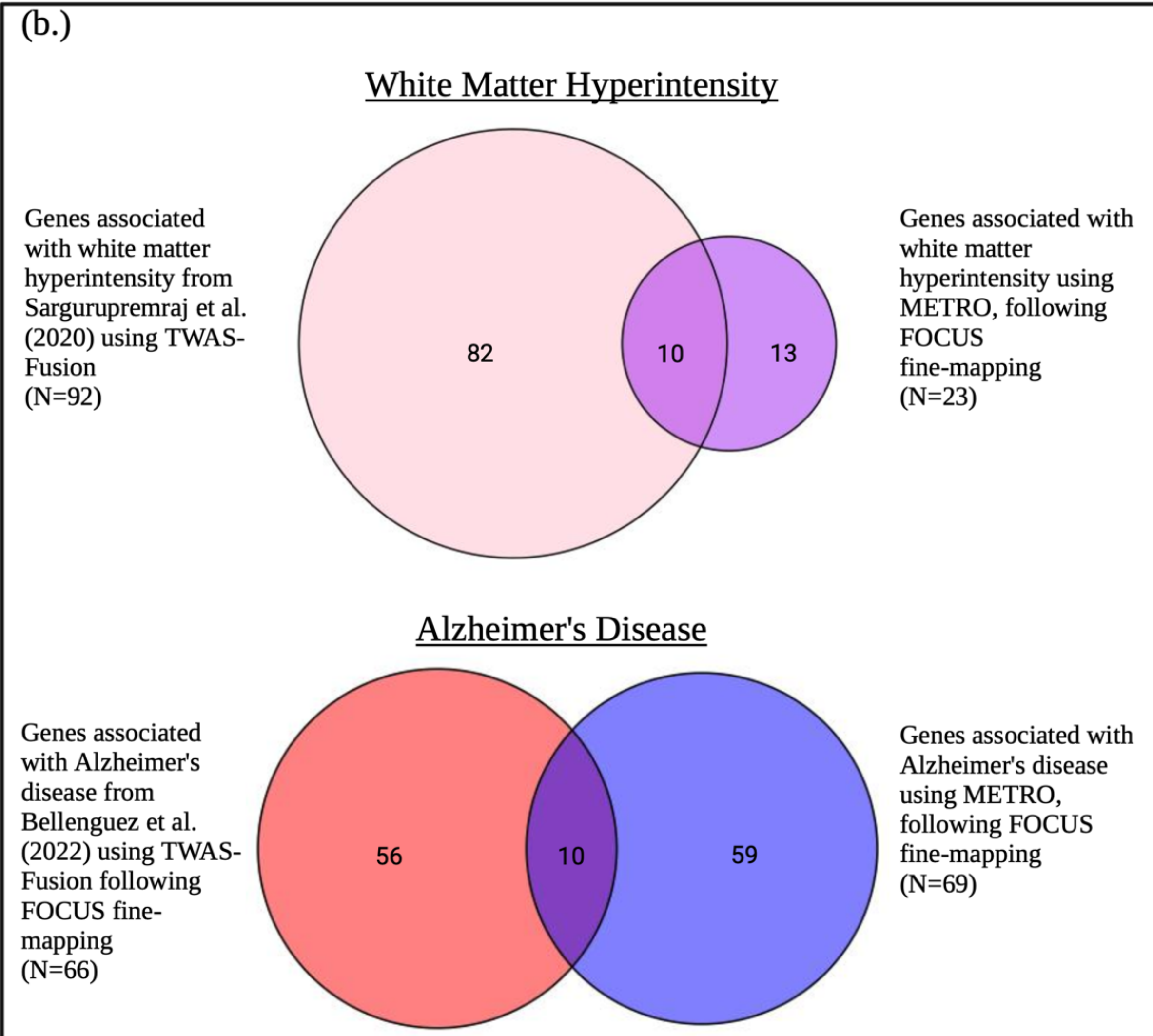


ID	Term ID	Term Name	Adjusted P-value
		GO:BP	
1	GO:1905906	regulation of amyloid fibril formation	1.87E-03
2	GO:0097242	amyloid-beta clearance	1.90E-03
3	GO:0001774	microglial cell activation	5.79E-03
4	GO:0050435	amyloid-beta metabolic process	1.07E-02
5	GO:0030450	regulation of complement activation, classical pathway	2.71E-02
6	GO:0050794	regulation of cellular process	3.60E-02
		GO: CC	
7	GO:0097418	neurofibrillary tangle	2.80E-04
8	GO:0005794	Golgi apparatus	4.39E-03

The top panel consists of a Manhattan plot that illustrates the enrichment analysis results. The x-axis represents functional terms that are grouped and color-coded by data sources, including Gene Ontology (GO): molecular function (MF; red), GO: biological process (BP; orange), GO: cellular component (CC; dark green), Kyoto Encyclopedia of Genes and Genomes (KEGG; pink), Reactome (REAC; dark blue), WikiPathways (WP; turquoise), Transfac (TF; light blue), MiRTarBase (MIRNA; emerald green), Human Protein Atlas (HPA; dark purple), CORUM protein complexes (light green), and Human Phenotype Ontology (HP; violet), in order from left to right. The y-axis shows the adjusted enriched -$\log_{10}$ p-values < 0.05. Multiple testing correction was performed using g:SCS method (Set Counts and Sizes) that takes into account overlapping terms. The top panel highlights driver GO terms identified using the greedy filtering algorithm in g:Profiler. The light circles represent terms that were not significant after filtering. The circle sizes are in accordance with the corresponding term size (i.e., larger terms have larger circles). The number in parentheses following the source name in the x-axis shows how many significantly enriched terms were from this source.

Figure 4-11. Venn diagram comparing METRO TWAS results prior to and following FOCUS fine-mapping with TWAS results from Sargurupremraj et al. (2020) and Bellenguez et al. (2022).





Venn diagram comparing METRO TWAS results (a) prior to and (b) following FOCUS[312] fine-mapping with TWAS results using Fusion for white matter hyperintensity from Sargurupremraj et al.[375] (2020) without fine-mapping and Alzheimer's disease from Bellenguez et al.[112] (2022) with FOCUS fine-mapping.

References

1. Davies, G., Lam, M., Harris, S.E., Trampush, J.W., Luciano, M., Hill, W.D., Hagenaars, S.P., Ritchie, S.J., Marioni, R.E., Fawns-Ritchie, C., et al. (2018). Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. Nat Commun *9*, 2098. 10.1038/s41467-018-04362-x.

2. International Network against Thrombosis (INVENT) Consortium, International Headache Genomics Consortium (IHGC), Sargurupremraj, M., Suzuki, H., Jian, X., Sarnowski, C., Evans, T.E., Bis, J.C., Eiriksdottir, G., Sakaue, S., et al. (2020). Cerebral small vessel disease genomics and its implications across the lifespan. Nat Commun *11*, 6285. 10.1038/s41467-020-19111-2.

3. Bellenguez, C., Küçükali, F., Jansen, I.E., Kleineidam, L., Moreno-Grau, S., Amin, N., Naj, A.C., Campos-Martin, R., Grenier-Boley, B., Andrade, V., et al. (2022). New insights into the genetic etiology of Alzheimer's disease and related dementias. Nat Genet, 1–25. 10.1038/s41588-022-01024-z.

4. Mancuso, N., Freund, M.K., Johnson, R., Shi, H., Kichaev, G., Gusev, A., and Pasaniuc, B. (2019). Probabilistic fine-mapping of transcriptome-wide association studies. Nature genetics *51*, 675–682.

4.10 Supplementary Material

Table 4-4. Genes associated with general cognitive function using METRO followed by fine-mapping with FOCUS (N=266 genes; $P<2.9x10^{-6}$)

