



Published in final edited form as:

*Mol Psychiatry*. 2023 March ; 28(3): 1293–1302. doi:10.1038/s41380-022-01890-3.

## African Ancestry GWAS of Dementia in a Large Military Cohort Identifies Significant Risk Loci

Richard Sherva<sup>1,2</sup>, Rui Zhang<sup>1</sup>, Nathan Sahelijo<sup>2</sup>, Gyungah Jun<sup>2</sup>, Tori Anglin<sup>3</sup>, Catherine Chanfreau<sup>3</sup>, Kelly Cho<sup>4,5</sup>, Jennifer R. Fonda<sup>6,7,8</sup>, J. Michael Gaziano<sup>4,5</sup>, Kelly M. Harrington<sup>4,6</sup>, Yuk-Lam Ho<sup>4</sup>, William Kremen<sup>9,10</sup>, Elizabeth Litkowski<sup>11,12</sup>, Julie Lynch<sup>3</sup>, Zoe Neale<sup>1,6</sup>, Panos Roussos<sup>13,14,15</sup>, David Marra<sup>4,6</sup>, Jesse Mez<sup>2,16,17</sup>, Mark W. Miller<sup>1,6</sup>, David H. Salat<sup>18</sup>, Debby Tsuang<sup>19</sup>, Erika Wolf<sup>1,6</sup>, Qing Zeng<sup>20</sup>, Matthew S. Panizzon<sup>9,10,21</sup>, Victoria C. Merritt<sup>9,21,22</sup>, Lindsay A. Farrer<sup>2,16,17,23,24,25</sup>, Richard L. Hauger<sup>9,10,22</sup>, Mark W. Logue<sup>1,2,6,23</sup>

<sup>1</sup>National Center for PTSD, Behavioral Sciences Division, VA Boston Healthcare System, Boston, MA, 02130, USA,

<sup>2</sup>Boston University School of Medicine, Biomedical Genetics, Boston, MA, 02118, USA,

<sup>3</sup>VA Informatics and Computing Infrastructure (VINCI), Salt Lake City, USA,

<sup>4</sup>Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC), VA Boston Healthcare System, Boston, MA, 02130,

<sup>5</sup>Division of Aging, Brigham & Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA,

<sup>6</sup>Department of Psychiatry, Boston University School of Medicine, Boston, MA, 02118 USA,

<sup>7</sup>Translational Research Center for TBI and Stress Disorders (TRACTS) and Geriatric Research, Educational and Clinical Center (GRECC), VA Boston Healthcare System, Boston, MA, 02130,

<sup>8</sup>Department of Psychiatry, Harvard Medical School, Boston, MA 02215,

<sup>9</sup>Department of Psychiatry, School of Medicine, University of California, San Diego, La Jolla, CA, USA,

<sup>10</sup>Center for Behavior Genetics of Aging, University of California, San Diego, La Jolla, CA, USA,

<sup>11</sup>Department of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, USA,

<sup>12</sup>VA Eastern Colorado Healthcare System, Aurora, CO, USA,

<sup>13</sup>Center for Disease Neurogenomics, Departments of Psychiatry and Genetics and Genomic Science, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA,

---

Corresponding Author: Mark W. Logue Ph.D., Statistician, National Center for PTSD, VA Boston Healthcare System, Associate Professor, Psychiatry, Boston University School of Medicine., Mail Stop 116B-2, VA Boston Healthcare System, 150 South Huntington Ave, Boston MA 02130, Phone: 857-364-4079.

Conflict of Interest

None of the authors have any conflicts of interest to disclose relating to this work.

<sup>14</sup>Center for Dementia Research, Nathan Kline Institute for Psychiatric Research, Orangeburg, NY 10962, USA,

<sup>15</sup>Mental Illness Research Education and Clinical Center (MIRECC), James J. Peters VA Medical Center, Bronx, New York, 10468, USA,

<sup>16</sup>Department of Neurology, Boston University School of Medicine, Boston, MA, USA,

<sup>17</sup>Boston University Alzheimer's Disease Research Center, Boston University School of Medicine, Boston, MA, USA,

<sup>18</sup>Neuroimaging Research for Veterans Center, VA Boston Healthcare System, Boston, MA, USA,

<sup>19</sup>Geriatric Research, Education, and Clinical Center, VA Puget Sound Health Care System, Seattle, WA, USA,

<sup>20</sup>VA Washington DC Healthcare System, Washington, DC, USA,

<sup>21</sup>Center of Excellence for Stress and Mental Health, VA San Diego Healthcare System, San Diego, CA, 92161,

<sup>22</sup>VA San Diego Healthcare System, 3350 La Jolla Village Dr, San Diego, CA, 92161,

<sup>23</sup>Department of Biostatistics, Boston University School of Public Health, Boston, MA, 02118, USA,

<sup>24</sup>Department of Ophthalmology, Boston University School of Medicine, Boston, MA, USA,.

<sup>25</sup>Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA,

## Abstract

While genome wide association studies (GWASs) of Alzheimer's Disease (AD) in European (EUR) ancestry cohorts have identified approximately 83 potentially independent AD risk loci, progress in non-European populations has lagged. In this study, data from the Million Veteran Program (MVP), a biobank which includes genetic data from more than 650,000 US Veteran participants, was used to examine dementia genetics in an African descent (AFR) cohort. A GWAS of Alzheimer's disease and related dementias (ADRD), an expanded AD phenotype including dementias such as vascular and non-specific dementia that included 4,012 cases and 18,435 controls age 60+ in AFR MVP participants was performed. A proxy dementia GWAS based on survey-reported parental AD or dementia (n=4,385 maternal cases, 2,256 paternal cases, and 45,970 controls) was also performed. These two GWASs were meta-analyzed, and then subsequently compared and meta-analyzed with the results from a previous AFR AD GWAS from the Alzheimer's Disease Genetics Consortium (ADGC). A meta-analysis of common variants across the MVP ADRD and proxy GWASs yielded GWAS significant associations in the region of *APOE* ( $p=2.48 \times 10^{-101}$ ), in *ROBO1* (rs11919682,  $p=1.63 \times 10^{-8}$ ), and RNA RP11-340A13.2 (rs148433063,  $p=8.56 \times 10^{-9}$ ). The MVP/ADGC meta-analysis yielded additional significant SNPs near known AD risk genes *TREM2* (rs73427293,  $p=2.95 \times 10^{-9}$ ), *CD2AP* (rs7738720,  $p=1.14 \times 10^{-9}$ ) and *ABCA7* (rs73505251,  $p=3.26 \times 10^{-10}$ ), although the peak variants observed in these genes differed from those previously reported in EUR and AFR cohorts. Of the genes in or near suggestive or genome-wide significant associated variants, nine (*CDA*, *SH2D5*, *DCBLD1*, *EML6*, *GOPC*, *ABCA7*, *ROS1*, *TMCO4*, and *TREM2*) were differentially expressed in the brains

of AD cases and controls. This represents the largest AFR GWAS of AD and dementia, finding non-*APOE* GWAS-significant common SNPs associated with dementia. Increasing representation of AFR participants is an important priority in genetic studies and may lead to increased insight into AD pathophysiology and reduce health disparities.

## Introduction

Late-onset Alzheimer's disease (AD) is the most common form of dementia<sup>1</sup>. The prevalence of AD in the United States (US) differs by ancestry group, with a higher proportion of AD in African ancestry (AFR) population than European ancestry (EUR) population (see e.g.<sup>2,3</sup>). The difference in prevalence between ancestries is not fully understood<sup>4</sup>, and may be due to a combination of factors, including higher rates of health and cardiovascular conditions that are associated with dementia risk<sup>5</sup> and societal factors which disadvantage AFR individuals in the US<sup>6</sup>.

