Biobank-scale characterization of Alzheimer's disease and related dementias identifies potential disease-causing variants, risk factors, and genetic modifiers across diverse ancestries

Marzieh Khani[1], Fulya Akçimen[2*], Spencer M. Grant[1*], S. Can Akerman[3, 4], Paul Suhwan Lee[1], Faraz Faghri[1,5], Hampton Leonard[1,5] , Jonggeol Jeffrey Kim[1], Mary B. Makarious[1,5], Mathew J. Koretsky[1,5], Jeffrey D Rothstein[3,4], Cornelis Blauwendraat[1,2,6], Mike A. Nalls[1,5], Andrew Singleton[1], Sara Bandres-Ciga[1,#]

Affiliations:

[1] Center for Alzheimer's and Related Dementias (CARD), National Institute on Aging and National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD 20892, USA

[2] Molecular Genetics Section, Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD 20892, USA

[3] Brain Science Institute, Johns Hopkins University School of Medicine, Baltimore, MD, 21205, USA

[4] Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

[5] Data Tecnica LLC, Washington, DC 20037, USA

[6] Aligning Science Across Parkinson's (ASAP) Collaborative Research Network, Chevy Chase, MD, USA

*These two authors contributed equally

#Corresponding author
Sara Bandres-Ciga
Center for Alzheimer's Disease and Related Dementias (CARD)
National Institute on Aging (NIA), National Institutes of Health (NIH)
9000 Rockville Pike, NIH Building T44
Bethesda, MD 20892.
Email: sara.bandresciga@nih.gov
Phone: +1 (301) 841-5095

Running head: Genetic characterization of ADRD across diverse ancestries

Abstract

Alzheimer's disease and related dementias (AD/ADRDs) pose a significant global public health challenge, underscored by the intricate interplay of genetic and environmental factors that differ across ancestries. To effectively implement equitable, personalized therapeutic interventions on a global scale, it is essential to identify disease-causing mutations and genetic risk and resilience factors across diverse ancestral backgrounds. Exploring genetic-phenotypic correlations across the globe enhances the generalizability of research findings, contributing to a more inclusive and universal understanding of disease. This study leveraged biobank-scale data to conduct the largest multi-ancestry whole-genome sequencing characterization of AD/ADRDs. We aimed to build a valuable catalog of potential disease-causing, genetic risk and resilience variants impacting the etiology of these conditions. We thoroughly characterized genetic variants from key genes associated with AD/ADRDs across 11 genetic ancestries, utilizing data from All of Us, UK Biobank, 100,000 Genomes Project, Alzheimer's Disease Sequencing Project, and the Accelerating Medicines Partnership in Parkinson's Disease, including a total of 25,001 cases and 93,542 controls. We prioritized 116 variants possibly linked to disease, including 18 known pathogenic and 98 novel variants. We detected previously described disease-causing variants among controls, leading us to question their pathogenicity. Notably, we showed a higher frequency of *APOE* ε4/ε4 carriers among individuals of African and African Admixed ancestry compared to other ancestries, confirming ancestry-driven modulation of *APOE*-associated AD/ADRDs. A thorough assessment of *APOE* revealed a disease-modifying effect conferred by the *TOMM40*:rs11556505, *APOE*:rs449647, *19q13.31*:rs10423769, *NOCT*:rs13116075, *CASS4*:rs6024870, and *LRRC37A*:rs2732703 variants among *APOE* ε4 carriers across different ancestries. In summary, we compiled the most extensive catalog of established and novel genetic variants in known genes increasing risk or conferring resistance to AD/ADRDs across diverse ancestries, providing clinical insights into their genetic-phenotypic correlations. The findings from this investigation hold significant implications for potential clinical trials and therapeutic interventions on a global scale. Finally, we present an accessible and user-friendly platform for the AD/ADRDs research community to help inform and support basic, translational, and clinical research on these debilitating conditions (https://niacard.shinyapps.io/MAMBARD_browser/).

Keywords: Alzheimer's disease, dementia, genetics, target prioritization, clinical trials, genetic risk factors, disease-causing variants, protective variants, disease-modifying variants, All of Us, UK Biobank, 100,000 Genomes Project, ADSP, AMP PD

Introduction

In 2023, the World Health Organization reported that dementia affects approximately 55 million people worldwide [1]. This number is expected to reach approximately 152.8 million (ranging from 130.8 to 175.9 million) by 2050 [2], placing a significant burden on healthcare infrastructure. Alzheimer's disease (AD), the most common form of dementia, represents roughly 60-70% of all cases [1]. Less prevalent forms, such as dementia with Lewy bodies (DLB) and frontotemporal dementia (FTD), each account for 10–15% of dementia cases [3,4].

Most of the research conducted thus far on the genetic underpinnings of dementia has primarily focused on populations of European ancestry, limiting the generalizability of findings [5]. Growing evidence indicates significant differences in the genetic architecture of disease among diverse ancestral populations, which raises concerns about the development of therapeutic interventions based on genetic targets primarily identified in a single population. Expanding research to include diverse ancestries is crucial for precision therapeutics. In the new era of personalized medicine, achieving accurate and effective disease-modifying treatments requires a comprehensive understanding of these diseases in a global context.

In recent years, researchers and healthcare institutions worldwide have undertaken ambitious efforts to create large-scale datasets encompassing diverse genetic ancestries, providing valuable insights into the genetic, environmental, and clinical factors influencing disease susceptibility and progression [6,7]. While more work remains in collecting diverse genetic datasets that are dementia-specific, existing efforts can provide valuable insights into dementia research. Currently, All of Us (AoU), UK Biobank (UKB), 100,000 Genomes Project (100KGP), Alzheimer's Disease Sequencing Project (ADSP), and the Accelerating Medicines Partnership in Parkinson's Disease (AMP PD) represent the largest and most prominent publicly available dementia datasets worldwide.

A priority in elucidating the etiology of AD and related dementias (AD/ADRDs) lies in defining cumulative risk; however, very little is known about genetic factors that enhance resistance to or protect against dementia. In genetics, protective variants reduce the risk of developing dementia or delay its onset. They confer protection via a loss-of-function or gain-of-function mechanism and can influence various biological pathways associated with the disease. Resilience variants (disease-modifying factors reducing the penetrance of risk loci) influence the development and course of the disease in individuals already at risk, potentially delaying symptom onset or reducing disease severity by interacting with pre-existing risk variants (genetic modifiers). To the best of our knowledge, 11 protective and 10 resilience variants have been reported in AD, with a particular focus on the role of genetic variation modulating AD risk among homozygous or heterozygous *APOE* ε4 carriers [8–14]. Understanding factors that confer protection or resilience can inform therapeutic strategies to reduce the overall burden of dementia, potentially decreasing healthcare costs and the societal impact of the disease.

In this study, we aimed to conduct the largest and most comprehensive multi-ancestry whole-genome sequencing characterization of AD/ADRDs potential disease-causing variation, as well as risk, protective, and disease-modifying factors leveraging biobank-scale data. We screened genetic variants in key genes linked to these conditions, including *APP*, *PSEN1*, *PSEN2*, *TREM2*, *MAPT*, *GRN*, *GBA*, *SNCA*, and *APOE* across a total of 25,001 AD/ADRD cases and 93,542 control individuals, collectively representing 11 ancestries. Furthermore, we assessed protective and disease-modifying variants among different *APOE* genotype carriers in those ancestry groups. This research is particularly relevant in the context of population-specific target prioritization for therapeutic interventions. Such advancements are crucial, as drug mechanisms supported by genetic insights have a 2.6 times higher likelihood of success than those without such support, underscoring the importance of including diverse genetic data to enhance therapeutic outcomes [15]. Here, we present genetic-phenotypic correlations among identified variants across all datasets and develop a user-friendly platform for the scientific community to help inform and support basic, translational, and clinical research on these debilitating conditions (https://niacard.shinyapps.io/MAMBARD_browser/).

Methods

Demographic information, including age and sex, was provided in the self-reported survey in AoU and through the UKB, ADSP, and AMP PD portals. Self-reported demographic data of participants in 100KGP were obtained from Data Release V18 (12/21/2023) using the LabKey application incorporated into the research environment. **Figure 1** displays the demographic characteristics of cohorts under study. **Figure 2** shows a summary of our workflow, which we explain in further detail below.