Gene	ENSG	alpha	w1	w2	P value	chr	Start	End
RNF123	ENSG00000164068	-0.23	0.58	0.42	3.20E-22	3	49726971	49758962
RABEP2	ENSG00000177548	-1.16	0.68	0.32	4.20E-22	16	28915742	28947847
GMPPB	ENSG00000173540	-0.28	1.00	0.00	8.35E-22	3	49754277	49761406
MST1	ENSG00000173531	-0.33	0.00	1.00	1.29E-21	3	49721380	49726934
APEH	ENSG00000164062	0.95	0.00	1.00	2.30E-21	3	49711447	49721404
IP6K1	ENSG00000176095	-0.76	0.62	0.38	4.32E-21	3	49761727	49823975
UBA7	ENSG00000182179	0.11	0.76	0.24	1.92E-20	3	49842642	49851386
TUFM	ENSG00000178952	-0.24	0.89	0.11	3.85E-20	16	28853732	28857669
FOXO6	ENSG00000204060	1.15	0.87	0.13	8.29E-20	1	41827594	41849262
NFKB2	ENSG00000077150	0.30	0.00	1.00	3.67E-19	10	104153867	104162281
PSD	ENSG00000059915	0.21	0.00	1.00	4.16E-19	10	104162374	104181296
STAU1	ENSG00000124214	0.17	0.52	0.48	9.53E-19	20	47729876	47804904
CSE1L	ENSG00000124207	-0.55	0.61	0.39	1.74E-18	20	47662783	47713497
ARFGEF2	ENSG00000124198	0.11	1.00	0.00	3.20E-18	20	47538248	47653230
USP4	ENSG00000114316	-0.76	0.27	0.73	4.14E-18	3	49314577	49378145
MEF2C	ENSG00000081189	-0.26	0.57	0.43	2.32E-17	5	88012934	88200074
PTPRD	ENSG00000153707	-0.70	1.00	0.00	1.03E-15	9	8314246	10613002
NEGR1	ENSG00000172260	-0.28	0.22	0.78	1.23E-15	1	71861626	72748222
SEMA3G	ENSG00000010319	-2.55	0.00	1.00	1.96E-15	3	52467051	52479119
ABT1	ENSG00000146109	0.61	0.39	0.61	2.24E-15	6	26597181	26600967
FOXP1	ENSG00000114861	1.10	0.93	0.07	3.34E-15	3	71003844	71633129
TET2	ENSG00000168769	-0.12	0.83	0.17	3.55E-15	4	106067032	106200973
ZNF322	ENSG00000181315	0.33	0.10	0.90	7.88E-15	6	26634611	26659980
HMGN4	ENSG00000182952	-0.09	0.00	1.00	1.08E-14	6	26538594	26547161
NISCH	ENSG00000010322	-0.25	1.00	0.00	5.24E-14	3	52489134	52527084
RBFOX1	ENSG00000078328	0.22	0.96	0.04	8.64E-14	16	5289803	7763342
ELAVL2	ENSG00000107105	-0.56	0.00	1.00	1.43E-13	9	23690102	23826335
IL27	ENSG00000197272	1.73	1.00	0.00	1.95E-13	16	28510683	28523372
TSNARE1	ENSG00000171045	-0.20	0.00	1.00	4.54E-13	8	143293441	143484543
KCNJ3	ENSG00000162989	0.34	0.00	1.00	5.53E-13	2	155554367	155714866
MTMR4	ENSG00000108389	-0.12	0.00	1.00	6.56E-13	17	56566890	56595266
AFF3	ENSG00000144218	-0.31	0.00	1.00	1.32E-12	2	100161881	100808890
LACE1	ENSG00000135537	-0.24	0.27	0.73	1.42E-12	6	108616195	108847999
ATXN1	ENSG00000124788	0.82	0.00	1.00	1.83E-12	6	16299343	16761722
PEF1	ENSG00000162517	0.11	0.79	0.21	2.25E-12	1	32095467	32110497
LONRF2	ENSG00000170500	0.45	0.00	1.00	3.21E-12	2	100888337	100938963
OR2J1	ENSG00000204702	-0.12	1.00	0.00	3.23E-12	6	29067267	29070478
HSF5	ENSG00000176160	-0.73	0.60	0.40	3.93E-12	17	56497528	56565769
ST3GAL3	ENSG00000126091	-0.09	0.90	0.10	5.88E-12	1	44171495	44396837
NKIRAS1	ENSG00000197885	-0.08	1.00	0.00	6.15E-12	3	23931442	23988082
COL16A1	ENSG00000084636	-1.07	0.00	1.00	6.99E-12	1	32117864	32169920
DCC	ENSG00000187323	-2.90	0.00	1.00	7.58E-12	18	49866567	51062273
FOXO3	ENSG00000118689	-1.70	0.00	1.00	1.09E-11	6	108881038	109005977
PRSS16	ENSG00000112812	0.30	0.53	0.47	1.11E-11	6	27215480	27224403
4-SEP	ENSG00000108387	-0.60	0.00	1.00	1.49E-11	17	56597611	56621729
QRICH1	ENSG00000198218	0.09	0.00	1.00	1.73E-11	3	49067140	49131796
RNF43	ENSG00000108375	-0.21	0.00	1.00	1.82E-11	17	56431037	56494956
ZNF193	ENSG00000137185	0.25	0.41	0.59	1.84E-11	6	28192664	28201265
ARF5	ENSG00000004059	0.15	1.00	0.00	2.40E-11	7	127228440	127231754
ZNF184	ENSG00000096654	0.21	1.00	0.00	2.54E-11	6	27418522	27440897
DPP4	ENSG00000197635	-1.30	0.02	0.98	2.70E-11	2	162848755	162930904
OR2J3	ENSG00000204701	-0.29	0.00	1.00	3.13E-11	6	29075835	29082547
SP4	ENSG00000105866	0.09	1.00	0.00	3.98E-11	7	21467661	21554440
FSCN3	ENSG00000106328	0.51	0.90	0.10	4.04E-11	7	127231463	127242198

GCC1	ENSG00000179562	0.13	0.88	0.12	4.72E-11	7	127220682	127233665
OR2H1	ENSG00000204688	-0.23	0.00	1.00	4.98E-11	6	29424932	29432105
SGCZ	ENSG00000185053	0.33	1.00	0.00	5.00E-11	8	13942354	15095940
FBXO41	ENSG00000163013	-0.21	0.99	0.01	5.10E-11	2	73481810	73511606
NR1D2	ENSG00000174738	-0.17	0.00	1.00	5.30E-11	3	23986777	24022108
DHODH	ENSG00000102967	-0.15	0.90	0.10	5.44E-11	16	72042487	72061563
HP	ENSG00000257017	-1.24	0.19	0.81	7.42E-11	16	72088404	72094954
ATF4	ENSG00000128272	-0.48	0.09	0.91	7.66E-11	22	39915700	39918688
TRIM27	ENSG00000204713	-0.92	0.00	1.00	9.36E-11	6	28870779	28891765
HIST1H2AG	ENSG00000196787	-0.17	0.00	1.00	1.00E-10	6	27100822	27101314
CCT7	ENSG00000135624	0.15	0.00	1.00	1.01E-10	2	73460548	73480149
THRB	ENSG00000151090	0.48	0.00	1.00	1.37E-10	3	24158644	24537247
ZKSCAN4	ENSG00000187626	0.42	0.23	0.77	1.44E-10	6	28209475	28220047
SND1	ENSG00000197157	-1.77	0.01	0.99	1.51E-10	7	127292248	127732661
PURA	ENSG00000185129	0.56	0.94	0.06	1.65E-10	5	139487362	139505204
SLC6A9	ENSG00000196517	0.25	0.37	0.63	1.71E-10	1	44457172	44497139
MGAT3	ENSG00000128268	0.38	0.00	1.00	1.91E-10	22	39853017	39888199
POU6F2	ENSG00000106536	0.55	0.00	1.00	1.92E-10	7	39017509	39532694
PPM1M	ENSG00000164088	0.12	0.58	0.42	2.21E-10	3	52279775	52284615
NPAS3	ENSG00000151322	0.72	1.00	0.00	2.44E-10	14	33403602	34290069
IST1	ENSG00000182149	0.17	0.00	1.00	2.75E-10	16	71879899	71965102
PRKAG1	ENSG00000181929	0.11	1.00	0.00	2.76E-10	12	49396057	49412590
HIST1H2BL	ENSG00000185130	-0.05	0.03	0.97	2.77E-10	6	27775257	27775707
PRADC1	ENSG00000135617	0.32	0.63	0.37	2.79E-10	2	73455138	73460367
CYSTM1	ENSG00000120306	0.35	0.00	1.00	2.82E-10	5	139554741	139661637
IPO9	ENSG00000198700	0.09	0.00	1.00	2.92E-10	1	201798277	201853419
EGR4	ENSG00000135625	0.95	0.87	0.13	3.14E-10	2	73518057	73520829
PDE4C	ENSG00000105650	-0.40	0.00	1.00	3.23E-10	19	18319462	18359010
HBEGF	ENSG00000113070	-0.50	0.00	1.00	3.38E-10	5	139712428	139726188
KIAA1683	ENSG00000130518	-0.19	0.00	1.00	3.52E-10	19	18367908	18385310
HPR	ENSG00000261701	-0.62	0.02	0.98	3.79E-10	16	72097047	72111145
NKAIN2	ENSG00000188580	24.44	0.00	1.00	5.73E-10	6	124125010	125146786
RNF39	ENSG00000204618	-0.53	0.66	0.34	5.95E-10	6	30038043	30043626
SRPK2	ENSG00000135250	-0.05	1.00	0.00	6.18E-10	7	104751151	105039755
SFXN5	ENSG00000144040	0.07	1.00	0.00	6.30E-10	2	73169165	73302747
MLL5	ENSG00000005483	0.25	1.00	0.00	6.40E-10	7	104581390	104755466
RBL2	ENSG00000103479	-0.11	0.00	1.00	6.89E-10	16	53467889	53525560
PFDN1	ENSG00000113068	0.16	0.76	0.24	7.66E-10	5	139624620	139682698
HIST1H2BJ	ENSG00000124635	0.04	1.00	0.00	7.95E-10	6	27093676	27100574
SPPL2C	ENSG00000185294	-0.82	0.41	0.59	8.29E-10	17	43922247	43924433
CDH8	ENSG00000150394	0.37	0.67	0.33	8.67E-10	16	61681146	62070939
FAM109B	ENSG00000177096	-0.13	1.00	0.00	1.03E-09	22	42470252	42475442
PTPRO	ENSG00000151490	-0.19	0.00	1.00	1.08E-09	12	15475191	15755109
SLC39A8	ENSG00000138821	0.18	0.00	1.00	1.35E-09	4	103172237	103352415
TNFRSF13C	ENSG00000159958	-0.01	0.00	1.00	1.37E-09	22	42318036	42322810
CPXM2	ENSG00000121898	-0.87	0.30	0.70	1.49E-09	10	125465723	125699783
PLCL1	ENSG00000115896	0.45	0.00	1.00	1.60E-09	2	198669317	199437305
NCOA2	ENSG00000140396	0.32	0.09	0.91	2.03E-09	8	71022017	71316043
PKD2L1	ENSG00000107593	0.26	0.70	0.30	2.28E-09	10	102047906	102090021
CDKAL1	ENSG00000145996	0.63	0.00	1.00	2.31E-09	6	20534688	21232635
SORT1	ENSG00000134243	-0.10	0.72	0.28	2.36E-09	1	109852190	109940540
TANK	ENSG00000136560	-0.43	0.69	0.31	2.37E-09	2	161993419	162092741
SUOX	ENSG00000139531	-0.08	0.57	0.43	2.38E-09	12	56390964	56400425
MLL2	ENSG00000167548	0.23	0.44	0.56	3.06E-09	12	49412758	49454577
LSM4	ENSG00000130520	0.43	0.55	0.45	4.06E-09	19	18417046	18433922
TIMM17A	ENSG00000134375	0.38	0.14	0.86	4.99E-09	1	201924631	201939792
RHEBL1	ENSG00000167550	-1.80	0.14	0.86	5.03E-09	12	49458459	49463808
CALN1	ENSG00000183166	-0.67	0.43	0.57	5.38E-09	7	71244476	71912136
PDE4D	ENSG00000113448	-0.19	1.00	0.00	6.24E-09	5	58264865	59817947
DDN	ENSG00000181418	-0.28	1.00	0.00	6.81E-09	12	49388932	49393158
CWF19L1	ENSG00000095485	0.07	0.00	1.00	7.48E-09	10	101992055	102027437
NKX2-1	ENSG00000136352	-0.42	0.56	0.44	8.23E-09	14	36985597	36990354