Genetic factors also influence AD risk. The largest-effect genetic risk factor for AD is the *APOE*  $\epsilon 4$  allele. Genome-wide association studies (GWASs) of EUR cohorts have increased our knowledge of the genetic architecture of AD beyond the *APOE* locus<sup>7-9</sup>. The most recent large-scale EUR GWAS of AD by Bellenguez et al.<sup>10</sup>, reported 75 AD risk loci in addition to *APOE*. Their study included 20,464 AD cases and 22,244 controls as well as  $n=46,828$  "proxy" dementia cases in which case status was determined based on a reported history of AD or dementia in one or both parents. The sample sizes of non-EUR AD GWASs has been much smaller. Although there are consistent genes implicated in AD risk across different populations, the specific risk variants for a particular locus and the direction of effects for associated variants may differ by ancestry<sup>11</sup>. For example, the effect of the *APOE*  $\epsilon 4$  locus in AFR cohorts is approximately half of that observed in EUR cohorts<sup>1</sup>. Thus, results obtained in one ancestral group may not generalize to other ancestries. Several AD GWASs have looked specifically at AFR cohorts<sup>11-15</sup>. The AFR AD GWAS by Reitz et al. (2013)<sup>15</sup> included a cohort of 1,968 AD cases and 3,928 control participants, and, in addition to the *APOE* locus, identified a genome-wide significant association with a SNP (rs115550680,  $p=2.2 \times 10^{-9}$ ) in *ABCA7*. A larger AFR AD GWAS was recently published by Kunkle et al. (2021)<sup>12</sup> that included 2,784 screened AD cases and 5,222 controls from the Alzheimer's Disease Genetics Consortium (ADGC) cohorts. Although this was a small-scale study compared to the current EUR AD GWASs, 4 novel common loci were identified at the suggestive significance level ( $p < 0.05 \times 10^{-7}$ ), with variants in *EDEMI*, *ALCAM*, *GPC6*, and *VRK3*. A secondary analysis in Kunkle et al. that covaried for *APOE* genotype identified a genome-wide significant rare variant in an intergenic region on chromosome 15. In addition, they reported on gene expression data from EUR brain tissue showing associations between expression of *ALCAM*, *ARAPI*, *GPC6*, and *RBFOX1* and brain  $\beta$ -amyloid load. Other studies have implicated AFR-specific rare variants in the gene *AKAP9* in association with AD<sup>16-18</sup>. A trans-ethnic meta-analysis from 2017 identified additional AD risk variants in *PFDNI/HBEGF*, *USP6NL/ECHDC3*, *BZRAPI-ASI*, *NFIC*, and *TPBG*<sup>11</sup>. The work by Mez et al. (2017)<sup>14</sup> used a conditional liability model to identify risk variants in *COBL* and *SLC10A2*.

To identify AD-associated variants in AFR populations, we performed a GWAS in AFR participants of a large, national healthcare system biorepository, the Department of Veterans Affairs (VA) Million Veteran Program (MVP). These participants are former US military service members who obtain their healthcare in the VA. MVP includes genetic data on MVP participants linked to their VA electronic medical record. While we are primarily interested in identifying AD risk variants, due to the limitations in the specificity associated with clinical diagnoses, we analyzed AD and related dementias (ADRD) cases and controls. We also conducted a “proxy dementia” GWAS of self-reported parental dementia, similar to the proxy GWAS included in recent EUR AD GWAS studies (e.g.<sup>10, 19</sup>). This represents the largest GWAS of dementia in an AFR cohort, as well as the first use of proxy cases in AFR dementia GWAS meta-analyses. To increase power, we also performed a meta-analysis of the MVP-based GWAS with the results from the ADGC AFR AD GWAS and performed gene-based tests and pathway analysis in MVP as well as MVP+ADGC.

## Methods

### MVP Subjects and Diagnostic Classification Procedures

Subjects for this study included AFR MVP participants as determined by the genetically informed harmonized ancestry and race/ethnicity (HARE) method<sup>20</sup>. This method combines self-report and genetically determined principal components in a machine learning model to assign participants to one of four groups: AFR, EUR, ASN (East Asian), and HIS (Hispanic). Individuals with indeterminate ancestry and individuals whose self-reported Race did not match genetically informed ancestry were excluded from classification. Consistent with other electronic medical record studies of dementia, ADRD cases were defined as participants with diagnosis codes for AD, non-specific dementia<sup>21</sup>, or other related dementias diagnosis codes (i.e., vascular dementia, Lewy body dementia, frontotemporal dementia, and normal pressure hydrocephalus<sup>22</sup>), which are included in Supplementary Table 1. Because AD accounts for more than 50% of individuals with dementia<sup>1</sup> and due to VA code usage<sup>21</sup>, it is likely that a large portion of MVP participants who were assigned ICD codes for non-specific dementia actually have AD. Further, we required two or more of these diagnosis codes and an age of onset (first diagnosis code) at or above age 60 to be included as a case. After exclusions for relatedness and other considerations (described in detail below), 4,012 participants met the ICD-code criteria for ADRD were included in the genetic analyses. A total of 18,435 individuals (representing approximately 4.5 times the number of ADRD cases) who were ages 65 and older without any dementia ICD codes, and who did not report a parental history of dementia in a health survey, were randomly selected as controls for the ADRD GWAS. The remainder were included as controls for the proxy GWAS described below.

The majority of MVP participants completed the “Baseline” survey which included questions regarding a wide variety of topics including employment, personality, psychiatric disorders, and family history of disease. For the proxy dementia GWAS, which excluded subjects used in the ADRD GWAS described above, participants who reported a history of “Alzheimer’s/ Other dementia” in their father or mother were classified as cases in the paternal or maternal proxy dementia GWASs respectively, and individuals with two affected

parents contributed to both the maternal and paternal proxy analyses. Because the specific type of dementia is not confirmed or even recorded for parents of participants, we refer to this as a proxy dementia analysis and refer to any meta-analysis including these subjects as a meta-analysis of dementia rather than AD or ADRD to reflect the ambiguity in proxy case diagnosis. However, EUR GWAS results indicate a strong genetic correlation between proxy dementia GWAS and non-proxy AD case-control results in ( $r_g=0.81$ )<sup>19</sup>. In order to limit the number of individuals with parents too young to be at risk for dementia, participants ages 45 and older without any dementia ICD codes (i.e. any codes from Supplementary Table 1) who were not included in the ADRD controls and did not report either parent as having dementia were included as proxy controls. Proxy cases were not screened on age. After applying these criteria, 4,385 maternal proxy cases, 2,256 paternal proxy cases, and a common proxy control cohort of 45,970 subjects remained for analysis.

### Genotype Data Generation and Processing

Genotype data processing and cleaning was performed by the MVP Bioinformatics core. The genotype data were generated using the MVP 1.0 custom Axiom array which assays 668,418 markers. Quality control (QC) included checks for sex concordance, advanced genotyping batch correction, and assessment for relatedness. The chip design and genotype cleaning pipeline have been described elsewhere<sup>23</sup>. The MVP Phase 4 genotype data release includes imputed genotype data for 62 million variants assessed for approximately 650,000 subjects. Imputation was performed using the African Genome Resources (AGR) panel from the Sanger Institute which includes the 1000 Genomes Phase 3 reference panel plus another 1,431 unrelated African-descent individuals to improve AFR imputation accuracy. Prior to imputation, SNPs with high missingness (20%), monomorphic, not in Hardy-Weinberg equilibrium, and that differed in frequency between batches were removed. Phasing was done with SHAPEIT4 v 4.1.3 and imputation was performed with MINIMAC4. One of each of a pair of related individuals, defined as a kinship coefficient of 0.09375 or higher, were removed from analysis using a scheme that prioritized a case over a control while the older control was selected if both were controls. When both members of a pair were ADRD cases, we selected a subject with AD-specific ICD codes or, in the absence of ICD codes, one individual was randomly selected. Principal components of ancestry were computed with FlashPCA2<sup>24</sup>, using only AFR individuals and a linkage-disequilibrium pruned set of 170,207 SNPs that excluded the major histocompatibility complex region of chromosome 6. *APOE* genotypes were determined using the “best guess” imputed genotypes (80% confidence threshold) for the rs7412 and rs429358 SNPs which were well imputed in the AFR cohort ( $r^2=0.87$  and  $0.99$ , respectively). The inclusion criteria for SNPs in SNP- and gene-based tests were set to that used in Kunkle et al. 2021<sup>12</sup>. Specifically, for the primary GWAS, SNPs with minor allele frequency (MAF) 1% and imputation quality ( $r^2$ ) 0.4 were included (N=16,589,632). For gene-based tests, SNPs with MAF 0.1% and  $r^2$  0.7 were included (N=20,289,645).

### Statistical Analysis

Association of each SNP with ADRD and proxy dementia was tested using logistic regression models implemented in PLINK 2.0. Firth logistic regression<sup>25</sup> was applied when the standard regression model failed to converge. Models were adjusted for sex and the first

ten ancestry principal components. GWAS for the maternal and paternal proxy AD/dementia outcomes were performed separately using a common set of controls, and the results from the maternal and paternal analyses were combined and meta-analyzed using the sample size weighted Z-score method as implemented in METAL including a correction for overlapping samples<sup>26</sup>. Results from the ADRD GWAS and proxy GWAS were then combined by weighted Z-score meta-analysis to generate the MVP AFR dementia GWAS results. As a sensitivity analysis, top hits were tested for association with a strict AD definition which included only cases with two or more AD ICD codes (see Supplementary Table 1 for codes, n=741) vs the non-demented controls described above. In addition, to match analyses performed in Kunkle et al,<sup>12</sup> we also performed a GWAS adjusting for *APOE*  $\epsilon$ 4 allele dosage.