Discovery phase: All of Us

The All of Us (AoU) Research Program (allofus.nih.gov) launched by the United States National Institutes of Health (NIH) endeavors to enhance precision health strategies by assembling rich longitudinal data from over one million diverse participants in the United States. The program emphasizes health equity by engaging underrepresented groups in biomedical research. This biobank includes a wide range of health information, including genetic, lifestyle data, and electronic health records (EHRs), among others, making it a valuable resource for studying the genetic and environmental determinants of various diseases, including AD/ADRDs [6,16].

Data Acquisition

We accessed the AoU data through the AoU researcher workbench cloud computing environment (https://workbench.researchallofus.org/), utilizing Python and R programming languages for querying. We used the online AoU data browser (https://databrowser.researchallofus.org/variants)

to extract genetic variants from short-read whole genome sequencing (WGS) data. The selected variants were filtered for protein-altering or splicing mechanisms for further analysis.

Cohort Creation

We generated WGS cohorts using the cohort-creating tool in the AoU Researcher Workbench. AD/ADRD cases were selected based on the condition domain in the EHRs. Controls were selected among individuals ≥ 65 years old without any neurological condition in their EHRs, family history, or neurological history in their self-reported surveys. In total, 539 AD cases, 1,655 related dementias, and 13,835 controls were included in the study.

Whole genome sequencing protocol and quality control assessment

WGS was conducted by the Genome Centers funded by the AoU Research Program [6,17], all of which followed the same protocols. Sequencing details are described elsewhere [18]. Phenotypic data, ancestry features, and principal components (PCs) were generated using Hail within the AoU Researcher Workbench (https://support.researchallofus.org/hc/en-us/articles/4614687617556-How-the-All-of-Us-Genomic-data-are-organized). Ancestry annotation and relatedness were determined using the PC-relate method in Hail and duplicated samples and one of each related participant pair with KINSHIP less than 0.1 being excluded from the Hail data [19] (https://support.researchallofus.org/hc/en-us/articles/4614687617556-How-the-All-of-Us-Genomic-data-are-organized). Flagged individuals and low-quality variants (qc.call_rate < 0.90) were removed from the analysis.

Variant Filtering and Analysis

We utilized protein-altering and splicing variants ('WGS_EXOME_SPLIT_HAIL_PATH') for our analysis, obtained following the tutorial 'How to Work with AoU Genomic Data (Hail - Plink) (v7).' The largest intervals for genomic positions were obtained from the UCSC Genome Browser (https://genome.ucsc.edu/).

Variant datasets were obtained as described in the related workspace (see “How to Work with AoU Genomic Data (Hail - Plink) (v7)” for further details). Genomic positions (GRCh38) for each gene were extracted from the Hail variant dataset. Variant-level quality assessments were applied as described in the Manipulate Hail Variant Dataset tutorial (see the “How to Work with AoU Genomic Data” workspace for further details). VCF files containing the cohorts in the current study were generated using BCFtools v1.12 [20]. Allele frequency and zygosity of each resulting variant were calculated per ancestry using PLINK v2.0 [21] in each of the AD, related dementias, and control cohorts.

Discovery phase: UK Biobank

The UK Biobank (UKB) (https://www.ukbiobank.ac.uk/) is a large-scale biomedical dataset containing detailed genetic, clinical, and lifestyle information from over 500,000 participants aged 40 to 69 years in the United Kingdom. Each participant's profile includes a diverse array of phenotypic and health-related information. Additionally, the participants' health has been followed long-term, primarily through linkage to a wide range of health-related records, enabling the validation and characterization of health-related outcomes [7]. This dataset has been instrumental in advancing research on various health conditions, including AD/ADRDs, by facilitating large-scale genome-wide association studies and rare, deleterious variant analyses [7].

Cohort Creation

We accessed UKB data (https://www.ukbiobank.ac.uk/) through the DNAnexus cloud computing environment, utilizing the Python programming language for querying. Three experimental cohorts were defined: AD, related dementias, and controls. The AD cohort was defined by the UKB field ID 42020, using diagnoses according to the UKB's algorithmically defined outcomes v2.0 (https://biobank.ndph.ox.ac.uk/ukb/refer.cgi?id=460). The related dementia cohort was defined by the UKB field ID 42018, using the UKB's algorithmically defined "Dementia" classification, with the added step of excluding any individuals in the aforementioned AD cohort. The control cohort includes individuals ≥ 65 years without any neurological condition or family history of neurological disorders. Relatedness was calculated with KING [22], and individuals closer than cousins were removed by KINSHIP > 0.0884 to ensure no pair of participants across all three cohorts were related. In total, 4,225 AD cases, 5,306 related dementias, and 56,741 controls were included in the study.

Whole genome sequencing protocol and quality control assessment

Sequencing was conducted using the NovaSeq 6000 platform [23]. These data were then analyzed with the DRAGEN v3.7.8 (Illumina, San Diego, CA, USA) software. Alignment was performed against the GRCh38 reference genome. Further details on quality control metrics can be found at https://biobank.ndph.ox.ac.uk/showcase/label.cgi?id=187.

Data Acquisition

WGS data are stored in the UKB as multi-sample aggregated pVCF files, each representing distinct 20 kbp segments for all participants. Genomic ranges were defined for each gene of interest using Ensembl (https://useast.ensembl.org/index.html). Those pVCF files containing any variants within these genomic ranges were included for analysis. Left alignment and normalization were performed on each of these variants using BCFtools v1.15.1 [24]. Then, ANNOVAR [25] was used to annotate the normalized variants.

Variant Filtering and Analysis

We filtered variants to include only those within our genes of interest, annotated as either protein-altering or splicing variants and present in any AD and/or related dementia cases. Allele frequency and zygosity of each resulting variant were calculated per ancestry using PLINK v2.0 [21] in each of the AD, related dementias, and control cohorts.

Discovery phase: 100,000 Genomes Project

The 100,000 Genomes Project (100KGP) (https://www.genomicsengland.co.uk/) has sequenced and analyzed genomes from over 75,000 participants with rare diseases and family members. Early onset dementia (encompassing FTD) is one of the rare diseases studied by the Neurology and Neurodevelopmental Disorders group within the rare disease domain. Participants were recruited by healthcare professionals and researchers from 13 Genomic Medicine Centres in England. The probands were enrolled in the project if they or their guardian provided written consent for their samples and data to be used in research. Probands and, if feasible, other family members were enrolled according to eligibility criteria set for certain rare disease conditions.

WGS data were utilized from 180 unrelated cases with early-onset dementia (encompassing FTD and prion disease) or Parkinson's disease (PD) with dementia phenotype and 3,479 unrelated controls ≥ 65 years at the time of the analysis. Sequencing and quality control analyses for the 100KGP were previously described elsewhere [26] (https://re-docs.genomicsengland.co.uk/sample_qc/). Protein-altering or splicing variants were obtained using the Exomiser variant prioritization application [27]. Candidate variants were extracted from a multi-sample aggregated VCF provided in the Genomics England research environment.

Replication phase: Alzheimer's Disease Sequencing Project

The Alzheimer's Disease Sequencing Project (ADSP) (https://adsp.niagads.org/), supported by the National Institute on Aging and the National Human Genome Research Institute, aims to generate data associated with AD/ADRDs. This dataset includes genetic data from thousands of individuals with and without AD, facilitating the discovery of novel genetic risk factors and pathways underlying the disease.

We used data from the ADSP dataset (v4) for this study, which included a total of 10,566 AD cases and 16,217 controls. The control cohort includes individuals ≥ 65 years without any neurological condition or family history of neurological disorders. Samples were excluded from further analysis if the sample call rate was less than 95%, the genetically determined sex did not match the sex reported in clinical data, or excess heterozygosity was detected (|F| statistics > 0.25). For quality control purposes, an MAF threshold of 0.1% was used. The missingness rate and allele

frequency of these variants were calculated for each ancestry using PLINK v2.0 [21] and PLINK v1.9 [28]. Variant quality control included removing variants with Hardy-Weinberg Equilibrium $P < 1 \times 10^{-4}$ in control samples, differential missingness by case-control status at $P \leq 1 \times 10^{-4}$, and non-random missingness by haplotype at $P \leq 1 \times 10^{-4}$. Relatedness was calculated with KING [22], and individuals closer than cousins were removed by KINSHIP > 0.0884. Duplicated samples were also removed. ADSP includes a range of cohorts, including extensive family cohorts and cohorts with progressive supranuclear palsy (PSP), corticobasal degeneration, mild cognitive impairment (MCI), and DLB patients. Only samples labeled as definite AD or control were included in this analysis. We meticulously screened for identified genetic variants with a CADD score > 20 that were present across any of the three discovery datasets (AoU, 100KGP, and UKB) in the ADSP cohort.