GDF15	ENSG00000130513	0.10	0.00	1.00	9.47E-09	19	18485541	18499986
NFIX	ENSG00000008441	0.24	0.03	0.97	1.02E-08	19	13106289	13209610
SNX29	ENSG00000048471	0.16	0.00	1.00	1.03E-08	16	12070591	12668144
AUTS2	ENSG00000158321	-0.25	0.00	1.00	1.16E-08	7	69063282	70258492
CHUK	ENSG00000213341	0.34	0.00	1.00	1.21E-08	10	101948057	101989353
RALYL	ENSG00000184672	-0.34	0.00	1.00	1.66E-08	8	85095022	85834079
PSMA5	ENSG00000143106	0.00	0.50	0.50	1.87E-08	1	109941664	109969070
PRKAR2B	ENSG00000005249	0.41	0.45	0.55	1.93E-08	7	106685150	106802256
EPS8	ENSG00000151491	-1.55	0.33	0.67	2.10E-08	12	15773068	16035263
LYL1	ENSG00000104903	0.25	0.00	1.00	2.25E-08	19	13209847	13213975
PSMC3	ENSG00000165916	-2.36	0.04	0.96	2.29E-08	11	47440320	47448024
LRRC14	ENSG00000160959	0.11	0.00	1.00	2.34E-08	8	145743376	145750556
WNT10B	ENSG00000169884	0.42	0.00	1.00	2.49E-08	12	49359123	49365518
RUNX1T1	ENSG00000079102	0.42	0.94	0.06	2.51E-08	8	92967195	93115514
PET112	ENSG00000059691	-0.05	0.90	0.10	3.00E-08	4	152591656	152682159
NKX2-8	ENSG00000136327	-0.20	0.00	1.00	3.10E-08	14	37049209	37051819
GLYCTK	ENSG00000168237	-0.13	0.67	0.33	3.20E-08	3	52321105	52329273
CAMK2N1	ENSG00000162545	0.18	0.59	0.41	3.50E-08	1	20808884	20812703
CKB	ENSG00000166165	0.10	0.12	0.88	3.68E-08	14	103986004	103989170
MYLK	ENSG00000065534	0.47	0.47	0.53	3.71E-08	3	123328896	123603179
CALR	ENSG00000179218	0.76	0.00	1.00	3.80E-08	19	13049392	13055303
JMJD1C	ENSG00000171988	0.30	0.31	0.69	4.49E-08	10	64926981	65281610
LRRC25	ENSG00000175489	0.13	1.00	0.00	4.61E-08	19	18501947	18508432
PPP1R16A	ENSG00000160972	0.31	0.45	0.55	4.66E-08	8	145703352	145727504
KCNJ6	ENSG00000157542	0.28	0.93	0.07	5.22E-08	21	38979675	39493439
EIF2B5	ENSG00000145191	0.96	0.45	0.55	5.93E-08	3	183852826	183863915
GADD45GIP1	ENSG00000179271	-0.41	1.00	0.00	5.97E-08	19	13063933	13068037
TONSL	ENSG00000160949	0.40	0.48	0.52	6.17E-08	8	145654158	145669823
ELK4	ENSG00000158711	-0.33	1.00	0.00	6.92E-08	1	205566684	205601139
NMNAT2	ENSG00000157064	1.60	0.76	0.24	7.81E-08	1	183217372	183387515
RAD23A	ENSG00000179262	0.16	1.00	0.00	7.92E-08	19	13056669	13064456
MACROD2	ENSG00000172264	-0.12	1.00	0.00	8.80E-08	20	13976015	16033842
DBN1	ENSG00000113758	0.07	1.00	0.00	9.91E-08	5	176883609	176901402
FAM193A	ENSG00000125386	0.09	0.10	0.90	1.04E-07	4	2538374	2734300
FARSA	ENSG00000179115	0.11	1.00	0.00	1.13E-07	19	13033293	13044851
NRBF2	ENSG00000148572	0.25	0.80	0.20	1.16E-07	10	64893007	64914791
ZSWIM6	ENSG00000130449	-1.09	0.00	1.00	1.18E-07	5	60628085	60841999
CCDC14	ENSG00000175455	-0.13	0.00	1.00	1.28E-07	3	123616152	123680255
CDH13	ENSG00000140945	-0.67	1.00	0.00	1.29E-07	16	82660570	83834245
MFSD4	ENSG00000174514	-0.07	1.00	0.00	1.31E-07	1	205538013	205572046
EPHA5	ENSG00000145242	0.29	0.00	1.00	1.53E-07	4	66185281	66536213
TSHZ3	ENSG00000121297	-0.79	0.55	0.45	1.54E-07	19	31640885	31840342
SLC39A4	ENSG00000147804	-1.75	0.00	1.00	1.65E-07	8	145635126	145642228
DAND5	ENSG00000179284	1.23	0.00	1.00	1.66E-07	19	13075973	13085574
PDCL3	ENSG00000115539	-0.11	1.00	0.00	1.79E-07	2	101179455	101193201
DCAF11	ENSG00000100897	-0.77	0.03	0.97	1.84E-07	14	24583404	24594451
GCDH	ENSG00000105607	0.18	1.00	0.00	1.96E-07	19	13001974	13025021
TMEM180	ENSG00000138111	-0.09	1.00	0.00	2.03E-07	10	104221152	104236802
AGAP1	ENSG00000157985	-0.62	0.33	0.67	2.35E-07	2	236402687	237040444
KLF1	ENSG00000105610	-1.72	0.18	0.82	2.49E-07	19	12995236	12998015
CDH4	ENSG00000179242	0.36	0.51	0.49	2.63E-07	20	59827317	60515673
CFB	ENSG00000243649	-1.58	0.93	0.07	2.64E-07	6	31913427	31919861
MAML2	ENSG00000184384	0.24	0.00	1.00	2.71E-07	11	95709762	96076359
IL34	ENSG00000157368	0.44	0.86	0.14	2.75E-07	16	70613798	70694585
CPEB1	ENSG00000214575	0.12	0.00	1.00	2.90E-07	15	83211951	83317612
MTSS1L	ENSG00000132613	0.48	1.00	0.00	2.95E-07	16	70695107	70719956
GRK6	ENSG00000198055	-0.23	0.67	0.33	3.02E-07	5	176830205	176869902
IL17D	ENSG00000172458	0.41	0.50	0.50	3.23E-07	13	21276266	21297237
FMNL3	ENSG00000161791	-0.26	0.44	0.56	3.28E-07	12	50030282	50101948
RORA	ENSG00000069667	0.25	0.01	0.99	3.44E-07	15	60780483	61521501
FBXL17	ENSG00000145743	0.37	0.77	0.23	3.45E-07	5	107194736	107717799
NCAM1	ENSG00000149294	-0.89	0.72	0.28	3.58E-07	11	112831969	113149158

CSRNP3	ENSG00000178662	0.34	0.53	0.47	3.60E-07	2	166326157	166545917
SSBP2	ENSG00000145687	0.42	0.63	0.37	3.82E-07	5	80708623	81047616
CPNE6	ENSG00000100884	0.26	1.00	0.00	4.05E-07	14	24540046	24547309
FITM1	ENSG00000139914	-0.42	0.00	1.00	4.07E-07	14	24599868	24602058
SEMA3E	ENSG00000170381	0.20	0.16	0.84	4.08E-07	7	82992554	83278455
NDUFAF2	ENSG00000164182	-0.17	1.00	0.00	4.27E-07	5	60241004	60450358
LDB2	ENSG00000169744	0.23	0.00	1.00	4.45E-07	4	16503164	16900301
MGST2	ENSG00000085871	0.07	0.76	0.24	4.92E-07	4	140586922	140661899
RNF4	ENSG00000063978	0.05	0.29	0.71	5.04E-07	4	2463947	2627047
RAI1	ENSG00000108557	-0.08	0.00	1.00	5.43E-07	17	17584772	17714767
IGSF9B	ENSG00000080854	0.26	1.00	0.00	5.50E-07	11	133766333	133826863
ANKS1B	ENSG00000185046	-0.26	0.70	0.30	5.51E-07	12	99120235	100378714
C6orf108	ENSG00000112667	-0.05	0.75	0.25	5.66E-07	6	43193367	43197219
C6orf25	ENSG00000204420	-0.18	0.20	0.80	5.80E-07	6	31686371	31694491
CDH9	ENSG00000113100	-0.28	1.00	0.00	6.16E-07	5	26880706	27121257
TNRC6A	ENSG00000090905	-0.14	1.00	0.00	6.19E-07	16	24621530	24838953
CRAT	ENSG00000095321	-0.03	0.00	1.00	6.32E-07	9	131856421	131873468
C18orf1	ENSG00000168675	0.69	1.00	0.00	6.32E-07	18	13217497	13652754
C17orf59	ENSG00000196544	-0.35	0.55	0.45	6.45E-07	17	8091663	8093498
SLC34A1	ENSG00000131183	-0.33	1.00	0.00	6.89E-07	5	176806236	176825849
CEP192	ENSG00000101639	0.03	0.10	0.90	6.97E-07	18	12991361	13125051
FBXL4	ENSG00000112234	-0.11	1.00	0.00	7.02E-07	6	99316411	99395882
NFAM1	ENSG00000235568	0.17	0.00	1.00	7.09E-07	22	42776413	42828409
C20orf173	ENSG00000125975	-0.19	0.70	0.30	7.20E-07	20	34111014	34117481
PRR7	ENSG00000131188	0.51	0.00	1.00	7.29E-07	5	176873446	176883287
SCN2A	ENSG00000136531	0.14	0.00	1.00	8.00E-07	2	166051503	166248820
SP2	ENSG00000167182	0.16	0.00	1.00	8.10E-07	17	45973516	46006323
CPD	ENSG00000108582	-0.08	0.55	0.45	8.37E-07	17	28705945	28797007
CRIP3	ENSG00000146215	0.31	1.00	0.00	8.49E-07	6	43267448	43276564
SHISA9	ENSG00000237515	-0.53	0.28	0.72	8.58E-07	16	12995455	13334273
SEMA6D	ENSG00000137872	-0.76	1.00	0.00	8.84E-07	15	47476298	48066425
XXbac-BPG32J3.19	ENSG00000250641	0.27	0.71	0.29	8.98E-07	6	31674681	31685695
MAML3	ENSG00000196782	0.15	0.88	0.12	9.55E-07	4	140637907	141075338
SMG7	ENSG00000116698	0.00	0.50	0.50	9.78E-07	1	183441351	183567381
SLC22A7	ENSG00000137204	-0.21	0.00	1.00	1.03E-06	6	43263432	43273276
CADM2	ENSG00000175161	-0.11	0.82	0.18	1.04E-06	3	85008140	86123579
PSME2	ENSG00000100911	-0.72	0.04	0.96	1.04E-06	14	24612571	24616779
GPD2	ENSG00000115159	0.20	1.00	0.00	1.13E-06	2	157291802	157470247
PTPRT	ENSG00000196090	-0.66	0.22	0.78	1.16E-06	20	40701392	41818610
AKAP6	ENSG00000151320	0.00	0.50	0.50	1.17E-06	14	32798504	33306890
MTMR2	ENSG00000087053	1.64	0.00	1.00	1.18E-06	11	95554930	95658479
C19orf81	ENSG00000235034	-0.61	0.62	0.38	1.22E-06	19	51152702	51162567
RMI1	ENSG00000178966	-0.28	0.63	0.37	1.28E-06	9	86595713	86618989
CALML5	ENSG00000178372	0.37	0.01	0.99	1.28E-06	10	5540660	5541533
ATP5H	ENSG00000167863	6.63	0.02	0.98	1.28E-06	17	73034958	73043080
TMBIM6	ENSG00000139644	0.05	1.00	0.00	1.32E-06	12	50101508	50158717
RGSL1	ENSG00000121446	-0.14	1.00	0.00	1.34E-06	1	182378327	182529734
AVL9	ENSG00000105778	-0.08	0.00	1.00	1.38E-06	7	32535038	32628338
TYW5	ENSG00000162971	0.06	0.42	0.58	1.38E-06	2	200793636	200820459
PSME1	ENSG00000092010	-1.04	0.00	1.00	1.49E-06	14	24605372	24608176
KCNK3	ENSG00000171303	1.06	0.36	0.64	1.51E-06	2	26915590	26956288
HSPA1A	ENSG00000204389	-0.48	0.06	0.94	1.65E-06	6	31783320	31785723
LMF1	ENSG00000103227	-0.04	0.93	0.07	1.71E-06	16	903634	1031318
MCRS1	ENSG00000187778	-0.04	1.00	0.00	1.76E-06	12	49950327	49961928
SPAG4	ENSG00000061656	-0.12	0.25	0.75	1.82E-06	20	34203751	34209016
SLC5A11	ENSG00000158865	0.18	1.00	0.00	1.94E-06	16	24857162	24922949
EXT1	ENSG00000182197	-0.09	0.00	1.00	1.94E-06	8	118806729	119124065
MAST4	ENSG00000069020	0.11	0.97	0.03	1.94E-06	5	65892208	66465421
ROMO1	ENSG00000125995	-0.34	0.37	0.63	1.97E-06	20	34287194	34288906
TMEM170B	ENSG00000205269	-0.06	0.81	0.19	1.98E-06	6	11537982	11583757
RPS17L	ENSG00000182774	0.23	0.83	0.17	2.00E-06	15	83205501	83209210
LRRC9	ENSG00000131951	-0.14	0.00	1.00	2.01E-06	14	60386431	60530277