Gene-based tests, along with functional mapping and annotation, expression quantitative trait locus (eQTL) analyses, gene set enrichment tests, and canonical pathway analysis, were conducted using the FUMA web portal<sup>27</sup> which uses the same methodology implemented in MAGMA<sup>28</sup>. Manhattan and regional plots were produced using FUMA. All presented genomic coordinates are according to GRCh37/hg19. Pairwise LD values were calculated based on 1000 genomes AFR population data using LDlink (<https://ldlink.nci.nih.gov>).

### Expression Profiling

We selected genes harboring variants with suggestive or genome-wide significant variants, as well as genes containing a SNP in LD ( $r^2 > 0.6$ ) with a suggestive SNP (including rare variants) identified by FUMA for further evaluation. As no comparably sized post-mortem AFR AD case-control cohorts exist, differential expression of these genes was assessed using RNA sequencing data generated from autopsied brains of 526 EUR AD cases and 456 controls from the Religious Orders Study/Rush Memory and Aging Project (ROS/MAP), the Mayo-Mount Sinai Brain Bank, the Framingham Heart Study, and the Boston University Alzheimer's Disease Research Center Brain Bank<sup>29</sup>. As these were EUR cohorts, these data could not be used to validate specific associated loci as eQTLs, but were instead used to examine variations in expression profile of specific genes potentially relevant to dementia pathogenesis. The details of data generation, processing, and quality control were described elsewhere<sup>29</sup>. In brief, expression of each gene/isoform was compared between AD cases and controls using the LIMMA program including age of death, sex, and RNA integrity number as covariates. A Bonferroni correction was used to adjust for the number of genes examined. We also examined the top associated variants and variants in LD for possible cis-acting eQTL effects in a publicly available eQTL database derived from AFR non-demented brain-expression profiles<sup>30</sup>.

### Results

The mean age at onset of ADRD cases was 73.8 years, which was similar to age at last exam for the controls (73.0 years). As typical for a cohort of older US Veterans, the proportion of women in the study was small; 3.0% of cases (n=120) and 5.9% of controls (n=1,090) were women. See Supplementary Figure 1 for a plot of the first two ancestry PCs of MVP AFR participants and 1000 Genomes Phase 3<sup>31</sup> reference samples. Supplementary Figure 2

contains a plot of the ancestry proportions of the MVP AFR participants as estimated using the program ADMIXTURE<sup>32</sup> based on several 1000 genomes subpopulations as proxies for continental ancestry groups. The mean proportion of African ancestry estimated for the MVP AFR cohort was 82% followed by 16% European ancestry.

### Association Findings in MVP

There was little evidence of genomic inflation in the ADRD GWAS ( $\lambda=1.01$ ), the proxy GWAS ( $\lambda=1.006$ ), or in the combined results from the ADRD and proxy GWAS ( $\lambda=1.11$ ). Figure 1 shows the Manhattan and quantile-quantile (QQ) plots for the meta-analysis of the ADRD and proxy dementia GWASs.

Genome-wide significant association was observed with *APOE* (Figure 2 and Supplementary Table 2). The odds ratios (ORs) of *APOE*  $\epsilon 2/\epsilon 4$  and  $\epsilon 3/\epsilon 4$  subjects were 1.45 ( $p=0.0001$ ) and 1.44 ( $p=2.66\times 10^{-18}$ ), respectively, compared to  $\epsilon 3$  homozygotes as a reference, which was approximately half of the effect size of *APOE*  $\epsilon 4/\epsilon 4$  (OR=2.90) compared to the reference genotype (Supplementary Table 2). No genome-wide significant associations were observed for variants with MAF>1% with ADRD outside the *APOE* region.

After meta-analysis with the MVP maternal-paternal proxy dementia results, two regions outside the *APOE* region showed genome-wide significant association: rs11919682 in the gene encoding roundabout guidance receptor 1 (*ROBO1*) on chromosome 3 ( $p=1.63\times 10^{-8}$ ) and rs148433063 in pseudogene *RP11-340A13.2* (aka *LINC02429*) on chromosome 4 ( $p=4.36\times 10^{-9}$ ; Supplementary Figure 3). There were also 15 suggestive ( $p<5.0\times 10^{-7}$ ) associations including SNPs near *CD2AP* (rs7738720), *TREM2* (rs2234253), and *ABCA7* (rs73505251) that were previously established as AD genes in EUR subjects (see Table 1 for details). Importantly, none of these were the same SNPs identified in recent GWASs in EUR cohorts<sup>7-10</sup>. Gene-based tests yielded one genome-wide significant association with *TREM2* (Bonferroni-corrected  $p=0.02$ ), consistent with the significant *TREM2* gene-based association observed by Kunkle et al.<sup>12</sup> Adjustment for *APOE*  $\epsilon 4$  did not yield any genome-wide significant associations, but seven variants reached the suggestive level (Supplementary Table 3), including SNPs in or near *LHX6* (rs181518405,  $p=3.90\times 10^{-7}$ ), *CFAP46* (rs146777408,  $p=3.34\times 10^{-7}$ ), and *OCSTAMP* (rs202380,  $p=2.98\times 10^{-7}$ ).

As a sensitivity analysis, to examine whether the dementia-associated loci may be associated with AD specifically, we examined these SNPs in association with the strict AD phenotype. The *ROBO1* and *LINC02429* genome-wide significant SNPs and 9 of the 15 suggestive associations for ADRD were at least nominally associated with AD with the same effect direction (Table 1), noting, however, that the number of subjects with the strict AD phenotype was much smaller than those with the broader ADRD phenotype. Among the genome-wide significant and suggestive SNPs from the previous ADGC AFR GWAS implicated in either the *APOE*  $\epsilon 4$ -adjusted or unadjusted analysis, ten were evaluated in MVP, of which SNP rs9516245 in *GPC6* was nominally significant ( $p=0.03$ ). The rare *IGFR1* SNP reported as genome-wide significant in the ADGC *APOE*  $\epsilon 4$ -adjusted analysis was not present in the MVP imputed results. Apart from *TREM2*, none of the other gene-based tests were nominally significant including those surpassing the  $p<1\times 10^{-4}$  threshold

in the ADGC (Kunkle et al.) GWAS (i.e., *ARAPI*, *TRANK1*, *FABP2*, *LARP1B*, *TSRM*, *STARD10*, *SPHK1*, and *SERPINB13*).

### Dementia GWAS in Combined Datasets

Figure 2 shows the Manhattan and QQ plots for the full meta-analysis of the MVP and ADGC cohorts. After combining the GWAS results from the MVP and ADGC cohorts, variants in three known AD loci outside of the *APOE* region that were significant at the suggestive level in MVP alone emerged as genome-wide significant in the meta-analysis: *CD2AP* (rs7738720,  $p=1.14\times 10^{-9}$ ), a SNP upstream of *TREM2* (rs73427293,  $p=2.95\times 10^{-9}$ ) that is in high LD with rs2234253 (the peak variant in the MVP-alone analysis), and *ABCA7* (rs73505251,  $p=3.26\times 10^{-10}$ ; Figure 2, Table 2, Supplementary Figure 4). Of the two genome-wide significant associations observed in the MVP GWAS, rs148433063 on chromosome 4 was not significant in the ADGC dataset ( $p=0.9$ ) or in the total sample ( $p=1.54\times 10^{-7}$ ). No result was available for the *ROBO1* SNP rs11919682 in the ADGC dataset. Notably, rs11919682 is a tri-allelic variant and may have been excluded from analysis in the ADGC dataset, although no copies of the rare alternate allele were present in the MVP imputed data. The second most significantly associated SNP in the *ROBO1* region, rs61053911 ( $p=9.36\times 10^{-8}$ ), is moderately correlated with the peak *ROBO1* SNP ( $r^2=0.35$ ) but was not significant in the ADGC dataset ( $p=0.66$ ). SNPs in 17 additional regions were significant at the suggestive level including rs4607615 in *MSRA* that was nearly genome-wide significant ( $p=7.37\times 10^{-8}$ ) and rs509334, which is approximately 45KB upstream from the known AD risk gene *SORL1*<sup>10, 33</sup> (Supplementary Figure 5). Results for three of these variants (rs567572378 in *KCNH8-EFHB*, rs192764155 in *RP11-506N2.1*, and rs2234253 near *TREM2*) were not available in the ADGC dataset. Seven rare variants (MAF<1%) showed suggestive evidence of association (Supplementary Table 4).

No additional genome-wide significant associations were found in the combined results from *APOE*  $\epsilon 4$ -adjusted models for ADRD in MVP and AD in ADGC, but there were two suggestive associations not observed in Model 1 including *RP11-144F15.1* (rs114572990,  $p=4.97\times 10^{-7}$ ) and *DNAI2* (rs9893381,  $p=4.89\times 10^{-7}$ ; Supplementary Table 5). Gene based tests for *TREM2* (Bonferroni corrected  $p=1.9\times 10^{-5}$ ) and *ABCA7* (Bonferroni corrected  $p=0.005$ ) were significant in the total sample.