Replication phase: Accelerating Medicines Partnership in Parkinson's Disease

Accelerating Medicines Partnership (AMP) (https://fnih.org/our-programs/accelerating-medicines-partnership-amp/) is a public-private initiative that aims to transform the current model for developing new diagnostics and treatments by jointly identifying and validating promising biological targets for therapeutics. It was launched in 2014 by the NIH, the U.S. Food and Drug Administration, multiple biopharmaceutical and life science companies, and several non-profit organizations. AMP PD focuses on advancing research into PD-related disorders and leverages cutting-edge technologies and large-scale data analysis to identify key genetic variants, biomarkers, and therapeutic targets associated with PD-related disorders, with the ultimate goal of developing novel treatments and improving patient outcomes.

We used AMP PD Release 3 genomic data, focusing specifically on DLB cases and controls. Samples were excluded from further analysis if the sample call rate was less than 95%, the genetically determined sex did not match the sex reported in clinical data, or excess heterozygosity was detected ($|F|$ statistics > 0.25). Variant quality control included removing variants with missingness above 0.05%. Relatedness was calculated with KING [22], and individuals closer than first cousins were removed by KINSHIP > 0.0884. After quality control and ancestry prediction, this dataset contains a total of 2,530 DLB cases and 3,270 controls, characterized as individuals ≥ 65 years without any neurological condition or family history of any neurological disorders. We screened for identified variants with a CADD score > 20 that were present across all three discovery datasets (AoU, 100KGP, and UKB) within AMP PD. Allele frequency of these variants per ancestry was calculated using PLINK v2.0 [21] and PLINK v1.9 [28].

Ancestry Prediction Analysis

All samples in AoU, UKB, ADSP, and AMP PD datasets underwent a custom ancestry prediction pipeline included in the GenoTools package (https://github.com/dvitale199/GenoTools) [29]. In brief, ancestry was defined using reference panels from the 1000 Genomes Project, the Human

Genome Diversity Project, and an Ashkenazi Jewish population dataset. We used a panel of 4,008 samples from 1000 Genomes Project and the Gene Expression Omnibus database (www.ncbi.nlm.nih.gov/geo; accession no. GSE23636) to define ancestry reference populations. The reference panel was then reduced to exclude palindromic SNPs (AT or TA or GC or CG). SNPs with minor allele frequency (MAF) < 0.05, genotyping call rate < 0.99, and HWE P < 1E-4 in the reference panel were further excluded. Variants overlapping between the reference panel SNP set and the samples of interest were then extracted. Any missing genotypes were imputed using the mean of that particular variant in the reference panel. The reference panel samples were split into an 80/20 train/test set, and then PCs were fitted using the set of overlapping SNPs described previously. The PCs were then transformed via UMAP to represent global genetic population substructure and stochastic variation. A classifier was then trained on these UMAP transformations of the PCs (linear support vector). Based on the test data from the reference panel and at 5-fold cross-validation, 11 ancestries were predicted consistently with balanced accuracies greater than 0.95.

Genetic ancestry in 100KGP was estimated by generating PCs for 1000 Genomes Project phase 3 samples and projecting all participants onto the super populations in the 1000 Genomes Project, as described elsewhere (https://re-docs.genomicsengland.co.uk/ancestry_inference/). Despite our efforts to utilize GenoTools, we encountered significant challenges during its implementation in Genomics England's High-Performance Computing Cluster (HPC). Consequently, GenoTools and Genomics England's HPC were incompatible in this context. PCA plots across all biobanks are shown in **Supplementary Figure 1**.

Evaluation of potential disease-causing mutations, risk factors, and disease risk modifiers across ancestries

In the discovery phase, variants were filtered out based on their presence in control individuals across biobanks. To prioritize potential disease-causing mutations, we followed the American College of Medical Genetics and Genomics (ACMG) guidelines (https://wintervar.wglab.org/), leveraging existing clinical and population databases and pathogenicity predictors including the Human Gene Mutation Database (HGMD) (https://www.hgmd.cf.ac.uk/ac/index.php), dbSNP (https://www.ncbi.nlm.nih.gov/snp/), gnomAD (https://gnomad.broadinstitute.org/), ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/), PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/), and Combined Annotation Dependent Depletion (CADD) scores (GRCh38-v1.7) (https://cadd.gs.washington.edu/).

Secondly, we investigated *APOE*, the major risk factor for AD/ADRDs, across diverse ancestries. We used PLINK (v1.9 and v2.0) [21,28] to extract genotypes for two *APOE* variants, rs429358 (chr19:44908684-44908685) and rs7412 (chr19:44908821-44908823), as a proxy for *APOE* allele status (ε1, ε2, ε3, and ε4) in the AoU, UKB, ADSP, and AMP PD datasets. Data analysis was conducted as reported elsewhere (https://github.com/neurogenetics/APOE_genotypes). In the

100KGP dataset, *APOE* genotypes were analyzed using PLINK v2.0 [21] in a multi-sample aggregated VCF provided in the Genomics England research environment. Subsequently, we calculated the number of individuals with each genotype per ancestry and their frequency percentages.

Finally, we assessed disease modifiers for *APOE* ε4 homozygous and heterozygous carriers specifically. A total of 21 variants, previously identified as either protective (n=11) or resilient (n=10), were extracted from all datasets using the same protocol previously described. Among them, *ABCA7*:rs72973581-A, *APP*:rs466433-G, *APP*:rs364048-C, *NOCT*:rs13116075-G, *SORL1*:rs11218343-C, *SLC24A4*:rs12881735-C, *CASS4*:rs6024870-A, *EPHA1*:rs11762262-A, *SPPL2A*:rs59685680-G, *APP*:rs63750847-T, *PLCG2*:rs72824905-G are protective, while *19q13.31*:rs10423769-A, *APOE*:rs449647-T, *FN1*:rs140926439-T, *FN1*:rs116558455-A, *RELN*:rs201731543-C, *TOMM40*:rs11556505-T, *RAB10*:rs142787485-G, *LRRC37A*:rs2732703-G, *NFIC*:rs9749589-A, and the *APOE3* Christchurch:rs121918393-A variant are reported to be resilient. These variants were then checked across all *APOE* genotypes and ancestries. Carrier frequencies (either heterozygous or homozygous) were calculated for each *APOE* genotype and ancestry and were then combined across each of the datasets. In AoU, a variant dataset in Hail format (WGS_VDS_PATH) was used for the analysis. PLINK v2.0 [21] and R v4.3.1 (https://www.r-project.org/) were used to assess the protective model (which evaluates the effect of each protective/disease-modifying variant on the phenotype), conditional model (which evaluates the effect of each protective/disease-modifying variant on the phenotype in the presence of *APOE* (ε4, ε4/ε4, ε3/ε3)), R^2 model (which evaluates the correlation of each protective/disease-modifying variant with *APOE* (ε4, ε4/ε4, ε3/ε3)), and interaction model (which evaluates putative interactions between each protective/disease-modifying variant and *APOE* (ε4, ε4/ε4, ε3/ε3) on the phenotype). Logistic and linear regression analyses, adjusting for *APOE* status, sex, age, and PCs, were applied in the most well-powered dataset (ADSP) to explore these effects.

Results

Large-scale genetic characterization nominates known and novel potential disease-causing variants associated with Alzheimer's disease and related dementias

A summary of the identified variants can be found in **Figure 3**. We identified a total of 159 variants in the *APP, PSEN1, PSEN2, TREM2, GRN, MAPT, GBA1*, and *SNCA* genes within the AoU dataset. All variants and their allele frequencies across different ancestries are available in **Supplementary Table 1**. Among these, 30 genetic variants were present only in cases and had a CADD score > 20 (CADD score > 20 means that the variant is among the top 1% most pathogenic in the genome, as a proxy for its deleteriousness). All 30 of these identified variants were heterozygous. Of these, four were previously reported in AD or FTD (**Table 1**), while 26 were novel (**Table 2**). Of the four known variants, two were found in cases of European ancestry, one of African ancestry, and one of American Admixed ancestry. Among the 26 novel variants, 18

were found in cases of European ancestry, three in cases of African ancestry, two in cases of American ancestry, one in a case of Ashkenazi Jewish ancestry, and two in cases of African Admixed ancestry.