ZNF638	ENSG00000075292	-0.11	0.00	1.00	2.04E-06	2	71503691	71662199
C9orf64	ENSG00000165118	-0.07	1.00	0.00	2.05E-06	9	86553226	86571901
SH3RF3	ENSG00000172985	0.08	0.83	0.17	2.09E-06	2	109745661	110262211
BCL11A	ENSG00000119866	-0.28	0.58	0.42	2.11E-06	2	60677655	60781602
MACROD1	ENSG00000133315	0.21	1.00	0.00	2.14E-06	11	63766030	63933585
PURG	ENSG00000172733	0.15	0.14	0.86	2.15E-06	8	30853318	30891231
RP11-468E2.4	ENSG00000259529	-2.84	0.00	1.00	2.15E-06	14	24616757	24635661
CUL9	ENSG00000112659	0.20	0.29	0.71	2.21E-06	6	43149922	43192325
LARGE	ENSG00000133424	-0.18	0.92	0.08	2.21E-06	22	33558212	34318829
VEGFA	ENSG00000112715	0.18	0.12	0.88	2.21E-06	6	43737921	43754224
PHF20	ENSG00000025293	-0.10	0.99	0.01	2.25E-06	20	34359896	34538292
SETBP1	ENSG00000152217	0.25	0.89	0.11	2.25E-06	18	42260138	42648475
C2orf47	ENSG00000162972	0.31	0.48	0.52	2.29E-06	2	200820040	200873263
KBTBD2	ENSG00000170852	-0.90	0.00	1.00	2.48E-06	7	32907784	32933743
CNGB3	ENSG00000170289	3.10	1.00	0.00	2.56E-06	8	87566205	87755903
PCK2	ENSG00000100889	0.35	0.00	1.00	2.56E-06	14	24563262	24579807
DCAF5	ENSG00000139990	0.16	1.00	0.00	2.64E-06	14	69517598	69619867
PRPF38A	ENSG00000134748	0.20	0.17	0.83	2.67E-06	1	52870274	52886508
SREBF1	ENSG00000072310	-0.35	0.41	0.59	2.67E-06	17	17713713	17740316
NCKAP5L	ENSG00000167566	0.40	0.75	0.25	2.67E-06	12	50184929	50222533
FAM76B	ENSG00000077458	0.10	1.00	0.00	2.72E-06	11	95502117	95523573
MED27	ENSG00000160563	0.20	1.00	0.00	2.74E-06	9	134728315	134955254
CPNE3	ENSG00000085719	0.21	0.63	0.37	2.74E-06	8	87526664	87573726
SYT3	ENSG00000213023	-0.58	0.27	0.73	2.82E-06	19	51124564	51143138
DCDC2	ENSG00000146038	0.12	1.00	0.00	2.84E-06	6	24171983	24358287
CHCHD3	ENSG00000106554	-0.11	0.31	0.69	2.89E-06	7	132469631	132766850

Abbreviations: ENSG, Ensembl gene ID; alpha, overall effect of the GreX; w1, contribution weight for African ancestry; w2, contribution weight for European ancestry; chr, chromosome.

Table S4-5. Genes associated with white matter hyperintensity using METRO followed by fine-mapping with FOCUS (N=23 genes; $P<2.9x10^{-6}$)

Gene	ENSG	alpha	w1	w2	P value	chr	Start	End
WBP2	ENSG00000132471	1.52	0.00	1.00	7.92E-54	17	73841780	73852588
SH3PXD2A	ENSG00000107957	0.00	0.50	0.50	5.78E-20	10	105353784	105615342
DCAKD	ENSG00000172992	3.93	0.00	1.00	3.13E-19	17	43100706	43138499
NMT1	ENSG00000136448	-0.44	0.00	1.00	1.41E-16	17	43035360	43186384
EFEMP1	ENSG00000115380	0.89	1.00	0.00	1.35E-15	2	56093102	56151274
ICA1L	ENSG00000163596	-0.27	0.36	0.64	5.44E-12	2	203637873	203736489
NBEAL1	ENSG00000144426	0.57	0.00	1.00	4.91E-11	2	203879331	204091101
WDR12	ENSG00000138442	-0.64	0.00	1.00	5.73E-11	2	203738984	203879521
KLHL24	ENSG00000114796	-0.99	0.00	1.00	1.01E-10	3	183353398	183402307
NEURL	ENSG00000107954	0.93	0.44	0.56	3.28E-10	10	105253462	105352303
CALCRL	ENSG00000064989	1.08	0.94	0.06	7.96E-10	2	188206691	188313187
HAAO	ENSG00000162882	-0.24	1.00	0.00	1.18E-08	2	42994229	43019733
ARMS2	ENSG00000254636	-0.62	0.00	1.00	1.64E-08	10	124214169	124216868
GJC1	ENSG00000182963	-0.06	0.00	1.00	3.57E-08	17	42875816	42908184
OXER1	ENSG00000162881	0.53	1.00	0.00	7.02E-08	2	42989639	42991275
HTRA1	ENSG00000166033	7.07	1.00	0.00	2.53E-07	10	124218067	124274423
LRRC37A3	ENSG00000176809	-0.44	0.00	1.00	2.54E-07	17	62850248	62915598
FBXO31	ENSG00000103264	1.11	0.00	1.00	3.05E-07	16	87360593	87425748
PDCD7	ENSG00000090470	-1.81	0.00	1.00	3.54E-07	15	65409717	65426146
CLPX	ENSG00000166855	0.40	1.00	0.00	3.54E-07	15	65440557	65477680
EFTUD2	ENSG00000108883	0.00	0.50	0.50	4.49E-07	17	42927316	42976813
UBAP1L	ENSG00000246922	-1.26	0.01	0.99	5.91E-07	15	65385098	65407538
MAP1LC3B	ENSG00000140941	0.37	0.13	0.87	1.94E-06	16	87417559	87438385

Abbreviations: ENSG, Ensembl gene ID; alpha, overall effect of the GreX; w1, contribution weight for African ancestry; w2, contribution weight for European ancestry; chr, chromosome.

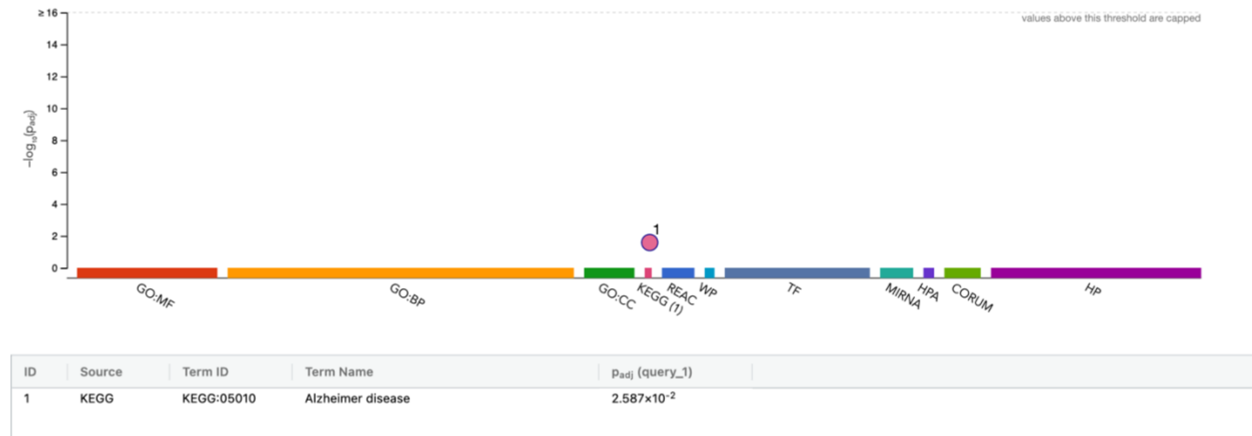
Table 4-6. Genes associated with Alzheimer's disease using METRO followed by fine-mapping with FOCUS (N=69 genes; $P<2.9x10^{-6}$)