### Previously implicated AD variants in African Americans

Reitz et al.<sup>15</sup> reported a genome-wide significant association with rs115550680 in *ABCA7* ( $p=2.2\times 10^{-9}$ ) but this SNP was not significant in a larger GWAS conducted by the ADGC (Kunkle et al.<sup>12</sup>) including the Reitz et al. sample ( $p=3.27\times 10^{-5}$ ). This SNP has again reached genome-wide significant ( $p=1.17\times 10^{-8}$ ) in our analysis combining the MVP and ADGC datasets. Of note, this SNP (rs115550680 in *ABCA7*) is not correlated with the top-ranked *ABCA7* SNP in the MVP study (rs73505251,  $r^2=0.013$ ), suggesting the possibility that these SNPs tag distinct functional variants. Neither the *ABCA7* frameshift deletion reported in Cukier et al.<sup>34</sup> (rs142076058), nor the insertion/deletion polymorphism in the gene reported by Logue et al. (rs567222111)<sup>16</sup> were nominally significant in the combined MVP sample ( $p>0.05$ ). *TREM2* SNPs previously reported to be associated with AD in AFR studies, including a stop-gain mutation<sup>35</sup>, were significant in the combined MVP and ADGC



datasets (rs7748513:  $p=6.06\times 10^{-6}$ , rs2234256:  $p=4.635\times 10^{-9}$ , rs2234258:  $p=5.48\times 10^{-7}$ ). All of these are in high LD with rs73427293, the peak *TREM2* SNP in the MVP+ADGC meta-analysis in terms of  $D'$  (approximately 1.0), and moderate to strong LD in terms of  $r^2$  (0.14–0.99). Neither of the genome-wide significant SNPs in *COBL* and *SLC10A2* identified by Mez et al.<sup>14</sup> were nominally significant in the full meta-analysis. Two rare AFR-specific non-synonymous *AKAP9* missense variants that were associated with AD in a smaller AFR sample<sup>18</sup> were nominally associated with the combined MVP and ADGC datasets in a model adjusting for *APOE*  $\epsilon 4$  (rs144662445,  $p=0.04$ ; rs149979685,  $p=0.03$ ), but neither was significant in the non-adjusted model. Among the tests of aggregated rare variants in genes ascribed to 74 distinct genome-wide significant loci reported in Bellenguez et al.<sup>10</sup>, only *TREM2*, *TREML2*, *CD2AP*, and *ABCA7* were significant in the combined MVP and ADGC datasets after Bonferroni correction ( $p<0.05$ ).

### Functional Annotation

Because the most significant GWAS SNPs are often not the functional variants, we examined non-synonymous coding variants and eQTLs in LD with variants with genome-wide significant or suggestive evidence for association with AD. Excluding the *APOE* region, we identified two non-synonymous *ABCA7* variants (rs73505232,  $p=4.40\times 10^{-10}$ , LD with peak SNP  $r^2=0.87$ , and rs59851484,  $p=3.22\times 10^{-8}$ ,  $r^2=0.66$ ) that are in LD with rs73505251, the peak *ABCA7* SNP. FUMA identified two candidate causal variants in LD with peak *TREM2* SNP: the previously-identified *TREM2* stop mutation and another nonsynonymous SNP (rs2234256 previously noted in<sup>35</sup>,  $p=4.64\times 10^{-9}$ , LD with peak SNP  $r^2=0.99$  and rs2234253,  $p=2.23\times 10^{-7}$ , LD with peak SNP  $r^2=0.99$ ). Supplementary Table 6 shows the association results for non-synonymous variants in the region of the peak variants. We additionally identified significant eQTLs in the region of the *ABCA7* and *MSRA* gene peaks according to MetaBrain Network AFR cortex data<sup>30</sup> (see Supplementary Table 7). Several SNPs in LD with the peak *ABCA7* SNP, rs34606911, are associated with *ABCA7* expression, most significantly rs115550680 (eQTL  $p$ -value= $1.78\times 10^{-7}$ ;  $p$ -value in dementia meta= $1.17\times 10^{-8}$ ,  $r^2=0.69$  to peak *ABCA7* SNP). “A” alleles of this variant are associated with lower *ABCA7* expression and also a lower risk of dementia. We also indicated ten dementia-associated ( $p=0.002$  to  $7.37\times 10^{-8}$ ) eQTLs in LD with the peak *MSRA* SNP ( $r^2$  from 0.63 to 0.86) which are associated with expression of the flanking gene, serine protease 51 (*PRSS51*), most significantly rs3750314 (eQTL  $p=1.60\times 10^{-22}$ , dementia meta  $p=9.4\times 10^{-6}$ ). The rs3750314 minor allele (C) is associated with higher *PRSS51* expression in post-mortem brains and higher risk of dementia.

### Pathway and Gene Set Enrichment Analyses

Gene set enrichment analyses in FUMA based on the gene-based tests indicated that a set including 8 genes involved in regulation of resting membrane potential was significantly enriched (Bonferroni corrected  $p=0.03$ ). The enrichment was accounted for primarily by *TREM2* (gene-based  $p=1.00\times 10^{-9}$ ), although the gene-based tests for *PSENI* ( $p=0.04$ ), *KCNG2* ( $p=0.06$ ), *KCNK6* ( $p=0.07$ ), and *KCNK1* ( $p=0.07$ ) also contributed to the enrichment signal. Five significant canonical pathways were identified and mostly accounted for by genes in the *APOE* region. These pathways were related to lipid metabolism, except for one involving immunoregulatory interactions between lymphoid and non-lymphoid cells

for which the major contributors included *PVRL2*, *TREM2*, *TREML1*, and *TREML2* (Supplementary Figure 6). No significant enrichment was observed for gene sets or canonical pathways using results from the model adjusting for *APOE*  $\epsilon 4$ .

### Differential Expression

We examined differential expression of 30 genes with genome-wide significant or suggestively significant rare or common variants (Tables 2 and S4) or SNPs in LD with these variants in the post-mortem EUR AD case control cohort. Eighteen of 29 genes for which results were available showed significantly different expression between AD cases and controls, and nine remained significant after Bonferroni correction ( $p < 0.05$ ): *CDA*, *SH2D5*, *DCBLD1*, *EML6*, *GOPC*, *ABCA7*, *ROS1*, *TMCO4*, and *TREM2*. *CDA*, *ABCA7*, *TMCO4*, and *TREM2* were upregulated in AD cases and the other genes were downregulated (Table 3).

### Discussion

Increasing the representation of non-European populations in GWASs has been identified as a critical scientific and equity issue in genetic studies.<sup>36, 37</sup> The differences in sample sizes between EUR and non-EUR studies to date could contribute to health disparities as a function of ancestry and the application of genetic information derived from EUR populations has reduced relevance and accuracy when applied to non-EUR populations<sup>37, 38</sup>. Moreover, the diversity gap in GWAS studies represents a missed opportunity because studies of traits in multiple populations have the opportunity to yield population-specific risk loci that can implicate new molecular pathways and potential therapeutic targets, as well as help to identify causal variants. This study, which utilized a large US Veteran cohort of AFR from the MVP project, therefore represents an important milestone in dementia genetics research.

This genetic study of ADRD, which included a sample that is more than twice the size of the previously largest AFR AD GWAS yielded several important findings relevant to AFR populations. First, using a large AFR cohort which included 4,012 ADRD cases, we examined the impact of the *APOE* AD risk locus genotypes. Whereas in the past, the literature had been inconsistent on whether AFR heterozygous carriers of the high risk  $\epsilon 4$  allele were at increased AD risk (e.g.<sup>39,40</sup>), here, it is clear that compared to the common  $\epsilon 3/\epsilon 3$  genotypes, those with either the  $\epsilon 3/\epsilon 4$  and  $\epsilon 2/\epsilon 4$  genotypes are indeed at increased ADRD risk (OR=1.44 and 1.45 respectively  $p < 0.001$ ). This risk was intermediate to  $\epsilon 4/\epsilon 4$  risk of OR=2.66. However, we did not confirm the effect of the uncommon  $\epsilon 2$  allele, which has been shown to be protective against AD in AFR and EUR. In our GWAS analyses, we confirmed the previously observed genome-wide significant AFR association with *APOE* and *ABCA7*. Additionally, we obtained the first genome-wide significant evidence for association in AFR cohorts with *TREM2* and *CD2AP*, known loci from EUR GWASs. This study also was the first to identify suggestive association in the region of the *SORL1* gene. The peak AFR SNP (rs509334) was not in LD ( $r$  less than 0.01) with either of the two EUR risk variants (rs11218343 and rs74685827) in AFR or combined reference panels,

suggesting independent risk variants may exist. Finally, we found novel genome-wide significant associations with *ROBO1* and *RP11-340A13.2* loci.