Within the UKB, we identified a total of 650 variants in the *APP, PSEN1, PSEN2, TREM2, GRN, MAPT, GBA1,* and *SNCA* genes (**Supplementary Table 2**). Among these, 87 variants were present only in cases and had a CADD score > 20. All 87 identified variants were heterozygous. Of these, 16 were previously reported as disease-causing in AD, FTD, frontotemporal lobar degeneration (FTLD), and Gaucher disease (**Table 1**), while 71 were novel (**Table 2**). A majority (n = 82) of the variants were identified in individuals of European genetic ancestry, two in cases of African ancestry, and one each in cases of South Asian, East Asian, and Ashkenazi Jewish ancestries, respectively. The allele frequencies of the variants across different ancestries are reported in **Supplementary Table 2**.

We identified a total of 11 variants in the *APP*, *PSEN1*, *PSEN2*, *GRN*, and *GBA1* genes within the 100KGP data (**Supplementary Table 3**). Among cases, no variants were identified in the *MAPT*, *TREM2*, and *SNCA* genes. Of the 11 variants, three were only present in cases and had a CADD score > 20. All three identified variants were heterozygous and previously reported in individuals of European ancestry. Among these three variants, *PSEN1* p.R269H had been previously reported as a cause of AD **(Table 1)**, while the remaining two variants in the *APP* and *PSEN2* genes were novel (**Table 2)**. The allele frequency of each variant is presented in **Supplementary Table 3**.

Replication analyses support the relevance of identified genetic variation across diverse ancestries

Six variants identified in AoU, 16 identified in UKB, and three identified in 100KGP were replicated in AD cases in the ADSP cohort **(Table 3)**. Among the six variants found in AoU that were replicated in ADSP, two variants — *APP* p.A713T and *PSEN1* p.A79V — had been previously reported, while four variants — *APP* p.L597W, *MAPT* p.G701R, *MAPT* p.G750S, and *SNCA* p.Q99R — were novel, with *APP* p.L597W being found in African, African Admixed, and Complex Admixture History ancestries. Searching for other dementia cases resulted in the identification of *APP* p.A713T in one DLB case and *PSEN1* p.A79V in a possible AD case according to ADSP diagnosis criteria. The allele frequency of each variant per genetic ancestry in cases and controls is reported in **Table 3**.

Among the three variants identified in the 100KGP dataset that were replicated in the ADSP cohort, *PSEN1* p.R269H was previously reported while *PSEN2* p.D320N and *APP* p.Y407H were novel. We observed the previously reported *PSEN1* p.R269H variant in five cases and no control participants. This variant was found in European ancestry individuals from the 100KGP cohort and was also observed in individuals of African Admixed ancestry (two cases) and European

ancestry (three cases) in the ADSP dataset. Of the novel variants, the *PSEN2* p.D320N variant was found in five controls and was not observed in any cases, while the *APP* p.Y407H variant was observed in two cases and one control. Searching for additional cases led to the discovery of *PSEN1* p.R269H in a patient with MCI.

Among the 16 variants identified in the UKB cohort that were replicated in the ADSP dataset, five variants — *PSEN1* p.A79V (five cases and one control), *PSEN1* p.R269H (five cases), *GRN* p.R493X (three cases), *MAPT* p.R406W (four cases and one control), and *GBA1* p.W223R (one case and one control) — have been previously reported. The remaining 11 variants were novel. Most of the novel variants were found in European cases in the UKB. The *PSEN1* p.R269H variant was found in cases of both African Admixed and European ancestries, and *GBA1* p.W223R was found in a case of Complex Admixture History ancestry and a control of African ancestry. The three remaining known variants were observed in individuals of European ancestry in the ADSP cohort. Novel variants identified in non-European participants include *PSEN1* p.R220Q (one African case, two European cases, and one African Admixed control), *PSEN1* p.T291A (one African Admixed case, one American Admixed control, and two controls with Complex Admixture History), *MAPT* p.A60G (two African Admixed cases, one African Admixed control, two African controls, and four controls with Complex Admixture History), *GRN* p.V490M (one African Admixed case), *MAPT* p.R103W (one American Admixed control), and *APP* p.P251S (one South Asian control). Searching for other cases resulted in the identification of *PSEN1* p.A79V in one possible AD patient, *PSEN1* p.R269H and *MAPT* p.A60G in two independent MCI patients, and *GBA1* p.W223R and *APP* p.A209T in two independent PSP patients **(Table 3)**.

We identified a novel *SNCA* variant (p.Q99R) in the AoU dataset, while the UKB dataset revealed two additional variants in *SNCA*: p.P90H and p.A91S. Both p.P90H and p.A91S were predicted to be likely pathogenic according to prediction estimates and have not been previously reported as disease-causing. Notably, the *SNCA* p.Q99R variant was replicated in the ADSP cohort. All three variants were heterozygous, and none of these variants were found in any controls across these datasets. However, the age at onset of these variant carriers is not consistent with a potential disease-causing deleterious effect.

Our analyses of multiple datasets identified 11 variants (three from AoU, seven from UKB, and one from 100KGP) that were absent in the ADSP control cohort.

Across each of our discovery datasets, we identified five candidate variants — *APP* p.A713T, *MAPT* p.G750S, *GRN* p.V490M, *GRN* p.R493X, and *APP* p.D516N — present in AMP PD. *GRN* p.V490M was present in one control and no cases, *GRN* p.R493X was present in one case and no controls, while the other three were present across both cases and controls. The allele frequencies of these variants are detailed in **Table 4**.

Among the 116 variants identified in this study, 13 were found exclusively in non-European ancestries. Notably, *APP*:p.L597W and *MAPT*:p.A60G were replicated in African and African Admixed ancestries across different datasets. These data highlight the potential significance of these variants in groups that are often underrepresented in genomic studies.

Supplementary Figure 2 shows the allele frequencies of all identified known and novel variants with CADD > 20 in the discovery and replication phases across all ancestries in each biobank**.**

Previously reported disease-causing variants raise questions about potential pathogenicity

Although the *SNCA* p.H50Q variant was initially identified as a pathogenic mutation in PD [30], subsequent research has challenged its pathogenicity [31]. Our study confirms that it is not pathogenic across other synucleinopathies such as DLB, based on its occurrence in five European controls in AoU and 28 European controls in UKB.

Additionally, several research studies have reported the *APP* p.A713T variant to be disease-causing [32,33]. In our study, we found this variant in heterozygous state in five control individuals: two in UKB, two in ADSP, and one in AMP PD. Interestingly, the *APP* p.E665D variant, which has been widely reported to cause AD [34,35], was found in one control in AoU in her late 70s. However, it is possible that the variant shows incomplete penetrance, or that this individual may harbor unidentified resilient genetic variation. Another previous study evaluating the role of *APP* p.E665D questioned the pathogenicity of this variant [36].

GBA1 coding variants in heterozygous state generally exhibit incomplete penetrance and act as genetic risk factors. Homozygous *GBA1* variants, including the p.T75del and c.115+1G>A mutations have been reported to cause Gaucher disease [37–40]. We found these two variants in a heterozygous state in one case and one control in AoU. *GBA1* p.T75del was found in individuals of African ancestry, and *GBA1* c.115+1G>A was found in individuals of European ancestry in both a case and a control. The *GBA1* c.115+1G>A variant was also found in nine European controls in the UKB cohort. Thirteen additional heterozygous variants in *GBA1* — p.R502C, p.A495P, p.L483R, p.D448H, p.E427X, p.G416S, p.N409S, p.R398X, p.R296Q, p.G241R, p.N227S, p.S212X, and p.R159W — were identified in our study, and have been reported as disease-causing for Gaucher disease in homozygous state. Three variants in *GRN*, including two loss of function variants (including p.Q130fs and p.Y294X) and one splicing variant (c.708+6_708+9del), have been previously reported to cause FTD, FTLD, and neurodegenerative disease [37,41–55]. Each of these 16 variants were found in several control individuals **(Supplementary Table 4)**.