Gene	ENSG	alpha	w1	w2	P value	chr	Start	End
CLU	ENSG00000120885	0.48	0.29	0.71	4.77E-39	8	27454434	27472217
TOMM40	ENSG00000130204	-19.78	1.00	0.00	4.77E-39	19	45393826	45406946
APOE	ENSG00000130203	67.11	0.28	0.72	4.77E-39	19	45409048	45412650
APOC4	ENSG00000224916	-17.64	0.91	0.09	4.77E-39	19	45445495	45452822
PICALM	ENSG00000073921	-0.15	1.00	0.00	9.33E-36	11	85668218	85780924
CCDC83	ENSG00000150676	0.28	1.00	0.00	1.01E-26	11	85566144	85631064
CR2	ENSG00000117322	1.26	0.00	1.00	1.87E-25	1	207627575	207663240
MS4A6A	ENSG00000110077	-0.29	0.58	0.42	1.48E-19	11	59939488	59952139
MS4A2	ENSG00000149534	1.51	0.22	0.78	2.29E-17	11	59855734	59865940
RIN3	ENSG00000100599	3.19	1.00	0.00	7.51E-17	14	92980125	93155339
MEPCE	ENSG00000146834	-0.22	1.00	0.00	5.62E-16	7	100025945	100031749
PPP1R35	ENSG00000160813	0.31	1.00	0.00	2.40E-15	7	100032905	100034120
CYB561	ENSG00000008283	0.12	0.85	0.15	1.66E-14	17	61509665	61523715
KCNH6	ENSG00000173826	-0.27	0.62	0.38	2.33E-14	17	61600695	61626338
USP6	ENSG00000129204	-0.38	0.98	0.02	3.47E-14	17	5019327	5078329
TREM2	ENSG00000095970	1.10	0.00	1.00	7.75E-14	6	41126244	41130924
ACE	ENSG00000159640	0.65	0.24	0.76	8.53E-14	17	61554422	61575741
ZYX	ENSG00000159840	0.17	0.00	1.00	9.46E-14	7	143078388	143088204
FAM131B	ENSG00000159784	-0.28	1.00	0.00	2.78E-13	7	143050493	143059863
C6orf10	ENSG00000204296	-0.39	0.00	1.00	3.04E-13	6	32256303	32339689
BTNL2	ENSG00000204290	0.60	0.76	0.24	7.66E-13	6	32361116	32374958
ATG16L1	ENSG00000085978	-0.12	0.00	1.00	1.33E-12	2	234118697	234204320
TREML1	ENSG00000161911	-1.07	0.96	0.04	2.31E-12	6	41117075	41122085
ZNF594	ENSG00000180626	0.98	0.40	0.60	2.44E-12	17	5082830	5095163
EPHA1	ENSG00000146904	-0.58	0.52	0.48	3.79E-12	7	143087382	143105949
APP	ENSG00000142192	-0.13	0.00	1.00	1.09E-11	21	27252861	27543446
USP6NL	ENSG00000148429	0.08	0.40	0.60	1.33E-10	10	11502509	11653665
SLC39A13	ENSG00000165915	-0.12	0.00	1.00	2.95E-10	11	47428683	47438047
PPP4C	ENSG00000149923	0.03	0.92	0.08	3.65E-10	16	30087299	30096697
SLC52A1	ENSG00000132517	0.78	0.27	0.73	3.69E-10	17	4935895	4955304
AC008394.1	ENSG00000233828	0.17	0.00	1.00	5.07E-10	5	86512423	86534822
FAM210B	ENSG00000124098	-0.14	0.00	1.00	9.14E-10	20	54934030	54943719
GCNT7	ENSG00000124091	-3.29	0.98	0.02	9.16E-10	20	55066548	55100981
RAPSN	ENSG00000165917	0.14	0.99	0.01	1.37E-09	11	47459315	47470695
CD55	ENSG00000196352	-0.47	1.00	0.00	1.84E-09	1	207494864	207560149
HLA-DQA2	ENSG00000237541	-0.06	0.00	1.00	1.87E-09	6	32709168	32714975
TBX6	ENSG00000149922	-0.11	0.96	0.04	1.95E-09	16	30097114	30103245
DYDC2	ENSG00000133665	-0.44	0.33	0.67	2.42E-09	10	82104501	82127829
CSTF1	ENSG00000101138	-4.15	0.00	1.00	2.62E-09	20	54967427	54981418
CASS4	ENSG00000087589	-0.18	0.37	0.63	4.41E-09	20	54987168	55035443
DYDC1	ENSG00000170788	0.30	0.39	0.61	4.65E-09	10	82095861	82116511
LILRA6	ENSG00000244482	0.21	0.13	0.87	5.48E-09	19	54720737	54746649
MME	ENSG00000196549	-0.23	1.00	0.00	5.99E-09	3	154741913	154901493
NIT1	ENSG00000158793	-0.24	0.32	0.68	7.12E-09	1	161087876	161095235
KLF16	ENSG00000129911	0.18	1.00	0.00	1.03E-08	19	1852398	1863578
DEDD	ENSG00000158796	0.15	0.00	1.00	1.17E-08	1	161090764	161102478
MAF1	ENSG00000179632	-0.11	0.00	1.00	1.52E-08	8	145159364	145162514
TP53INP1	ENSG00000164938	-0.05	0.66	0.34	2.05E-08	8	95938200	95961606
KANSL1	ENSG00000120071	0.00	0.50	0.50	3.81E-08	17	44107282	44302755
LACTB	ENSG00000103642	0.06	0.41	0.59	4.42E-08	15	63414032	63434260
SHARPIN	ENSG00000179526	0.23	1.00	0.00	9.56E-08	8	145153536	145163027

MAPT	ENSG00000186868	-0.54	0.51	0.49	1.76E-07	17	43971893	44105700
IKZF1	ENSG00000185811	-0.13	1.00	0.00	3.33E-07	7	50343664	50472799
CPSF3	ENSG00000119203	0.19	1.00	0.00	3.85E-07	2	9563780	9613230
CCNE2	ENSG00000175305	-0.18	0.35	0.65	4.65E-07	8	95891998	95908906
FAM108A1	ENSG00000129968	-0.81	0.68	0.32	4.69E-07	19	1876809	1885495
GRN	ENSG00000030582	0.18	0.84	0.16	6.31E-07	17	42422614	42430474
ABI3	ENSG00000108798	-0.05	0.00	1.00	9.18E-07	17	47287773	47300587
SLTM	ENSG00000137776	0.17	1.00	0.00	9.82E-07	15	59171249	59225878
RNF111	ENSG00000157450	-0.39	0.06	0.94	1.04E-06	15	59157374	59389618
EPDR1	ENSG00000086289	0.04	1.00	0.00	1.16E-06	7	37723446	37991538
SUPT4H1	ENSG00000213246	0.06	0.76	0.24	1.27E-06	17	56422536	56430454
STYX	ENSG00000198252	0.10	0.00	1.00	1.38E-06	14	53196884	53241707
SIGLEC9	ENSG00000129450	-0.17	0.00	1.00	1.82E-06	19	51628163	51639908
DDX54	ENSG00000123064	0.22	0.36	0.64	2.02E-06	12	113594978	113623284
TRIB1	ENSG00000173334	0.42	0.00	1.00	2.41E-06	8	126442600	126450645
BLNK	ENSG00000095585	0.40	0.00	1.00	2.55E-06	10	97948927	98031344
SLC25A39	ENSG00000013306	-0.58	1.00	0.00	2.57E-06	17	42396993	42402238
ICA1L	ENSG00000163596	-0.06	0.41	0.59	2.68E-06	2	203637873	203736489

Abbreviations: ENSG, Ensembl gene ID; alpha, overall effect of the GreX; w1, contribution weight for African ancestry; w2, contribution weight for European ancestry; chr, chromosome.

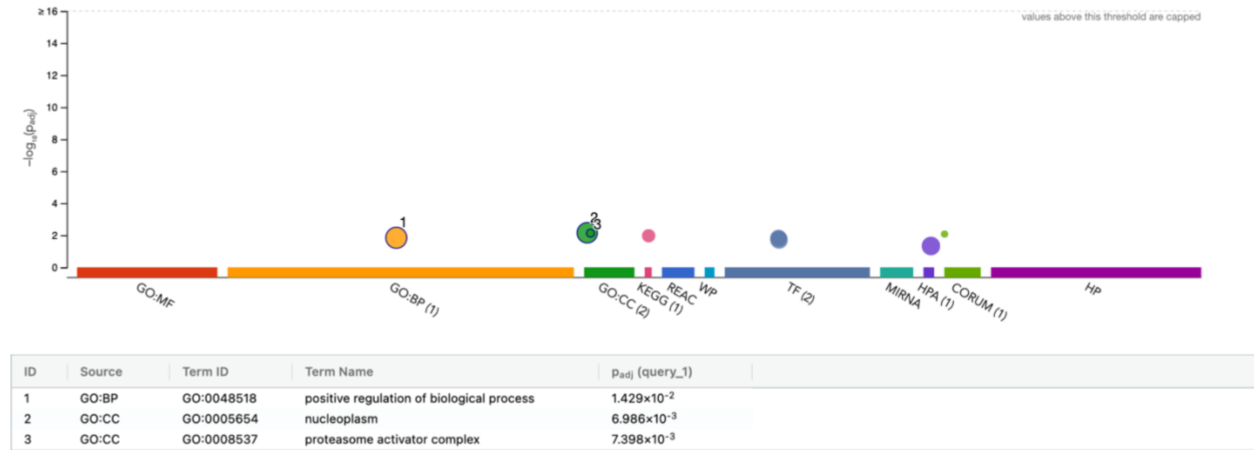
Figure 4-12. Functional enrichment analysis on the gene set identified by METRO for general cognitive function and AD (N=22 genes; P<2.90x10^{-6}).



ID	Source	Term ID	Term Name	p_{adj} (query_1)
1	KEGG	KEGG:05010	Alzheimer disease	2.587×10^{-2}

The top panel consists of a Manhattan plot that illustrates the enrichment analysis results. The x-axis represents functional terms that are grouped and color-coded by data sources, including Gene Ontology (GO): molecular function (MF; red), GO: biological process (BP; orange), GO: cellular component (CC; dark green), Kyoto Encyclopedia of Genes and Genomes (KEGG; pink), Reactome (REAC; dark blue), WikiPathways (WP; turquoise), Transfac (TF; light blue), MiRTarBase (MIRNA; emerald green), Human Protein Atlas (HPA; dark purple), CORUM protein complexes (light green), and Human Phenotype Ontology (HP; violet), in order from left to right. The y-axis shows the adjusted enriched $-\log_{10}$ p-values <0.05. Multiple testing correction was performed using g:SCS method (Set Counts and Sizes) that takes into account overlapping terms. The top panel highlights driver GO terms identified using the greedy filtering algorithm in g:Profiler. The light circles represent terms that were not significant after filtering. The circle sizes are in accordance with the corresponding term size (i.e., larger terms have larger circles). The number in parentheses following the source name in the x-axis shows how many significantly enriched terms were from this source.