The MVP GWAS meta-analysis also implicated several other genes whose impact on AD risk is less established. Not much is known about the function of *RP11-340A13.2* (aka *LINC02429*) that encodes a long intergenic non-protein coding RNA. However, there is strong biological rationale to suggest that *ROBO1* is an AD risk locus. *ROBO1* encodes the ROBO1 transmembrane receptor that binds SLIT proteins to regulate axon guidance and prevent axons from crossing the midline of the CNS<sup>41, 42</sup>. There is an extensive literature on the SLIT/ROBO signaling that mediates formation and maintenance of neural circuits in the hippocampus<sup>43</sup>. Furthermore, SLIT proteins bind to both ROBO1 and APP receptors and their direct interaction triggers APP processing and ectodomain shedding, which can dysregulate APP-mediated axon pathfinding and contribute to AD pathophysiology<sup>44, 45</sup>. Additionally, *ROBO1* was recently implicated among the top genes in a study that examined single-variant and spatial clustering-based testing on rare variants in a whole-genome sequencing study of AD in a EUR population<sup>46</sup>, although this is the first time it has been identified as genome-wide significant. Because results for the associated *ROBO1* SNP were not available in the ADGC analysis and the association was not significant in the analysis of the strict AD phenotype (p=0.10), it is possible that this association may not be specific for AD but rather for dementia more broadly.

Several suggestive level associated SNPs in the MVP+ADGC meta-analysis are located in or near biologically plausible candidate genes including *MSRA*, *EML6*, *CDA*, and *GOPC*. Six SNPs that are in high LD with the peak *MSRA* SNP are significant eQTLs for *MSRA* in multiple brain tissues. *MSRA* reduces oxidative stress, and *MSRA* upregulation resulted in a marked reduction in aging phenotypes in transgenic mice<sup>47</sup>. Active in mitochondria, *MSRA*'s primary function is to reduce methionine sulfoxide, a potentially harmful reactive oxygen species produced during oxidative bursts, common during the neutrophil- and macrophage-mediated response to bacterial infection. Alternately, SNPs in this gene may affect dementia risk by altering the expression of the neighboring gene, *PRSS51*, although little is known about this gene other than that has endopeptidase and proteolytic functions. We found that *EML6* was significantly downregulated in prefrontal cortex tissue from AD brains. *EML6* may modify the assembly dynamics of microtubules, making them longer and more dynamic, and is associated with sodium dependent hypertension<sup>48</sup>. In addition, we identified three genes in or near rare variants associated with ADRD at the suggestive level that are also differentially expressed in brain tissue between AD cases and controls: *CDA*, *SH2D5*, and *GOPC*. *CDA* catalyzes the hydrolytic deamination of cytidine and deoxycytidine to uridine and deoxyuridine and is a marker of monocyte to macrophage differentiation<sup>49</sup>. *SH2D5* encodes a mammalian-specific adaptor-like protein highly enriched in the brain and, based on the function of its homologue in mice, may modulate synaptic plasticity by regulating breakpoint cluster region protein Rac-GTP levels<sup>50</sup>. *GOPC* plays a role in intracellular protein trafficking and degradation and may affect AD pathogenesis by serving as a scaffold protein forming a complex between the CD46 receptor and Beclin 1<sup>51</sup>. Beclin 1 expression is downregulated in the entorhinal and frontal cortex during early AD, which would be expected to impair neuronal autophagy, dysregulate APP processing and increase  $\beta$ -amyloid deposition, and promote neurodegeneration and microgliosis<sup>52</sup>. In

addition, GOPC suppresses complement attacks and inflammation<sup>53</sup>, is cleaved by the  $\gamma$ -secretase complex in response to microbial infection<sup>54</sup>, and impairment of these activities in AD could reduce microglial phagocytic capacity and amyloid- $\beta$  clearance<sup>55, 56</sup>.

## Strengths and Limitations

This work represents the largest GWAS of AD and related dementias in individuals of AFR to date and is the first to focus on a veteran population. Veterans may be at increased risk for dementia due to higher rates of obesity-related health conditions, traumatic brain injury, and posttraumatic stress disorder<sup>57, 58</sup>. In addition, we obtained evidence that several of the most significantly associated genes are differentially expressed in the prefrontal cortex from AD cases and controls. Finally, SNPs in two novel loci (notably in *MSRA*) showed evidence for association in both the MVP and ADGC cohorts. However, several limitations to the study should be noted. First, because the sample size available for these analyses is small compared to AD GWAS in EUR GWAS studies, the opportunity for novel discovery is reduced and even some of the highly significant findings may not be robust. We did not observe corresponding association with the *ROBO1* and *RP11-340A13.2* in the ADGC data. This may be due to a lack of power, or from heterogeneity between the two cohorts. The MVP and ADGC cohorts differ in several important respects, including the proportion of males, the analysis phenotype, and the ascertainment scheme. It is also possible that these results represent false positives. However, the genome-wide significant associations observed with SNPs in AD loci previously established in other populations increases confidence that the findings with novel loci represent reliable associations. Second, because our findings were obtained from a sample including overlapping but non-identical phenotypes of AD, ADRD, and proxy dementia, the associations with novel loci may not be AD-specific. However, it is likely that most of these findings are AD-related given that AD is by far the most common form of dementia, the majority of the significant associations were observed with loci that were previously established in AD-specific cohorts, and ten of the 23 independent lead SNPs outside the *APOE* region were at least nominally associated with a strict AD diagnosis despite the substantially reduced sample size. Nonetheless, the novel genes identified in this study may not directly affect AD-related pathology but rather processes common to multiple types of dementia such as general cognitive function or memory. Finally, discordance of association findings between the MVP and ADGC cohorts in this study may reflect sex-specific influences noting that proportion of women differs widely between the MVP (approximately 10%) and the ADGC (approximately 60%) cohorts. Additional studies are needed to confirm and extend our findings in an enlarged AFR sample, as well as assess whether heterogeneity in the results between the MVP and ADGC AFR cohorts is due to the influence of unique environmental exposures for a largely male US Veteran cohort.

## Conclusions

The findings of this study build on the two known genome-wide significant common variant AD associations from AFR cohorts, *APOE* and possibly *ABCA7*, to include *TREM2*, *CD2AP*, *ROBO1*, and *RP11-340A13.2*. In addition, the top-ranked SNPs for *ABCA7*, *TREM2* and *CD2AP* differ between AFR and EUR studies suggesting the possibility of

distinct causal variants in these populations. The loci and associations identified in this study represent a substantial increase in the knowledge of the genetic architecture of dementia risk in AFR populations. The insights they provide will pave the way for more accurate risk assessment for individuals of AFR. Moreover, the newly identified risk loci can contribute to a broader understanding of dementia pathogenesis which may provide targets for AD and dementia treatment.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Funding

This research is based on data from the Million Veteran Program, Office of Research and Development, Veterans Health Administration, and was supported by VA BLR&D grant 1 I01 BX004192 (MVP015). ZN was supported by National Institute of Mental Health award T32MH019836. JRF was supported by VA CSR&D grant 1HK2CX002192-01A2. This publication does not represent the views of the Department of Veteran Affairs or the United States Government.

## Data Availability

The data underlying this publication is accessible to researchers with MVP data access. MVP is currently only accessible to researchers an approved funded MVP project, either through a VA Merit Award or a career development award. See <https://www.research.va.gov/funding/Guidance-MVP-Data-Access-Merit-Award.pdf> and for details. Summary statistics of the ADRD and proxy GWASs will be made publicly available on dbGAP.