Genetic-phenotypic correlations provide valuable clinical insights

Clinical data for the identified variants are summarized in **Supplementary Tables 5 and 6**. Here, we briefly explain the main findings.

The *GRN* p.R493X variant is this gene's most reported pathogenic mutation. This variant has been associated with several types of dementia, including FTD, FTLD, primary progressive aphasia, AD, and corticobasal degeneration. It is known to be more frequently identified among FTD cases, particularly in early-onset forms [56]. In one study investigating the genetics underlying disease etiology in 1,118 DLB patients, this variant was reported in a single case, presenting a wide range of neurological phenotypes that could not lead to a conclusive diagnosis. Severe dementia, parkinsonism, and visual hallucinations suggested a clinical diagnosis of AD or mixed vascular dementia. However, the final neuropathological diagnosis was suggested to be AD, DLB, and argyrophilic grain disease [57]. We identified this variant in four European AD patients, three of whom presented with early onset in their fifties. Interestingly, we also identified this variant in a DLB patient in her early 60s. Neuropathological data and McKeith criteria [58] strongly supported a diagnosis of DLB in this patient. Although this variant has been widely reported across different types of dementia, our finding is the first report of this variant in DLB with a McKeith criteria of "high likelihood of DLB," expanding the etiological spectrum of *GRN* variation **(Supplementary Table 5)**.

GRN p.C222Y was previously reported in a familial AD case from Latin American (Caribbean Hispanic) ancestry [59,60]. While the AAO for this patient was not reported, the mean AAO for the cohorts under study was 56.9 years (SD = 7.29), with a range between 40–73 years. In our study, we identified this variant in an individual of American Admixed ancestry with dementia in his late 40s and a disease duration of 11 years to date. This finding reinforces the role of this variant in early-onset disease.

There are several other interesting findings regarding variants in *GRN*. The *GRN* c.708+1G>A variant was previously reported in several FTD, FTLD, and corticobasal syndrome (CBS) cases, mostly early-onset [55,61]. We identified this variant in two European AD cases, both diagnosed in their 70s, marking the first report of this variant in late-onset Alzheimer's disease (LOAD). The *GRN* p.P166fsX variant was previously reported in an early-onset behavioral variant FTD case [62]. In our study, we identified this variant in a European dementia case diagnosed in her mid 70s with a disease duration of 8 years to date. The *GRN* p.R418X variant is identified in the literature in two cases of FTLD with ubiquitin-positive inclusions (FTLD-U) with an AAO of 49 and 60 years [63]. We identified this variant in a European dementia case in her early 70s. Both findings represent the first report of these variants in late-onset dementia.

PSEN1 R269H is a known pathogenic variant causing early-onset Alzheimer's disease (EOAD) [64,65]. However, it has been previously reported in only two LOAD cases [66,67]. In our study,

we identified this variant in European and African Admixed ancestries in a total of 12 cases (eight AD and four related dementias), six of which were early-onset (≤65 years) and six were late-onset (>65 years). This finding underscores the potential for *PSEN1* p.R269H to contribute to LOAD with reduced penetrance. Additionally, one EOAD case that presented with hallucinations [68] and another that manifested a behavioral presentation [69] have been reported to carry this variant. In this study, we identified *PSEN1* p.R269H in one FTD patient in the 100KGP cohort, marking the first report of this variant in FTD.

MAPT p.R406W has been reported in several familial cases of FTD with parkinsonism, all with early onset [70]. There are only two articles related to this variant in AD. The first describes a family with AD-like symptoms, with an average AAO of 61 years [71], and the other reports a familial AD case with an AAO of 50 years [72]. In our study, we identified this variant in nine AD cases, with a mean AAO of 61 years. This finding underscores the role of this variant in EOAD.

Several variants in *GBA1*, such as p.F298L, p.V230G, p.W223R, and p.P198L, have been previously reported in Gaucher disease patients. In our study, we identified these variants in heterozygous state in one AD case and five dementia cases, all with late onset. *GBA1* p.W223R was found in one AD case of Complex Admixture History ancestry. *GBA1* mutations are known to confer an increased risk for dementia in PD and DLB. Notably, they have not been previously suggested to contribute to AD.

Similarly, *APP* p.E693Q has been reported in a few AD cases. In our study, we identified it in a related dementia case and no controls. This finding suggests that this variant may also be implicated in other types of dementia.

Several known variants identified in this study confirm previous findings related to disease type and onset. For example, the *APP* p.V717L variant has been reported in numerous AD cases, primarily in early-onset forms [73,74]. In our study, we identified this variant in two cases of EOAD with AAO ranges of 51-55 years and 56-60 years, respectively. Additional examples are reported in **Supplementary Tables 5 and 6.**

In AoU, the *SNCA* p.Q99R variant was found in a female patient diagnosed in her late 60s with unspecified dementia without behavioral disturbance. In ADSP, the variant was identified in a male patient diagnosed with pure AD in his early 70s. *SNCA* p.P90H and p.A91S were found in two males in their late 70s in the UKB cohort. All four patients were of European ancestry. Previously reported mutations in *SNCA* are known to cause early-onset PD and DLB [75,76]. The mean AAO in patients carrying *SNCA* mutations in this study is 72.75 years. These data suggest that these variants may not be disease-causing but could represent rare risk factors despite their absence in controls and the replication of p.Q99R across datasets.

Novel variants found in this study that may potentially be associated with early-onset dementia include: p.L597W, p.V375I, p.L364F, p.A209T, p.D460N, p.R409C, and p.V227L variants in *APP;* p.R54X and p.M457V in *PSEN1*; p.H169R, p.D320N, and p.G349R in *PSEN2;* p.R556C and p.V28fs in *GRN*; p.G332LfsX64 and p.G701R in *MAPT*; p.G103D and p.A42T in *GBA1;* and *TREM2* p.W44X. Among these variants, *APP* p.L364F and *PSEN1* p.R54X were found in vascular dementia cases with AAO ranges of 41-45 years and 46-50 years, respectively. Additionally, *PSEN2* p.D320N was found in an FTD case with an AAO in his mid-50s. Another notable finding is the identification of *GRN* p.R556C in a dementia case with an AAO in her mid-30s.

***APOE* drives different population-attributable risk for Alzheimer's disease and related dementias**

The summary of our findings on ancestry-specific effects of *APOE* on AD/ADRDs is depicted in **Figure 4**, **Table 5**, and **Supplementary Figure 3**. In AoU, UKB, and 100KGP, the *APOE* ε4/ε4 genotype exhibits a higher frequency among both AD patients and control individuals of African and African Admixed ancestries compared to Europeans. Related dementia patients show similar results in UKB. In AoU, related dementia cases of African Admixed ancestry show a higher frequency than Europeans, while frequencies are similar between Africans and Europeans, likely due to the limited number of individuals with African ancestry in this dataset. In ADSP, the frequencies of this genotype among AD patients are similar across the three ancestries. Among control individuals in ADSP, *APOE* ε4/ε4 is more frequent in African Admixed and African ancestries than in Europeans, as previously reported [77]. Notably, the *APOE* ε4/ε4 genotype was absent from African and African Admixed DLB cases and controls in the AMP PD dataset. The frequency of *APOE* ε4/ε4 in Europeans was higher in cases compared to controls in AMP PD.

When combining results across all datasets, the frequency of *APOE* ε4/ε4 in African and African Admixed AD patients is still higher than in Europeans, but the values are not significantly different. However, the frequency of *APOE* ε4/ε4 in control individuals of African and African Admixed ancestries was found to be substantially higher than in controls of European ancestry. Additionally, the frequency of *APOE* ε4/ε4 in Finnish individuals was found to be higher in AD cases and lower in controls compared to Europeans.