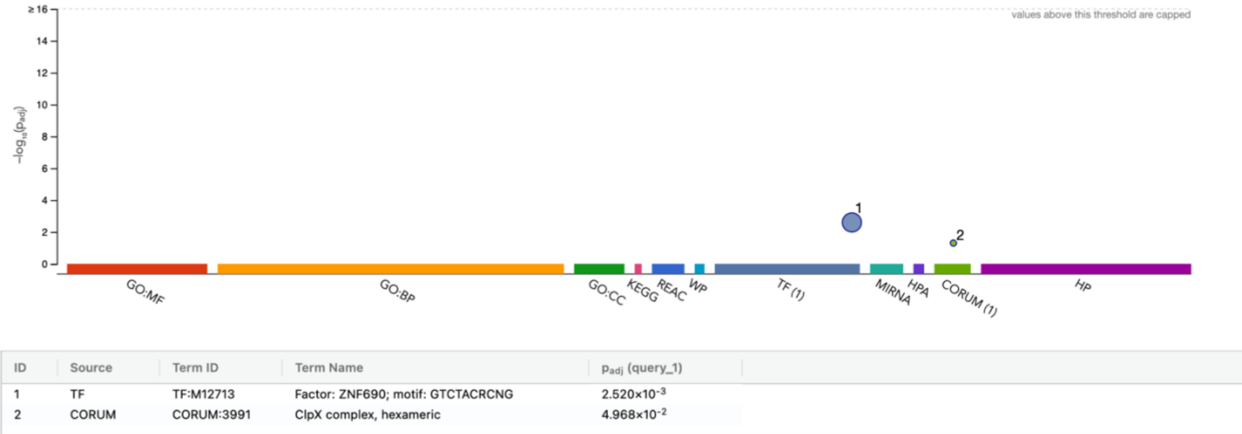
Figure 4-13. Functional enrichment analysis on the fine-mapped gene set not previously identified by Davies et al. (2018)[15] for general cognitive function using METRO TWAS (N=82 genes).



ID	Source	Term ID	Term Name	p_{adj} (query_1)
1	GO:BP	GO:0048518	positive regulation of biological process	1.429×10^{-2}
2	GO:CC	GO:0005654	nucleoplasm	6.986×10^{-3}
3	GO:CC	GO:0008537	proteasome activator complex	7.398×10^{-3}

The top panel consists of a Manhattan plot that illustrates the enrichment analysis results. The x-axis represents functional terms that are grouped and color-coded by data sources, including Gene Ontology (GO): molecular function (MF; red), GO: biological process (BP; orange), GO: cellular component (CC; dark green), Kyoto Encyclopedia of Genes and Genomes (KEGG; pink), Reactome (REAC; dark blue), WikiPathways (WP; turquoise), Transfac (TF; light blue), MiRTarBase (MIRNA; emerald green), Human Protein Atlas (HPA; dark purple), CORUM protein complexes (light green), and Human Phenotype Ontology (HP; violet), in order from left to right. The y-axis shows the adjusted enriched $-\log_{10}$ p-values <0.05. Multiple testing correction was performed using g:SCS method (Set Counts and Sizes) that takes into account overlapping terms. The top panel highlights driver GO terms identified using the greedy filtering algorithm in g:Profiler. The light circles represent terms that were not significant after filtering. The circle sizes are in accordance with the corresponding term size (i.e., larger terms have larger circles). The number in parentheses following the source name in the x-axis shows how many significantly enriched terms were from this source.

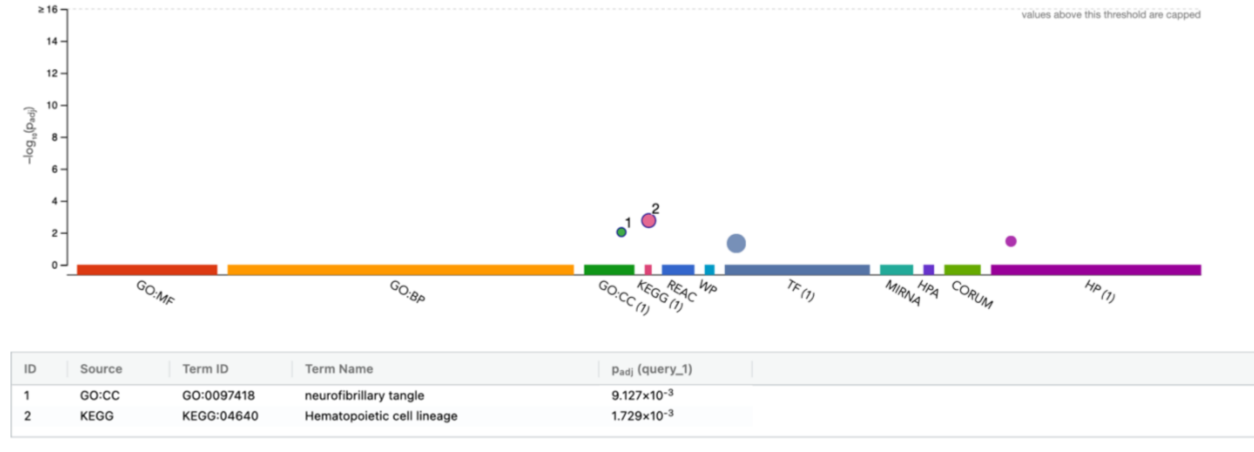
Figure 4-14. Functional enrichment analysis on the fine-mapped gene set not previously identified by Sargurupremraj et al. (2020)[305] for white matter hyperintensity using METRO TWAS (N=12 genes).



ID	Source	Term ID	Term Name	p_{adj} (query_1)
1	TF	TF:M12713	Factor: ZNF690; motif: GTCTACRCNG	2.520×10^{-3}
2	CORUM	CORUM:3991	ClpX complex, hexameric	4.968×10^{-2}

The top panel consists of a Manhattan plot that illustrates the enrichment analysis results. The x-axis represents functional terms that are grouped and color-coded by data sources, including Gene Ontology (GO): molecular function (MF; red), GO: biological process (BP; orange), GO: cellular component (CC; dark green), Kyoto Encyclopedia of Genes and Genomes (KEGG; pink), Reactome (REAC; dark blue), WikiPathways (WP; turquoise), Transfac (TF; light blue), MiRTarBase (MIRNA; emerald green), Human Protein Atlas (HPA; dark purple), CORUM protein complexes (light green), and Human Phenotype Ontology (HP; violet), in order from left to right. The y-axis shows the adjusted enriched $-\log_{10}$ p-values <0.05. Multiple testing correction was performed using g:SCS method (Set Counts and Sizes) that takes into account overlapping terms. The top panel highlights driver GO terms identified using the greedy filtering algorithm in g:Profiler. The light circles represent terms that were not significant after filtering. The circle sizes are in accordance with the corresponding term size (i.e., larger terms have larger circles). The number in parentheses following the source name in the x-axis shows how many significantly enriched terms were from this source.

Figure 4-15. Functional enrichment analysis on the fine-mapped gene set not previously identified by Bellenguez et al. (2022)[112] for Alzheimer's disease using METRO TWAS (N=45 genes).



ID	Source	Term ID	Term Name	p_{adj} (query_1)
1	GO:CC	GO:0097418	neurofibrillary tangle	9.127×10^{-3}
2	KEGG	KEGG:04640	Hematopoietic cell lineage	1.729×10^{-3}

The top panel consists of a Manhattan plot that illustrates the enrichment analysis results. The x-axis represents functional terms that are grouped and color-coded by data sources, including Gene Ontology (GO): molecular function (MF; red), GO: biological process (BP; orange), GO: cellular component (CC; dark green), Kyoto Encyclopedia of Genes and Genomes (KEGG; pink), Reactome (REAC; dark blue), WikiPathways (WP; turquoise), Transfac (TF; light blue), MiRTarBase (MIRNA; emerald green), Human Protein Atlas (HPA; dark purple), CORUM protein complexes (light green), and Human Phenotype Ontology (HP; violet), in order from left to right. The y-axis shows the adjusted enriched $-\log_{10}$ p-values <0.05. Multiple testing correction was performed using g:SCS method (Set Counts and Sizes) that takes into account overlapping terms. The top panel highlights driver GO terms identified using the greedy filtering algorithm in g:Profiler. The light circles represent terms that were not significant after filtering. The circle sizes are in accordance with the corresponding term size (i.e., larger terms have larger circles). The number in parentheses following the source name in the x-axis shows how many significantly enriched terms were from this source.

References

1. Davies, G., Lam, M., Harris, S.E., Trampush, J.W., Luciano, M., Hill, W.D., Hagenaars, S.P., Ritchie, S.J., Marioni, R.E., Fawns-Ritchie, C., et al. (2018). Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. Nat Commun *9*, 2098. 10.1038/s41467-018-04362-x.

2. Sargurupremraj, M., Suzuki, H., Jian, X., Sarnowski, C., Evans, T.E., Bis, J.C., Eiriksdottir, G., Sakaue, S., Terzikhan, N., Habes, M., et al. (2020). Cerebral small vessel disease genomics and its implications across the lifespan. Nature Communications *11*, 6285. 10.1038/s41467-020-19111-2.

3. Bellenguez, C., Küçükali, F., Jansen, I.E., Kleineidam, L., Moreno-Grau, S., Amin, N., Naj, A.C., Campos-Martin, R., Grenier-Boley, B., Andrade, V., et al. (2022). New insights into the genetic etiology of Alzheimer′s disease and related dementias. Nat Genet, 1–25. 10.1038/s41588-022-01024-z.

Chapter 5 . Conclusion

5.1 Summary and Implications of Main Findings

The objectives of this dissertation were to: (1) investigate whether previously identified SNPs, epigenetic variants and/or their interactions in the *ABCA7* region were associated with general cognitive function in AA participants without preliminary evidence of dementia, using linear mixed models; (2) assess whether DNA methylation from peripheral blood leucocytes mediates the relationships between neighborhood factors and cognitive function/WMH outcomes in cognitively healthy AA, using high-dimensional mediation methods; and (3) conduct TWAS that leverage gene expression data collected from EA and AA, through a joint likelihood-based inference framework, to identify genes associated with general cognitive function, WMH and AD. This dissertation uses data from the GENOA study, a large and well-characterized cohort of AA (as well as EA) with rich multi-omic data and neighborhood measures. This body of work advances our knowledge of the relationships between genetic variants, methylation, and transcriptomic mechanisms, as well as their interactions with socio-contextual factors, underlying neurocognitive outcomes and structural brain measures in older adults.