## References

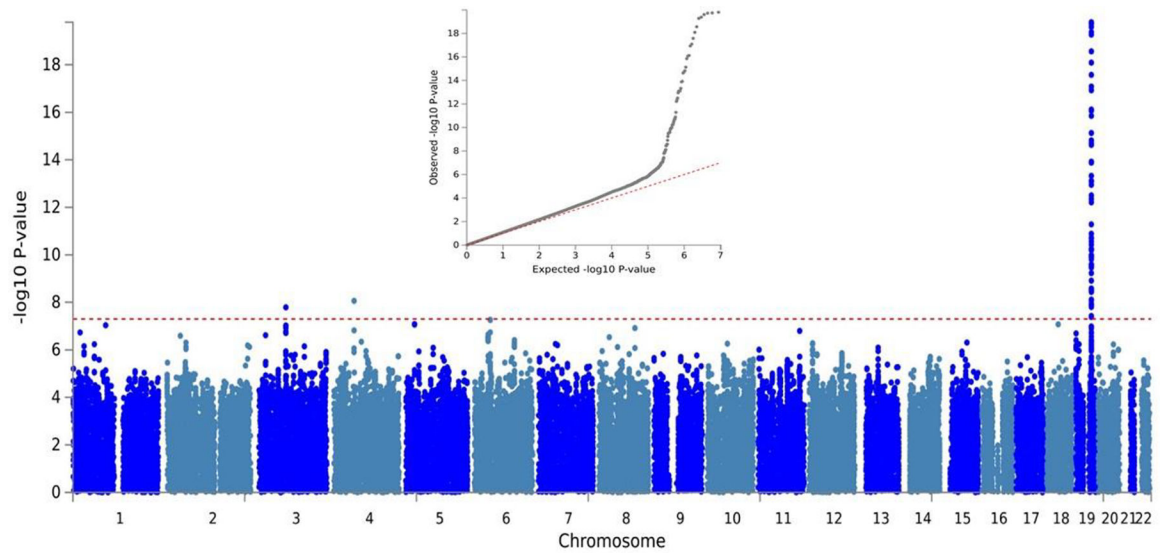
1. 2020 Alzheimer's disease facts and figures. *Alzheimers Dement* 2020.
2. Mayeda ER, Glymour MM, Quesenberry CP, Whitmer RA. Inequalities in dementia incidence between six racial and ethnic groups over 14 years. *Alzheimers Dement* 2016; 12(3): 216–224. [PubMed: 26874595]
3. Matthews KA, Xu W, Gaglioti AH, Holt JB, Croft JB, Mack D et al. Racial and ethnic estimates of Alzheimer's disease and related dementias in the United States (2015–2060) in adults aged  $\geq 65$  years. *Alzheimers Dement* 2019; 15(1): 17–24. [PubMed: 30243772]
4. Barnes LL. Alzheimer disease in African American individuals: increased incidence or not enough data? *Nat Rev Neurol* 2022; 18(1): 56–62. [PubMed: 34873310]
5. Clark LR, Norton D, Berman SE, Johnson SC, Bendlin BB, Wieben O et al. Association of Cardiovascular and Alzheimer's Disease Risk Factors with Intracranial Arterial Blood Flow in Whites and African Americans. *J Alzheimers Dis* 2019; 72(3): 919–929. [PubMed: 31658057]
6. Yaffe K, Falvey C Fau - Harris TB, Harris Tb Fau - Newman A, Newman A Fau - Satterfield S, Satterfield S Fau - Koster A, Koster A Fau - Ayonayon H et al. Effect of socioeconomic disparities on incidence of dementia among biracial older adults: prospective study. *BMJ* 2013; 347: f7051. [PubMed: 24355614]
7. Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* 2013; 45(12): 1452–1458. [PubMed: 24162737]
8. Kunkle BW, Grenier-Boley B, Sims R, Bis JC, Damotte V, Naj AC et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. *Nat Genet* 2019; 51(3): 414–430. [PubMed: 30820047]

9. Naj AC, Jun G, Beecham GW, Wang LS, Vardarajan BN, Buross J et al. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat Genet* 2011; 43(5): 436–441. [PubMed: 21460841]
10. Bellenguez C, Kucukali F, Jansen IE, Klei L, Moreno-Grau S, Amin N et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat Genet* 2022.
11. Jun GR, Chung J, Mez J, Barber R, Beecham GW, Bennett DA et al. Transethnic genome-wide scan identifies novel Alzheimer's disease loci. *Alzheimers Dement* 2017; 13(7): 727–738. [PubMed: 28183528]
12. Kunkle BW, Schmidt M, Klein HU, Naj AC, Hamilton-Nelson KL, Larson EB et al. Novel Alzheimer Disease Risk Loci and Pathways in African American Individuals Using the African Genome Resources Panel: A Meta-analysis. *JAMA Neurol* 2021; 78(1): 102–113. [PubMed: 33074286]
13. Logue MW, Schu M, Vardarajan BN, Buross J, Green RC, Go RC et al. A comprehensive genetic association study of Alzheimer disease in African Americans. *Arch Neurol* 2011; 68(12): 1569–1579. [PubMed: 22159054]
14. Mez J, Chung J, Jun G, Kriegl J, Bourlas AP, Sherva R et al. Two novel loci, COBL and SLC10A2, for Alzheimer's disease in African Americans. *Alzheimers Dement* 2017; 13(2): 119–129. [PubMed: 27770636]
15. Reitz C, Jun G, Naj A, Rajbhandary R, Vardarajan BN, Wang LS et al. Variants in the ATP-binding cassette transporter (ABCA7), apolipoprotein E 4, and the risk of late-onset Alzheimer disease in African Americans. *Jama* 2013; 309(14): 1483–1492. [PubMed: 23571587]
16. Logue MW, Lancour D, Farrell J, Simkina I, Fallin MD, Lunetta KL et al. Targeted Sequencing of Alzheimer Disease Genes in African Americans Implicates Novel Risk Variants. *Front Neurosci* 2018; 12: 592. [PubMed: 30210277]
17. Ikezu T, Chen C, DeLeo AM, Zeldich E, Fallin MD, Kanaan NM et al. Tau Phosphorylation is Impacted by Rare AKAP9 Mutations Associated with Alzheimer Disease in African Americans. *Journal of Neuroimmune Pharmacology* 2018; 13(2): 254–264. [PubMed: 29516269]
18. Logue MW, Schu M, Vardarajan BN, Farrell J, Bennett DA, Buxbaum JD et al. Two rare AKAP9 variants are associated with Alzheimer's disease in African Americans. *Alzheimer's & Dementia* 2014; 10(6): 609–618.e611.
19. Jansen IE, Savage JE, Watanabe K, Bryois J, Williams DM, Steinberg S et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat Genet* 2019; 51(3): 404–413. [PubMed: 30617256]
20. Fang H, Hui Q, Lynch J, Honerlaw J, Assimes TL, Huang J et al. Harmonizing Genetic Ancestry and Self-identified Race/Ethnicity in Genome-wide Association Studies. *Am J Hum Genet* 2019; 105(4): 763–772. [PubMed: 31564439]
21. Cho K, Gagnon DR, Driver JA, Altincatal A, Kosik N, Lanes S et al. Dementia Coding, Workup, and Treatment in the VA New England Healthcare System. *Int J Alzheimers Dis* 2014; 2014: 821894.
22. Jaakkimainen RL, Bronskill SE, Tierney MC, Herrmann N, Green D, Young J et al. Identification of Physician-Diagnosed Alzheimer's Disease and Related Dementias in Population-Based Administrative Data: A Validation Study Using Family Physicians' Electronic Medical Records. *J Alzheimers Dis* 2016; 54(1): 337–349. [PubMed: 27567819]
23. Hunter-Zinck H, Shi Y, Li M, Gorman BR, Ji SG, Sun N et al. Genotyping Array Design and Data Quality Control in the Million Veteran Program. *Am J Hum Genet*; 106(4): 535–548. [PubMed: 32243820]
24. Abraham G, Qiu Y, Inouye M. FlashPCA2: principal component analysis of Biobank-scale genotype datasets. *Bioinformatics* 2017; 33(17): 2776–2778. [PubMed: 28475694]
25. Wang X. Firth logistic regression for rare variant association tests. *Front Genet* 2014; 5: 187. [PubMed: 24995013]
26. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010; 26(17): 2190–2191. [PubMed: 20616382]
27. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* 2017; 8(1): 1826. [PubMed: 29184056]

28. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* 2015; 11(4): e1004219.
29. Panitch R, Hu J, Chung J, Zhu C, Meng G, Xia W et al. Integrative brain transcriptome analysis links complement component 4 and HSPA2 to the APOE ε2 protective effect in Alzheimer disease. *Molecular Psychiatry* 2021.
30. de Klein N, Tsai EA, Vochteloo M, Baird D, Huang Y, Chen C-Y et al. Brain expression quantitative trait locus and network analysis reveals downstream effects and putative drivers for brain-related diseases. *bioRxiv* 2021: 2021.2003.2001.433439.
31. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM et al. A global reference for human genetic variation. *Nature* 2015; 526(7571): 68–74. [PubMed: 26432245]
32. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 2009; 19(9): 1655–1664. [PubMed: 19648217]
33. Rogava E, Meng Y, Lee JH, Gu Y, Kawarai T, Zou F et al. The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease. *Nat Genet* 2007; 39(2): 168–177. [PubMed: 17220890]
34. Cukier HN, Kunkle BW, Vardarajan BN, Rolati S, Hamilton-Nelson KL, Kohli MA et al. ABCA7 frameshift deletion associated with Alzheimer disease in African Americans. *Neurol Genet* 2016; 2(3): e79. [PubMed: 27231719]
35. Jin SC, Carrasquillo MM, Benitez BA, Skorupa T, Carrell D, Patel D et al. TREM2 is associated with increased risk for Alzheimer's disease in African Americans. *Molecular neurodegeneration* 2015; 10: 19. [PubMed: 25886450]
36. Mills MC, Rahal C. A scientometric review of genome-wide association studies. *Commun Biol* 2019; 2: 9. [PubMed: 30623105]
37. Sirugo G, Williams SM, Tishkoff SA. The Missing Diversity in Human Genetic Studies. *Cell* 2019; 177(4): 1080. [PubMed: 31051100]
38. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet* 2019; 51(4): 584–591. [PubMed: 30926966]
39. Farrer LA, Cupples LA, Haines JL, Hyman B, Kukull WA, Mayeux R et al. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. *Jama* 1997; 278(16): 1349–1356. [PubMed: 9343467]
40. Choi YY, Lee JJ, Choi KY, Seo EH, Choo IH, Kim H et al. The Aging Slopes of Brain Structures Vary by Ethnicity and Sex: Evidence From a Large Magnetic Resonance Imaging Dataset From a Single Scanner of Cognitively Healthy Elderly People in Korea. *Front Aging Neurosci* 2020; 12: 233. [PubMed: 32903525]
41. Kidd T, Brose K, Mitchell KJ, Fetter RD, Tessier-Lavigne M, Goodman CS et al. Roundabout controls axon crossing of the CNS midline and defines a novel subfamily of evolutionarily conserved guidance receptors. *Cell* 1998; 92(2): 205–215. [PubMed: 9458045]
42. Andrews WD, Barber M, Parnavelas JG. Slit-Robo interactions during cortical development. *J Anat* 2007; 211(2): 188–198. [PubMed: 17553100]
43. Sasaki T, Komatsu Y, Yamamori T. Expression patterns of SLIT/ROBO mRNAs reveal a characteristic feature in the entorhinal-hippocampal area of macaque monkeys. *BMC Research Notes* 2020; 13(1): 262. [PubMed: 32460877]
44. Wang B, Li H, Mutlu SA, Bowser DA, Moore MJ, Wang MC et al. The Amyloid Precursor Protein Is a Conserved Receptor for Slit to Mediate Axon Guidance. *eNeuro* 2017; 4(3).
45. Bai G, Chivatakarn O, Bonanomi D, Lettieri K, Franco L, Xia C et al. Presenilin-Dependent Receptor Processing Is Required for Axon Guidance. *Cell* 2011; 144(1): 106–118. [PubMed: 21215373]
46. Prokopenko D, Morgan SL, Mullin K, Hofmann O, Chapman B, Kirchner R et al. Whole-genome sequencing reveals new Alzheimer's disease-associated rare variants in loci related to synaptic function and neuronal development. *Alzheimers Dement* 2021; 17(9): 1509–1527. [PubMed: 33797837]

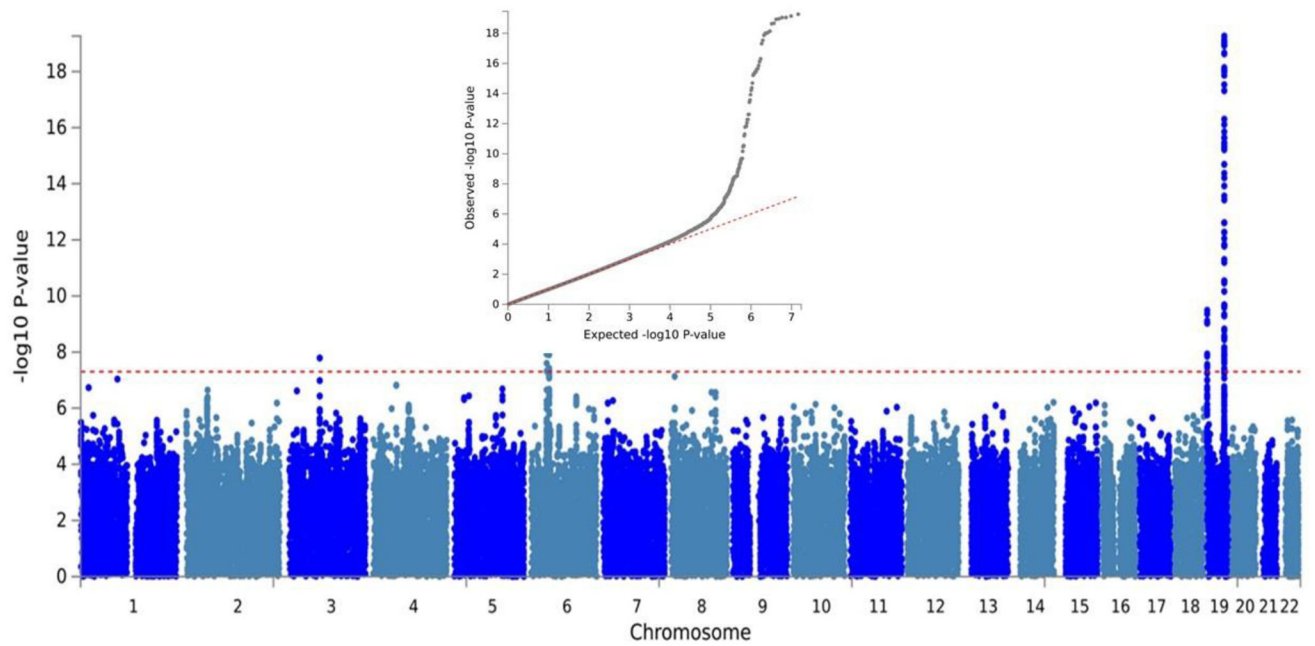
47. Moskovitz J, Bar-Noy S, Williams WM, Requena J, Berlett BS, Stadtman ER. Methionine sulfoxide reductase (MsrA) is a regulator of antioxidant defense and lifespan in mammals. *Proc Natl Acad Sci U S A* 2001; 98(23): 12920–12925. [PubMed: 11606777]
48. Kwon YJ, Kim JO, Park JM, Choi JE, Park DH, Song Y et al. Identification of Genetic Factors Underlying the Association between Sodium Intake Habits and Hypertension Risk. *Nutrients* 2020; 12(9).
49. Kuhn K, Bertling WM, Emmrich F. Cloning of a functional cDNA for human cytidine deaminase (CDD) and its use as a marker of monocyte/macrophage differentiation. *Biochem Biophys Res Commun* 1993; 190(1): 1–7. [PubMed: 8422236]
50. Gray EJ, Petsalaki E, James DA, Bagshaw RD, Stacey MM, Rocks O et al. Src homology 2 domain containing protein 5 (SH2D5) binds the breakpoint cluster region protein, BCR, and regulates levels of Rac1-GTP. *J Biol Chem* 2014; 289(51): 35397–35408. [PubMed: 25331951]
51. Joubert PE, Meiffren G, Gregoire IP, Pontini G, Richetta C, Flacher M et al. Autophagy induction by the pathogen receptor CD46. *Cell Host Microbe* 2009; 6(4): 354–366. [PubMed: 19837375]
52. Pickford F, Masliah E, Britschgi M, Lucin K, Narasimhan R, Jaeger PA et al. The autophagy-related protein beclin 1 shows reduced expression in early Alzheimer disease and regulates amyloid beta accumulation in mice. *J Clin Invest* 2008; 118(6): 2190–2199. [PubMed: 18497889]
53. Cardone J, Le Fric G, Kemper C. CD46 in innate and adaptive immunity: an update. *Clin Exp Immunol* 2011; 164(3): 301–311. [PubMed: 21488871]
54. Weyand NJ, Calton CM, Higashi DL, Kanack KJ, So M. Presenilin/gamma-secretase cleaves CD46 in response to *Neisseria* infection. *J Immunol* 2010; 184(2): 694–701. [PubMed: 20018629]
55. Heneka MT, Kummer MP, Stutz A, Delekate A, Schwartz S, Vieira-Saecker A et al. NLRP3 is activated in Alzheimer's disease and contributes to pathology in APP/PS1 mice. *Nature* 2013; 493(7434): 674–678. [PubMed: 23254930]
56. Jounai N, Kobiyama K, Shiina M, Ogata K, Ishii KJ, Takeshita F. NLRP4 negatively regulates autophagic processes through an association with beclin1. *J Immunol* 2011; 186(3): 1646–1655. [PubMed: 21209283]
57. Nelson KM. The burden of obesity among a national probability sample of veterans. *J Gen Intern Med* 2006; 21(9): 915–919 [PubMed: 16918734]
58. Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *Lancet (London, England)* 2020; 396(10248): 413–446. [PubMed: 32738937]





**Figure 1.** Manhattan and quantile-quantile plots for the meta-analysis of the Alzheimer's disease and related dementias and proxy dementia genome-wide association studies in Million Veteran Program African Americans.

The red dashed line on the Manhattan plot represents the genome-wide significance threshold ( $5.0 \times 10^{-8}$ ). Results from the *APOE* region on chromosome 19 are truncated at  $1.0 \times 10^{-20}$ .