Disease-modifying variants in *APOE* ε4 carriers modulate Alzheimer's and dementia risk across different ancestries

The summary of our findings for the frequencies of protective and disease-modifying variants under study, alongside *APOE* genotypes across all five datasets, is depicted in **Supplementary Tables 7–11.** The proportions of individuals carrying *APOE* ε4 homozygous or heterozygous genotypes alongside protective or disease-modifying variants, within the total population, total ε4/ε4 carriers, and total ε4 carriers across each ancestry, combined across all biobanks, in AD,

related dementias, and controls are reported in **Figure 5, Supplementary Figure 4** and **Supplementary Table 12.** Summaries of our findings for all the assessed models in *APOE* ε4, ε4ε4, and ε3ε3 are shown in **Figure 6, Table 6** and **Supplementary Tables 13–22.**

In brief, we observe higher frequencies of individuals carrying *APOE*:rs449647-T, *19q13.31*:rs10423769-A, *APP*:rs466433-G, or *APP*:rs364048-C protective variant alleles alongside either one or two copies of *APOE* ε4 among African and African Admixed ancestries compared to Europeans in AD, related dementias, and controls. Of note, carriers of *APOE*:rs449647-T and *19q13.31*:rs10423769-A are particularly noteworthy because *APOE*:rs449647-T displays the highest frequency among these ancestries, and the ratio of frequencies for *19q13.31*:rs10423769-A in both African and African Admixed ancestries compared to Europeans is substantially higher than the other protective/disease-modifying variants investigated across all three cohorts. In individuals of African ancestry, *19q13.31*:rs10423769-A was found to have a higher frequency in controls compared to both AD and related dementia cases among *APOE* ε4 homozygous or heterozygous carriers. In contrast, *APOE*:rs449647-T was found to have a lower frequency in controls compared to AD cases among *APOE* ε4 homozygous or heterozygous carriers in African ancestry but showed a higher frequency in controls carrying *APOE* ε4/ε4 versus cases in European populations (**Supplementary Table 12**).

The combination of *TOMM40*:rs11556505-T with either homozygous or heterozygous *APOE* ε4 was observed to have a higher frequency in Europeans and a lower frequency in Africans compared to most other ancestries across all three phenotypes. Additionally, the combination of *NOCT*:rs13116075-G and *APOE* ε4 homozygous or heterozygous was found to have a higher frequency in individuals of European and African Admixed ancestry versus Africans, in AD cases as compared to controls.

The protective model shows a modifying effect of *APOE*:rs449647-T in European, African, and Ashkenazi Jewish ancestries as well as a modifying effect of *TOMM40*:rs11556505-T in European, American Admixed, Ashkenazi Jewish, African Admixed, and individuals of Complex Admixture History. The R^2 model indicates that these variants are not in linkage disequilibrium with the *APOE* risk variants rs429358 and rs7412 **(Table 6)**.

Significant interactions were found between *APOE* ε4 and the following variants: *19q13.31*:rs10423769-A in Africans; *NOCT*:rs13116075-G in both African and African Admixed populations; *CASS4*:rs6024870-A in the Complex Admixed History group; *LRRC37A*:rs2732703-G in the American Admixed ancestry; *APOE*:rs449647-T in European and African Admixed ancestries; and *TOMM40*:rs11556505-T in Europeans. In *APOE* ε4/ε4 carriers, the interaction with *APOE*:rs449647-T was found to be significant in American Admixed and African Admixed populations, while *TOMM40*:rs11556505-T was significant in the European ancestry **(Table 6)**.

The interaction model in *APOE* ε3/ε3 shows no significant p-value for *19q13.31*:rs10423769-A in Africans, *NOCT*:rs13116075-G in African Admixed ancestry, *CASS4*:rs6024870-A in Complex Admixed History, and *LRRC37A*:rs2732703-G in American Admixed ancestry, but highly significant p-values for *APOE*:rs449647-T and *TOMM40*:rs11556505-T in European and *NOCT*:rs13116075-G in Africans with opposite directional effects compared to *APOE* ε4 carriers. These data confirm the role of these variants in modulating the effect of *APOE* ε4 in AD risk **(Table 6)**.

Discussion

We undertook the largest and most comprehensive characterization of potential disease-causing, risk, protective, and disease-modifying variants in known AD/ADRDs genes to date, aiming to create an accessible genetic catalog of both known and novel coding and splicing variants associated with AD/ADRDs in a global context. Our results expand our understanding of the genetic basis of these conditions, potentially leading to new insights into their pathogenesis, risk, and progression. A comprehensive genetic catalog can inform the development of targeted therapies and personalized medicine approaches in the new era of precision therapeutics.

We present a user-friendly platform for the AD/ADRDs research community that enables easy and interactive access to these results (https://niacard.shinyapps.io/MAMBARD_browser/). This browser may serve as a valuable resource for researchers, clinicians, and clinical trial design.

We identified 116 genetic variants (18 known and 98 novel) in AD/ADRDs across diverse populations. The successful replication of novel variants across different datasets increases the likelihood of these variants being pathogenic and warrants further validation through future functional studies, highlighting their potential broader applicability and significance in global genetics research.

We identified 20 potentially disease-causing variants in non-European ancestries, including 13 that were absent in individuals of European ancestry. This highlights the necessity of expanding genetics research to diverse populations, corroborating the notion that the genetic architecture of AD/ADRDs risk differs across populations. We identified a total of 21 variants in control individuals that had been previously reported as disease-causing. This scenario involves three possibilities: (i) the mutation is a non-disease-causing variant found by chance in an AD patient, (ii) the mutation is pathogenic but exhibits incomplete penetrance, or (iii) control individuals represent undiagnosed patients. These findings reveal the potential for conflicting reports and misinterpretations, emphasizing the need for careful analysis and functional validation in genetics research. It underscores the critical importance of caution in identifying and interpreting potentially-pathogenic variants, which is essential for ensuring accurate diagnosis, risk assessment, genetic counseling, and development of effective treatments.

We conducted genotype-phenotype correlations for both known and novel variants. Our findings involving known variants largely reinforce previous studies, while expanding the clinical spectrum for various types of dementia, exhibiting different AAO and/or additional clinical features not previously reported. While the genotype-phenotype correlations for newly identified variants require further investigation to fully understand their impact, we identified 19 novel variants that may potentially be associated with early-onset dementia and therefore warrant further study.

While several studies have conducted *APOE* genotyping across different age groups, sexes, and population ancestries [78,79], the differential role of *APOE* across 11 ancestries in a global context has not yet been explored. We found that individuals of African and African Admixed ancestries harbor a higher frequency of *APOE* ε4/ε4 carriers than individuals of European ancestry. Recent studies have shed light on the varying risk associated with *APOE ε4* alleles in populations of African ancestry. Indeed, it has been reported that individuals of African descent who carry the *APOE ε4* allele have a lower risk of developing AD compared to other populations with the same allele. This suggests that the genetic background of African ancestry around the *APOE* gene is linked to a reduced odds ratio for risk variants [77]. Furthermore, a recent study has suggested the presence of a resilient locus (*19q13.31)* potentially modifying *APOE ε4* risk in African-descent populations. This disease-modifying locus, located 2MB from *APOE*, significantly lowers the AD risk for African *APOE ε4* carriers, reducing the magnitude of the effect from 7.2 to 2.1 [10]. Our finding is in concordance with the largest AD meta-analysis conducted to date [77]. We identified several variants with high frequency among *APOE* ε4 homozygous or heterozygous carriers in African and African Admixed ancestries. Notably, individuals carrying both *APOE* ε4 homozygous or heterozygous and either *APOE*:rs449647-T or *19q13.31*:rs10423769-A exhibit higher frequencies in African and African Admixed ancestries compared to Europeans.

Considering all the models under study, we find that in the presence of *APOE* ε4, *APOE*:rs449647-T decreases the risk of AD in Europeans but increases it in Africans. *TOMM40*:rs11556505-T increases the risk of AD in Europeans. *TOMM40*:rs11556505-T also increases the risk of AD in American Admixed and Ashkenazi Jewish ancestries, though not through an interaction with *APOE*. An interaction effect with *APOE* was found for *19q13.31*:rs10423769-A, *NOCT*:rs13116075-G, *CASS4*:rs6024870-A, and *LRRC37A*:rs2732703-G. Specifically, *19q13.31*:rs10423769-A reduces the risk of AD in Africans, *NOCT*:rs13116075-G reduces the risk in Africans but increases it in African Admixed ancestry, *CASS4*:rs6024870-A reduces the risk in Complex Admixed History ancestry, and *LRRC37A*:rs2732703-G increases the risk in American Admixed ancestry.