In Chapter 2, we investigated whether previously identified risk SNPs in *ABCA7*, DNA methylation in *ABCA7*, and their interactions were associated with general cognitive function in older AA without dementia. To better understand the functional consequence of these risk factors at the molecular level, we also evaluated whether identified epigenetic or genetic risk factors were associated with transcript level *ABCA7* gene expression in transformed beta lymphocytes

from the same cohort. Although *ABCA7* sentinel SNPs and CpG sites were not associated with general cognitive function, we did see evidence of SNP-by-CpG interactions. We found that two sentinel SNPs in the *ABCA7* region (rs3764647 and rs115550680) may regulate the effects of methylation on cognitive function. In conclusion, while AD risk SNPs in ABCA7 were not associated with cognitive function in this sample, methylation at local CpGs may play an important role on cognitive function, depending on the genotype.

To better understand the functional consequences of these risk factors at the molecular level, we also evaluated whether identified epigenetic or genetic risk factors were associated with transcript level *ABCA7* gene expression in transformed beta-lymphocytes from the same cohort. We found that depending on the allele carried, identified SNPs may influence transcript expression levels that may affect cognitive function. This differential pattern may be due to different functions of the two transcripts instead of alternative splicing. Taken together, these results suggest that SNPs and CpG sites in *ABCA7* may interact to modulate the expression and/or function of *ABCA7* transcripts, and that some of the affected transcripts may influence cognitive function in older AA.

Our study findings are important because they fill in the gap in the current literature on the effect of *ABCA7* risk SNPs and their interplay with DNA methylation on cognitive function in older AA without dementia. While *ABCA7* has previously been implicated in AD in AA, this is the first study to date that has examined this gene in relation to cognitive function in AA. This is also the first study to examine whether SNP-by-CpG interactions, which have been shown to be an important mechanism underlying human complex diseases,[221–223] were associated with cognitive outcomes. By further incorporating transcriptomic data, we investigated whether these SNP-by-CpG interactions may influence cognitive function through alternative splicing or

modulation of expression of specific transcripts in the pathway. In summary, we demonstrated that an intricate interplay between genetic and epigenetic risk factors in the *ABCA7* region may play an important role in cognitive function. Future studies are needed to disentangle this complicated relationship.

In Chapter 3, we examined whether DNA methylation in peripheral blood leukocytes mediates the relationship between neighborhood characteristics and cognitive function or WMH in older AA participants without preliminary evidence of dementia. Greater simple densities of alcohol drinking places were associated with higher delayed recall, and greater densities of fast-food destination and unfavorable food stores with alcohol were associated with higher visual conceptual tracking in cognitively normal AA. However, we detected no significant mediating effects of DNA methylation on the associations between these neighborhood characteristics and cognitive function.

The direction of the total effects is surprising given that greater densities of destinations that may encourage unhealthy dietary choices (e.g., such as alcohol drinking places, fast-food destinations and unfavorable food stores), were associated with higher cognitive measures of delayed recall and visual conceptual tracking. These results may instead be due to increased access to neighborhood community resources and walking destinations that were positively associated with cognitive health through related to improved physical activity levels, social engagement, mental health or quality of life.[280] The plausible mechanisms and direction or presence of neighborhood-cognitive function association may depend on the neighborhood characteristic and cognitive domain being studied, and more than one mechanism may be at play.

To clarify the underlying potential biological mechanisms linking neighborhood factors and cognitive function/WMH, we further investigated whether DNA methylation may mediate

the relationship the pathways between the neighborhood environment and cognitive function/WMH. Previous studies have shown that the neighborhood context affects DNA methylation, even after adjusting for individual level factors, and that DNA methylation patterns in stress and inflammatory pathways may be responsive to interventions.[255] EWAS have also found multiple CpGs related to neurodegeneration associated with cognitive function.[118,257] Considering these factors and that past studies have found epigenetic markers mediating the relationship between neighborhood environment and various cardiovascular risk factors,[260–262] which are potential upstream factors of cognitive function and dementia, we expected to detect mediating CpG sites in the associations between neighborhood characteristics and cognitive function/WMH using Sobel-Comp method in older AA. Our results may indicate that either methylation is not a critical component of the mediating pathway, or that we do not have enough power to investigate CpG sites that may mediate the relationship between the neighborhood environment and cognitive function/WMH.

In Chapter 4, we conducted a multi-ancestry TWAS that leveraged gene expression data collected from EA and AA in GENOA, through a joint likelihood-based inference framework, to identify genes associated with general cognitive function, WMH and AD. We then fine-mapped the identified regions using FOCUS and characterized identified genes. We identified 266, 23, and 69 genes associated with general cognitive function, WMH, and AD, respectively (Bonferroni-corrected alpha-value $P<2.9 \times 10^{-6}$). Among those, METRO identified 82, 12 and 45 genes that were not previously identified by the GWAS studies for general cognitive function, WMH and AD, respectively. Our TWASs indicated overlapping genes associated with innate immunity, vascular dysfunction and neuroinflammation, suggesting that similar mechanisms drive the progression of dementia. The WMH and AD TWASs showed that downregulation of

ICA1L contributed to overlapping AD and VaD neuropathology and may be a target and/or preventative treatment for AD and VaD. To our knowledge, this study was the first TWAS analysis using expression mapping studies in multiple ancestries (EA and AA) to identify genes associated with cognitive function and neurocognitive disorders.

The significance of this study is that by leveraging the complementary information in gene expression prediction models constructed in EA and AA, as well as the uncertainty in SNP prediction weights, we were able to conduct a highly powered TWAS to identify important gene-trait associations and transcriptomic mechanisms underlying AD, VaD and general cognitive function. We also conducted FOCUS fine-mapping to narrow in on a list of putatively causal genes among multiple significant genes in a region. While our study may benefit from the inclusion of summary statistics from an AA GWAS, as well as eQTL data in brain tissue, our study sheds light on gene-trait associations using publicly available EA GWAS summary statistics and eQTL data in peripheral blood leucocytes from EA and AA. In addition, our use of eQTL data in EA and AA allows us to increase the effectiveness of TWAS and improve generalizability of gene-trait findings to non-EA ancestry groups using the largest AA eQTL to date. Also, our use of eQTL in blood in multiple ancestries provides insight into the systemic influences and transcriptomic mechanisms affecting cognitive function and dementia development. The results of this study are important because while there were similar pathways that contribute to healthy cognitive aging and progression of dementia, there were also distinct pathways that were unique to each neuropathology. By understanding overlapping and unique genes and transcriptomic mechanisms associated with each phenotype, we may identify possible targets for prevention and/or treatments for cognitive aging and dementia.

In this dissertation, we utilized a multi-omic approach to investigate how genetic, epigenetic and transcriptomic mechanisms affect and interact to affect the pathology of cognitive function and dementia. Our findings are particularly relevant to AA, an understudied population that has a greater burden of dementia compared to NHW.[21,23–25] This collective work sheds light on the many overlapping and interacting mechanisms that contribute to healthy cognitive aging and neurodegenerative processes of dementia. There seems to be an interplay between cognitive function and dementia that is related to cardiovascular processes, such as diabetes and obesity.[376] For example, there is evidence that hyperglycemia, insulin resistance, and increased tau may interact with amyloid-beta plaques to induce neurodegeneration.[376] In addition, changes in neuroinflammatory regulation processes may also contribute to dementia onset. If neuroinflammation is not initially resolved, chronic inflammation proceeds to initiate neurodegenerative responses in an unregulated, cascading manner.[377] Despite the previously-studied complex factors underlying dementia and cognitive function, these traits still have mysterious aspects that contribute to their uncontrollable processes. As such, multi-omic studies are a promising tool to investigate the global and dynamic molecular changes underlying dementia in the prodromal phase and cognitive decline in healthy individuals. Due to inaccessibility of the human brain, it is crucial to differentially diagnose and study the etiology of dementia and mild cognitive impairment (MCI) prior to onset using biomarkers in the blood, such as genetic and epigenetic variants, and transcriptomic markers. By investigating how these mechanisms interact with each other, as well as socio-contextual factors, we may understand how these factors converge to contribute to the pathogenesis and clinical expression of neurogenerative diseases. This may allow the identification of targets for intervention and treatment, especially in populations that are most at risk.[35]

5.2 Strengths and Limitations

While this dissertation contributes to our understanding of the roles of genetic variants, DNA methylation and transcriptomic mechanisms, and their interplay with socio-contextual factors in cognitive/WMH outcomes, it is not without limitations. First, recruitment in GENOA focused on obtaining a sample enriched with genetic variants related to hypertension among sibships. This may be a source of selection bias and may limit the generalizability of our study to the general population. To account for sibships, the genetic relatedness matrix was modeled as a random effect in all models. There may also have been residual confounding, as well as measurement error in cognitive function. We also used MMSE to exclude participants who may have had dementia. Several studies have found that the MMSE alone cannot be used to predict dementia, and its accuracy in measuring cognitive function could be affected by sociocultural variables, age, education, and other factors. [too] However, since we do not have data on dementia or AD diagnosis, this measure was adequate to exclude participants whose dementia symptoms were more pronounced.

A further limitation is that our gene expression measures were taken from transformed beta-lymphocytes from immortalized cell lines. While these cell lines optimize examination of the functional effects of genetic variation on gene expression due to homogeneity in the cellular environment, the transformation process causes epigenetic and downstream transcriptional changes to the immortalized cells that are not fully understood.[228] Since our DNA methylation was from peripheral blood leukocytes, a different cell type than the transcriptomic data, the inferences from our combined epigenetic and transcriptomic analyses should be verified in a

single cell type. In addition, methylation and transcription patterns may differ between blood and brain tissues;[191,215] However, blood cells touch every cell bed that affects the brain and are involved in chronic inflammation and oxidative stress, which are linked to cognitive performance.[216,217] Further, collecting blood cells is also relatively inexpensive and non-invasive, providing a means for investigating multi-omic relationships with neurocognitive outcomes in large samples of living participants. Thus, our results provide a starting point for multi-omic investigation of neurocognition that need to be further characterized in brain tissue.