**Figure 2.**

Manhattan and quantile-quantile plots for the meta-analysis of the Alzheimer's disease and related dementias and proxy dementia genome-wide association studies in Million Veteran Program African Americans and the Alzheimer's disease genome-wide association study in Alzheimer's Disease Genetics Consortium African Americans.

The red dashed line on the Manhattan plot represents the genome-wide significance threshold ( $5.0 \times 10^{-8}$ ). Results from the *APOE* region on chromosome 19 are truncated at  $1.0 \times 10^{-20}$ .

Common SNP associations ( $p < 5.0 \times 10^{-7}$ ) in the MVP ADRD+proxy dementia meta-analysis with corresponding results for AD in MVP and the ADGC

**Table 1:**

SNP	Chr	Pos	Gene	AI/A2	MVP Meta Z	MVP Meta P	ADGC B	ADGC P	MVP AD B	MVP AD P	MAF	r <sup>2</sup>
rs10197243	2	55052493	EML6	T/C	5.03	4.96E-07	0.09	1.60E-01	0.02	7.69E-01	0.15	0.89
rs11919682	3	78770102	ROBO1	C/G	-5.65	1.63E-08	NA	NA	-0.11	1.05E-01	0.27	0.85
rs148433063	4	59886649	RP11-340A13.2	T/C	-5.75	8.65E-09	0.02	9.04E-01	-0.64	2.99E-05	0.02	0.92
rs28377689	4	80967671	ANTXR2	T/G	-5.04	4.59E-07	0.04	4.22E-01	-0.15	1.91E-02	0.27	0.84
rs74852218	5	25701222	RNU6-374P	A/G	5.36	8.20E-08	0.06	7.53E-01	0.69	2.48E-04	0.01	0.86
rs2234253	6	41129105	TREM2	T/G	5.18	2.23E-07	NA	NA	0.23	2.05E-03	0.13	1.00
rs16894668	6	41388910	AL136967.1	T/C	5.13	2.91E-07	-0.01	9.18E-01	0.46	1.99E-03	0.02	0.99
rs7738720	6	47395399	AL355353.1	T/C	-5.43	5.48E-08	-0.20	5.51E-03	-0.16	1.24E-01	0.10	0.76
rs112395375	8	105698769	RP11-127H5.1	A/G	5.29	1.22E-07	0.11	4.86E-01	0.76	1.81E-02	0.01	0.80
rs509334	11	121277918	RP11-142I2.1-SORL1	A/G	5.24	1.59E-07	0.01	8.33E-01	0.16	3.54E-02	0.02	0.79
rs116620371	18	34179992	FHOD3	A/G	-5.36	8.39E-08	0.11	6.07E-01	-0.60	9.28E-02	0.01	0.92
rs73505251	19	1068095	ABCA7	A/T	5.20	2.03E-07	0.22	1.23E-04	0.17	2.19E-02	0.13	0.98

MAF and r<sup>2</sup> represent the minor allele frequency and imputation quality in the MVP AFR cohort. AI/A2=effect/non-effect allele

Common SNP associations ( $p < 5.0 \times 10^{-7}$ ) in the MVP ADRD+proxy dementia+ADGC AD meta-analysis

Table 2:

SNP	Chr	Pos	AI/A2	Nearest Gene(s)	ADRD OR	ADRD P	Proxy P	MVP P	ADGC P	All P	Dir	MAF	r2
rs116329346	2	38743886	A/G	AC016995.3	1.66	7.93E-04	9.13E-05	2.57E-07	6.97E-01	4.46E-06	++	0.01	0.77
rs58443395	2	53482134	A/G	AC010967.2-SCARNA16	1.13	1.35E-03	1.14E-03	5.39E-06	2.99E-02	4.75E-07	---	0.12	0.94
rs10197243	2	55052493	T/C	EML6	1.09	1.88E-02	4.79E-06	4.96E-07	1.60E-01	2.27E-07	+++	0.15	0.89
rs567572378	3	19633465	A/C	KCNH8-EFHB	1.89	2.21E-04	2.63E-04	2.42E-07	NA	2.42E-07	++?	0.01	0.75
rs11919682	3	78770102	C/G	ROBO1	0.86	6.70E-07	1.23E-03	1.63E-08	NA	1.63E-08	---?	0.27	0.85
rs192764155	4	59794471	A/G	RPI1-506N2.1	1.27	4.86E-03	7.43E-06	1.51E-07	NA	1.51E-07	---?	0.02	0.88
rs148433063	4	59886649	T/C	RPI1-340A13.2	1.28	3.20E-03	5.07E-07	8.65E-09	9.04E-01	1.54E-07	---	0.02	0.92
rs28377689	4	80967671	T/G	ANTXR2	1.10	1.63E-03	8.24E-05	4.59E-07	4.22E-01	1.53E-05	---	0.27	0.84
rs74852218	5	25701222	A/G	RNU6-374P	1.44	4.44E-04	5.01E-05	8.20E-08	7.53E-01	4.26E-07	+++	0.01	0.86
rs114681435	5	38361541	T/G	EGFLAM	1.37	2.12E-04	1.16E-03	1.17E-06	1.16E-01	3.64E-07	---	0.03	0.72
rs2234253	6	41129105	T/G	TREM2	1.20	6.05E-07	9.03E-03	2.23E-07	NA	2.23E-07	++?	0.13	1.00
rs73427293	6	41136611	A/T	TREM2-TREM2	1.19	2.70E-06	5.33E-03	2.95E-07	1.89E-03	2.95E-09	---	0.13	0.98
rs16894668	6	41388910	T/C	ALI36967.1-RPI1-328M4.2	1.39	1.05E-05	2.51E-03	2.91E-07	9.18E-01	2.89E-06	++	0.02	0.99
rs7738720	6	47395399	T/C	RPI1-157D6.1-CD2AP	0.83	1.35E-04	9.62E-05	5.48E-08	5.51E-03	1.14E-09	---	0.10	0.76
rs4607615	8	10279626	C/G	MSRA	1.13	4.76E-05	3.06E-03	1.13E-06	2.11E-02	7.37E-08	---	0.24	0.95
rs73581622	8	31258426	A/G	RPI1-566H8.3-RPI1-566H8.1	1.58	2.24E-03	3.73E-05	2.94E-07	6.98E-01	4.98E-06	++	0.01	0.86
rs112395375	8	105698769	A/G	RPI1-127H5.1	0.50	5.94E-07	5.91E-03	1.22E-07	4.86E-01	2.73E-07	+++	0.01	0.80
rs76427927	8	116123287	A/C	RPI1-192P9.1-TRPS1	1.10	8.30E-02	1.76E-05	9.81E-06	6.23E-03	2.76E-07	---	0.06	0.92
rs509334	11	121277918	A/G	BRP11-142I.1-SORL1	0.85	3.90E-06	2.73E-03	1.59E-07	8.33E-01	9.33E-07	+++	0.20	0.79
rs145008711	15	68726103	A/T	ITGAI1	1.58	3.03E-03	4.67E-05	4.94E-07	8.97E-01	4.75E-06	---	0.01	0.76
rs116620371	18	34179992	A/G	FHOD3	0.65	3.54E-03	5.71E-06	8.39E-08	6.07E-01	2.26E-06	---	0.01	0.92
rs73505251	19	1068095	A/T	ABC47	1.13	8.14E-04	7.03E-05	2.03E-07	1.23E-04	3.26E-10	+++	0.13	0.98
rs429358	19	45411941	T/C	APOE	1.58	2.90E-58	2.56E-47	2.48E-101	2.59E-60	6.09E-150	---	0.22	0.99

**Table3:**

Differential gene expression results for top genes in a meta-analysis of three European ancestry cohorts. P-values > 2.0E-04 not significant after multiple testing correction

Gene	Chr	Meta P <sub>expression</sub>
<i>CDA</i>	1	1.60E-08
<i>SH2D5</i>	1	6.70E-07
<i>DCBLD1</i>	6	3.20E-06
<i>EML6</i>	2	6.60E-06
<i>GOPC</i>	6	7.50E-06
<i>ABCA7</i>	19	7.20E-05
<i>ROS1</i>	6	9.40E-05
<i>TMCO4</i>	1	1.30E-04
<i>TREM2</i>	6	2.00E-04
<i>NUS1</i>	6	2.20E-03
<i>EIF4G3</i>	1	5.30E-03
<i>ARHGAP45</i>	19	1.40E-02
<i>CD2AP</i>	6	1.70E-02
<i>CNN2</i>	19	2.80E-02
<i>POLR2E</i>	19	3.00E-02
<i>RP11-127H5.1</i>	8	3.10E-02
<i>KIF17</i>	1	3.40E-02
<i>TREML1</i>	6	3.90E-02