Our findings support previous literature, which indicates that *19q13.31* is an African-ancestry-specific locus that reduces the risk effect of *APOE* ε4 for developing AD. *APOE*:rs449647-T is a polymorphism in the regulatory region of *APOE* that can modulate the risk of developing AD by altering its affinity to transcription factors, thus affecting gene expression. Our study demonstrates

its association with an increased risk of AD in *APOE* ε4 carriers of African ancestry and a decreased risk in *APOE* ε4 carriers of European ancestry. *TOMM40*:rs11556505-T has been variably associated with both risk and protective effects, likely depending on the phenotype being evaluated [11]. We show an association with an increased risk of AD in *APOE* ε4 carriers, particularly in Europeans. In addition, this study reveals the disease-modifying effect of *NOCT*:rs13116075-G, *CASS4*:rs6024870-A and *LRRC37A*:rs2732703-G across different ancestries. The interaction of these variants with *APOE* ε4 is not known, but identifying the mechanism(s) conferring protection could provide greater insights into the etiology of AD and inform potential ancestry-specific therapeutic interventions.

This comprehensive genetic characterization, the largest of its kind for AD/ADRDs across diverse populations, holds critical implications for potential clinical trials and therapeutic interventions in a global context, highlighting its significance for such efforts worldwide. For example, clinical trials for *GRN* have recently commenced [80] (https://www.theaftd.org/posts/1ftd-in-the-news/b-ftd-grn-gene-therapy-abio/). Understanding population-specific frequencies of genetic contributors to disease is vital in the design and implementation of clinical trials for several reasons. Firstly, it allows for targeted recruitment, ensuring that studies include an adequate number of participants concordant with their genetic makeup. Secondly, it promotes diversity in clinical trial populations, which is essential to understand disease globally, as well as treatment responses across different groups. Furthermore, knowledge of population-specific variant frequencies may inform treatment efficacy assessments in the future, as genetics y influence treatment outcomes. In summary, it plays a vital role in personalized medicine, guiding more targeted and effective treatments based on individuals' genetic profiles.

Despite our efforts, there are several limitations and shortcomings to consider in this study. A major limitation is the underrepresentation of certain populations, which leads to underpowered datasets limiting the possibility of drawing firm conclusions. Additionally, the reliance on currently available datasets may introduce biases, as these datasets often have varying levels of coverage, quality, and accurate clinical information. Another relevant consideration is the differing exclusion criteria for controls across cohorts, which may affect the comparability of results. Future research should aim to include more diverse populations and improve the quality of genetic data in addition to standardized data harmonization efforts. Moreover, functional validation of identified variants is necessary to confirm their pathogenicity and relevance to AD/ADRDs.

Lastly, ongoing collaborations between researchers, clinicians, and policy-makers are crucial to ensure that advancements in genetic research translate into equitable and effective clinical applications. Our study is a step towards addressing these limitations by providing a more diverse genetic characterization and highlighting the need for inclusive research practices. Future directions should continue to focus on enhancing the robustness and applicability of genetic findings in AD/ADRDs research, making knowledge globally relevant.

Acknowledgments

We thank Paige Brown Jarreau for her meticulous editing of this manuscript.

This work was supported in part by the Intramural Research Program of the NIH, the National Institute on Aging (NIA), National Institutes of Health, Department of Health and Human Services; project number ZO1 AG000535 and ZIA AG000949. This work utilized the computational resources of the NIH HPC Biowulf cluster. (http://hpc.nih.gov).

The All of Us Research Program is supported by the National Institutes of Health, Office of the Director: Regional Medical Centers: 1 OT2 OD026549; 1 OT2 OD026554; 1 OT2 OD026557; 1 OT2 OD026556; 1 OT2 OD026550; 1 OT2 OD 026552; 1 OT2 OD026553; 1 OT2 OD026548; 1 OT2 OD026551; 1 OT2 OD026555; IAA #: AOD 16037; Federally Qualified Health Centers: HHSN 263201600085U; Data and Research Center: 5 U2C OD023196; Biobank: 1 U24 OD023121; The Participant Center: U24 OD023176; Participant Technology Systems Center: 1 U24 OD023163; Communications and Engagement: 3 OT2 OD023205; 3 OT2 OD023206; and Community Partners: 1 OT2 OD025277; 3 OT2 OD025315; 1 OT2 OD025337; 1 OT2 OD025276. In addition, the All of Us Research Program would not be possible without the partnership of its participants.

This research has been conducted using the UK Biobank Resource under application number 33601.

This research was made possible through access to data in the National Genomic Research Library, which is managed by Genomics England Limited (a wholly owned company of the Department of Health and Social Care). The National Genomic Research Library holds data provided by patients and collected by the NHS as part of their care and data collected as part of their participation in research. The National Genomic Research Library is funded by the National Institute for Health Research and NHS England. The Wellcome Trust, Cancer Research UK and the Medical Research Council have also funded research infrastructure.

This research has been conducted using the Alzheimer's Disease Sequencing Project (ADSP) Resource under accession number NG00067. The data for this study were prepared, archived, and distributed by the National Institute on Aging Alzheimer's Disease Data Storage Site (NIAGADS) at the University of Pennsylvania (U24-AG041689), funded by the National Institute on Aging.

This work was supported in part by the Intramural Research Program of the National Institute on Aging and the National Institute of Neurological Disorders and Stroke (project number Z01-AG000949-02). Data used in the preparation of this article were obtained from the AMP PD Knowledge Platform. For up-to-date information on the study, please visit https://www.amp-pd.org. AMP PD—a public-private partnership—is managed by the FNIH and funded by Celgene, GSK, the Michael J. Fox Foundation for Parkinson's Research, the National Institute of

Neurological Disorders and Stroke, Pfizer, Sanofi, and Verily. Clinical data and biosamples used in the preparation of this article were obtained from the Parkinson's Progression Markers Initiative (PPMI), and the Parkinson's Disease Biomarkers Program (PDBP). PPMI—a public-private partnership—is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners, including full names of all of the PPMI funding partners found at http://www.ppmi-info.org/fundingpartners. The PPMI Investigators have not participated in reviewing the data analysis or content of the manuscript. For up-to-date information on the study, visit http://www.ppmi-info.org. The Parkinson's Disease Biomarker Program (PDBP) consortium is supported by the National Institute of Neurological Disorders and Stroke (NINDS) at the National Institutes of Health. A full list of PDBP investigators can be found at https://pdbp.ninds.nih.gov/policy. The PDBP Investigators have not participated in reviewing the data analysis or content of the manuscript. PDBP sample and clinical data collection is supported under grants by NINDS: U01NS082134, U01NS082157, U01NS082151, U01NS082137, U01NS082148, and U01NS082133.

Author Contributions

JDR, CB, MAN, AS and SBC, contributed to the study concept or design. MK, FA, SMG, SCA, PSL, FF, HL, JJK, MJK and MBM, were involved in the analysis of data across different biobanks. All authors contributed to the critical review and had final responsibility for the decision to submit for publication.

Potential Conflicts of Interest

FF, HL, MJK, MBM and MAN's participation in this project was part of a competitive contract awarded to Data Tecnica LLC by the US National Institutes of Health (NIH). The other authors declare that they have no conflict of interest.

Data and code availability

The code used in this study can be found online at https://github.com/NIH-CARD/ADRD-GeneticDiversity-Biobanks,10.5281/zenodo.13363465.

Funding

This work was supported in part by the Intramural Research Program of the NIH, the National Institute on Aging (NIA), National Institutes of Health, Department of Health and Human Services; project number ZO1 AG000535 and ZIA AG000949.

References

1. Sinha JK, Trisal A, Ghosh S, Gupta S, Singh KK, Han SS, et al. Psychedelics for alzheimer's disease-related dementia: Unveiling therapeutic possibilities and pathways. Ageing Res Rev. 2024;96: 102211.

2. GBD 2019 Dementia Forecasting Collaborators. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. Lancet Public Health. 2022;7: e105–e125.

3. McKeith I, Mintzer J, Aarsland D, Burn D, Chiu H, Cohen-Mansfield J, et al. Dementia with Lewy bodies. Lancet Neurol. 2004;3: 19–28.

4. Sagbakken M, Nåden D, Ulstein I, Kvaal K, Langhammer B, Rognstad M-K. Dignity in people with frontotemporal dementia and similar disorders - a qualitative study of the perspective of family caregivers. BMC Health Serv Res. 2017;17: 432.