This dissertation has several strengths. First, our studies were conducted among AA, where findings may help us better understand cognitive impairment, cognitive decline and dementia in a population that is typically underrepresented in multi-omic research. Additionally, our interpretations of functional consequences were improved through the use of gene- and transcript-level expression data. We also implemented state-of-the-art statistical methods that allowed us to assess high-dimensional DNA methylation pathways linking socio-contextual factors with cognitive function and WMH (Sobel-Comp; Aim 2), as well as leverage gene expression data from multiple ancestries to conduct a well-powered TWAS for cognitive function, WMH and AD (METRO; Aim 3). Lastly, with comprehensive cognitive function measures, we were able to assess multi-omic associations with general cognitive function, individual cognitive domains, and WMH in AA.

5.3 Future Directions

This dissertation sheds light on the importance of assessing how multi-omic layers of information interact with socio-contextual factors to affect cognitive function/WMH and dementia etiology in older adults. Our findings show that there may be an intricate network of

genetic variants, DNA methylation sites, and transcriptomic mechanisms that underly these complex traits and diseases. We were also able to show that the socio-contextual environment, which includes the effects of lifestyle and environmental exposures throughout the life course, may influence cognitive/WMH outcomes; however, their influence may not operate primarily through epigenetics.

While this body of work has important implications that may allow us to develop interventions and/or treatments for cognitive aging prior to dementia onset, replication studies are needed to characterize whether our findings are generalizable to other cohorts of AA, as well as other ancestries. Future studies with multi-ethnic cohorts and longitudinal multi-omic measurements, from early in life to older adulthood, may further elucidate differences in neuropathogenesis between groups and improve our understanding of the contribution of multi-omic and socio-contextual influences on cognitive function and dementia the prodromal period.

Future directions may further incorporate additional layers of “omic” data including proteomics and metabolomics that may further elucidate underlying biological mechanisms and processes identified by our findings. Also, future directions may include the use of data from brain tissue to understand how our findings apply in more relevant tissue to cognitive/WMH and dementia outcomes compared to peripheral blood leucocytes. Further, studies with larger sample sizes are necessary, especially important for epigenetic mediation. In Aim 2, we used a relatively small sample of 542 participants with available DNA methylation and neighborhood measures. It is possible that our lack of findings is due to lower power, and a larger sample could increase the statistical power to identify mediating CpG sites throughout the epigenome. In conclusion, future adequately powered studies with repeated measures would be beneficial to our understanding of the role of multi-omic information in cognitive aging.

5.4 Conclusion

The set of studies in this dissertation are among the first to take a multi-omic approach to examining neurocognitive outcomes in AA. Our thorough investigation of the relationships between these multi-omic layers and later-life cognitive function characterized the underlying etiology of cognitive/WMH outcomes and its interplay with socio-contextual factors in older adulthood, prior to dementia onset. This may allow the identification of targets for intervention and treatment for cognitive aging and dementia, especially among a highly vulnerable population.[35]

5.5 References

1. Tsuboi, K., Nagatomo, T., Gohno, T., Higuchi, T., Sasaki, S., Fujiki, N., Kurosumi, M., Takei, H., Yamaguchi, Y., Niwa, T., et al. (2017). Single CpG site methylation controls estrogen receptor gene transcription and correlates with hormone therapy resistance. J Steroid Biochem Mol Biol *171*, 209–217. 10.1016/j.jsbmb.2017.04.001.

2. Qiu, C., Shen, H., Fu, X., Xu, C., and Deng, H. (2018). Meta-Analysis of Genome-Wide Association Studies Identifies Novel Functional CpG-SNPs Associated with Bone Mineral Density at Lumbar Spine. Int J Genomics *2018*, 6407257. 10.1155/2018/6407257.

3. Gertz, J., Varley, K.E., Reddy, T.E., Bowling, K.M., Pauli, F., Parker, S.L., Kucera, K.S., Willard, H.F., and Myers, R.M. (2011). Analysis of DNA Methylation in a Three-Generation Family Reveals Widespread Genetic Influence on Epigenetic Regulation. PLoS Genet *7*, e1002228. 10.1371/journal.pgen.1002228.

4. Cohn-Schwartz, E. (2020). Pathways From Social Activities to Cognitive Functioning: The Role of Physical Activity and Mental Health. Innov Aging *4*, igaa015. 10.1093/geroni/igaa015.

5. Smith, J.A., Zhao, W., Wang, X., Ratliff, S.M., Mukherjee, B., Kardia, S.L.R., Liu, Y., Roux, A.V.D., and Needham, B.L. (2017). Neighborhood characteristics influence DNA methylation of genes involved in stress response and inflammation: The Multi-Ethnic Study of Atherosclerosis. Epigenetics *12*, 662–673. 10.1080/15592294.2017.1341026.

6. Marioni, R.E., McRae, A.F., Bressler, J., Colicino, E., Hannon, E., Li, S., Prada, D., Smith, J.A., Trevisi, L., Tsai, P.-C., et al. (2018). Meta-analysis of epigenome-wide association studies of cognitive abilities. Mol Psychiatry *23*, 2133–2144. 10.1038/s41380-017-0008-y.

7. McCartney, D.L., Hillary, R.F., Conole, E.L.S., Banos, D.T., Gadd, D.A., Walker, R.M., Nangle, C., Flaig, R., Campbell, A., Murray, A.D., et al. (2022). Blood-based epigenome-wide analyses of cognitive abilities. Genome Biology *23*, 26. 10.1186/s13059-021-02596-5.

8. Chu, S.H., Loucks, E.B., Kelsey, K.T., Gilman, S.E., Agha, G., Eaton, C.B., Buka, S.L., and Huang, Y.-T. (2018). Sex-specific epigenetic mediators between early life social disadvantage and adulthood BMI. Epigenomics *10*, 707–722.

9. Loucks, E.B., Huang, Y.-T., Agha, G., Chu, S., Eaton, C.B., Gilman, S.E., Buka, S.L., and Kelsey, K.T. (2016). Epigenetic mediators between childhood socioeconomic disadvantage and mid-life body mass index: the New England family study. Psychosomatic medicine *78*, 1053.

10. Wang, Y.Z., Zhao, W., Ammous, F., Song, Y., Du, J., Shang, L., Ratliff, S.M., Moore, K., Kelly, K.M., Needham, B.L., et al. (2022). DNA Methylation Mediates the Association Between Individual and Neighborhood Social Disadvantage and Cardiovascular Risk Factors. Front Cardiovasc Med *9*, 848768. 10.3389/fcvm.2022.848768.

11. Mayeda, E.R., Glymour, M.M., Quesenberry, C.P., and Whitmer, R.A. (2016). Inequalities in dementia incidence between six racial and ethnic groups over 14 years. Alzheimers Dement *12*, 216–224. 10.1016/j.jalz.2015.12.007.

12. Barnes, L.L., and Bennett, D.A. (2014). Alzheimer's disease in African Americans: risk factors and challenges for the future. Health Aff (Millwood) *33*, 580–586. 10.1377/hlthaff.2013.1353.

13. Brewster, P., Barnes, L., Haan, M., Johnson, J.K., Manly, J.J., Nápoles, A.M., Whitmer, R.A., Carvajal-Carmona, L., Early, D., Farias, S., et al. (2019). Progress and future challenges in aging and diversity research in the United States. Alzheimers Dement *15*, 995–1003. 10.1016/j.jalz.2018.07.221.

14. Rajan, K.B., Weuve, J., Barnes, L.L., Wilson, R.S., and Evans, D.A. (2019). Prevalence and incidence of clinically diagnosed Alzheimer's disease dementia from 1994 to 2012 in a population study. Alzheimers Dement *15*, 1–7. 10.1016/j.jalz.2018.07.216.

15. Shinohara, M., and Sato, N. (2017). Bidirectional interactions between diabetes and Alzheimer's disease. Neurochem Int *108*, 296–302. 10.1016/j.neuint.2017.04.020.

16. Whittington, R.A., Planel, E., and Terrando, N. (2017). Impaired Resolution of Inflammation in Alzheimer's Disease: A Review. Frontiers in Immunology *8*.

17. Cacabelos, R., and Torrellas, C. (2015). Epigenetics of Aging and Alzheimer's Disease: Implications for Pharmacogenomics and Drug Response. International Journal of Molecular Sciences *16*, 30483–30543. 10.3390/ijms161226236.

18. Bleecker, M.L., Bolla-Wilson, K., Kawas, C., and Agnew, J. (1988). Age-specific norms for the Mini-Mental State Exam. Neurology *38*, 1565–1565. 10.1212/WNL.38.10.1565.

19. BRAYNE, C., and CALLOWAY, P. (1990). The Association of Education and Socioeconomic Status with the Mini Mental State Examination and the Clinical Diagnosis of Dementia in Elderly People. Age and Ageing *19*, 91–96. 10.1093/ageing/19.2.91.

20. Crum, R.M., Anthony, J.C., Bassett, S.S., and Folstein, M.F. (1993). Population-Based Norms for the Mini-Mental State Examination by Age and Educational Level. JAMA *269*, 2386–2391. 10.1001/jama.1993.03500180078038.

21. Fridman, A.L., and Tainsky, M.A. (2008). Critical pathways in cellular senescence and immortalization revealed by gene expression profiling. Oncogene *27*, 5975–5987. 10.1038/onc.2008.213.

22. Yu, L., Chibnik, L.B., Yang, J., McCabe, C., Xu, J., Schneider, J.A., De Jager, P.L., and Bennett, D.A. (2016). Methylation profiles in peripheral blood CD4+ lymphocytes versus brain: The relation to Alzheimer's disease pathology. Alzheimers Dement *12*, 942–951. 10.1016/j.jalz.2016.02.009.

23. Lunnon, K., Smith, R., Hannon, E., De Jager, P.L., Srivastava, G., Volta, M., Troakes, C., Al-Sarraj, S., Burrage, J., Macdonald, R., et al. (2014). Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. Nat Neurosci *17*, 1164–1170. 10.1038/nn.3782.

24. Ashraf-ganjouei, A., Moradi, K., Bagheri, S., and Aarabi, M.H. (2020). The association between systemic inflammation and cognitive performance in healthy adults. Journal of Neuroimmunology *345*, 577272. 10.1016/j.jneuroim.2020.577272.

25. Weinstein, G., Lutski, M., Goldbourt, U., and Tanne, D. (2017). C-reactive protein is related to future cognitive impairment and decline in elderly individuals with cardiovascular disease. Archives of Gerontology and Geriatrics *69*, 31–37. 10.1016/j.archger.2016.11.002.