5. Jonson C, Levine KS, Lake J, Hertslet L, Jones L, Patel D, et al. Assessing the lack of diversity in genetics research across neurodegenerative diseases: a systematic review of the GWAS Catalog and literature. medRxiv. 2024. doi:10.1101/2024.01.08.24301007

6. All of Us Research Program Investigators, Denny JC, Rutter JL, Goldstein DB, Philippakis A, Smoller JW, et al. The "All of Us" Research Program. N Engl J Med. 2019;381: 668–676.

7. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562: 203–209.

8. Bhattarai P, Gunasekaran TI, Belloy ME, Reyes-Dumeyer D, Jülich D, Tayran H, et al. Rare genetic variation in fibronectin 1 (FN1) protects against APOEε4 in Alzheimer's disease. Acta Neuropathol. 2024;147: 70.

9. Arboleda-Velasquez JF, Lopera F, O'Hare M, Delgado-Tirado S, Marino C, Chmielewska N, et al. Resistance to autosomal dominant Alzheimer's disease in an APOE3 Christchurch homozygote: a case report. Nat Med. 2019;25: 1680–1683.

10. Rajabli F, Beecham GW, Hendrie HC, Baiyewu O, Ogunniyi A, Gao S, et al. A locus at 19q13.31 significantly reduces the ApoE ε4 risk for Alzheimer's Disease in African Ancestry. PLoS Genet. 2022;18: e1009977.

11. Andrews SJ, Fulton-Howard B, Goate A. Protective Variants in Alzheimer's Disease. Curr Genet Med Rep. 2019;7: 1–12.

12. Quiroz YT, Aguillon D, Aguirre-Acevedo DC, Vasquez D, Zuluaga Y, Baena AY, et al. Christchurch Heterozygosity and Autosomal Dominant Alzheimer's Disease. N Engl J Med. 2024;390: 2156–2164.

13. Lopera F, Marino C, Chandrahas AS, O'Hare M, Villalba-Moreno ND, Aguillon D, et al.

Resilience to autosomal dominant Alzheimer's disease in a Reelin-COLBOS heterozygous man. Nat Med. 2023;29: 1243–1252.

14. Sassi C, Nalls MA, Ridge PG, Gibbs JR, Ding J, Lupton MK, et al. ABCA7 p.G215S as potential protective factor for Alzheimer's disease. Neurobiol Aging. 2016;46: 235.e1–9.

15. Minikel EV, Painter JL, Dong CC, Nelson MR. Refining the impact of genetic evidence on clinical success. Nature. 2024;629: 624–629.

16. Multidirectional genetic and genomic data sharing in the All of Us research program. Genomic Data Sharing. Academic Press; 2023. pp. 39–69.

17. Harrison SM, Austin-Tse CA, Kim S, Lebo M, Leon A, Murdock D, et al. Harmonizing variant classification for return of results in the All of Us Research Program. Hum Mutat. 2022;43: 1114–1121.

18. Venner E, Muzny D, Smith JD, Walker K, Neben CL, Lockwood CM, et al. Whole-genome sequencing as an investigational device for return of hereditary disease risk and pharmacogenomic results as part of the All of Us Research Program. Genome Med. 2022;14: 34.

19. Conomos MP, Reiner AP, Weir BS, Thornton TA. Model-free Estimation of Recent Genetic Relatedness. Am J Hum Genet. 2016;98: 127–148.

20. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10. doi:10.1093/gigascience/giab008

21. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4: 7.

22. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen W-M. Robust relationship inference in genome-wide association studies. Bioinformatics. 2010;26: 2867–2873.

23. Li S, Carss KJ, Halldorsson BV, Cortes A, UK Biobank Whole-Genome Sequencing Consortium. Whole-genome sequencing of half-a-million UK Biobank participants. medRxiv. 2023. p. 2023.12.06.23299426. doi:10.1101/2023.12.06.23299426

24. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011;27: 2987–2993.

25. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38: e164.

26. 100,000 Genomes Project Pilot Investigators, Smedley D, Smith KR, Martin A, Thomas EA, McDonagh EM, et al. 100,000 Genomes Pilot on Rare-Disease Diagnosis in Health Care - Preliminary Report. N Engl J Med. 2021;385: 1868–1880.

27. Smedley D, Jacobsen JOB, Jäger M, Köhler S, Holtgrewe M, Schubach M, et al. Next-generation diagnostics and disease-gene discovery with the Exomiser. Nat Protoc. 2015;10: 2004–2015.

28. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81: 559–575.

29. Vitale D, Koretsky M, Kuznetsov N, Hong S, Martin J, James M, et al. GenoTools: An Open-Source Python Package for Efficient Genotype Data Quality Control and Analysis. bioRxiv. 2024. doi:10.1101/2024.03.26.586362

30. Appel-Cresswell S, Vilarino-Guell C, Encarnacion M, Sherman H, Yu I, Shah B, et al. Alpha-synuclein p.H50Q, a novel pathogenic mutation for Parkinson's disease. Mov Disord. 2013;28: 811–813.

31. Blauwendraat C, Kia DA, Pihlstrøm L, Gan-Or Z, Lesage S, Gibbs JR, et al. Insufficient evidence for pathogenicity of SNCA His50Gln (H50Q) in Parkinson's disease. Neurobiol Aging. 2018;64: 159.e5–159.e8.

32. Rossi G, Giaccone G, Maletta R, Morbin M, Capobianco R, Mangieri M, et al. A family with Alzheimer disease and strokes associated with A713T mutation of the APP gene. Neurology. 2004;63: 910–912.

33. Conidi ME, Bernardi L, Puccio G, Smirne N, Muraca MG, Curcio SAM, et al. Homozygous carriers of APP A713T mutation in an autosomal dominant Alzheimer disease family. Neurology. 2015;84: 2266–2273.

34. Abbatemarco JR, Jones SE, Larvie M, Bekris LM, Khrestian ME, Krishnan K, et al. Amyloid Precursor Protein Variant, E665D, Associated With Unique Clinical and Biomarker Phenotype. Am J Alzheimers Dis Other Demen. 2021;36: 1533317520981225.

35. Peacock ML, Murman DL, Sima AA, Warren JT Jr, Roses AD, Fink JK. Novel amyloid precursor protein gene mutation (codon 665Asp) in a patient with late-onset Alzheimer's disease. Ann Neurol. 1994;35: 432–438.

36. Hsu S, Pimenova AA, Hayes K, Villa JA, Rosene MJ, Jere M, et al. Systematic validation of variants of unknown significance in APP, PSEN1 and PSEN2. Neurobiol Dis. 2020;139: 104817.

37. Koprivica V, Stone DL, Park JK, Callahan M, Frisch A, Cohen IJ, et al. Analysis and classification of 304 mutant alleles in patients with type 1 and type 3 Gaucher disease. Am J Hum Genet. 2000;66: 1777–1786.

38. Arndt S, Heitner R, Lane A, Ramsay M. Glucocerebrosidase gene mutations in black South Africans with Gaucher disease. Blood Cells Mol Dis. 2009;43: 129–133.

39. Jeong S-Y, Park S-J, Kim HJ. Clinical and genetic characteristics of Korean patients with

Gaucher disease. Blood Cells Mol Dis. 2011;46: 11–14.

40. Beutler E, Gelbart T, Kuhl W, Zimran A, West C. Mutations in Jewish patients with Gaucher disease. Blood. 1992;79: 1662–1666.

41. Hong CM, Ohashi T, Yu XJ, Weiler S, Barranger JA. Sequence of two alleles responsible for Gaucher disease. DNA Cell Biol. 1990;9: 233–241.

42. Tammachote R, Tongkobpetch S, Srichomthong C, Phipatthanananti K, Pungkanon S, Wattanasirichaigoon D, et al. A common and two novel GBA mutations in Thai patients with Gaucher disease. J Hum Genet. 2013;58: 594–599.

43. Kurolap A, Del Toro M, Spiegel R, Gutstein A, Shafir G, Cohen IJ, et al. Gaucher disease type 3c: New patients with unique presentations and review of the literature. Mol Genet Metab. 2019;127: 138–146.

44. Beutler E, Gelbart T, Scott CR. Hematologically important mutations: Gaucher disease. Blood Cells Mol Dis. 2005;35: 355–364.

45. Amaral O, Lacerda L, Marcão A, Pinto E, Tamagnini G, Sá Miranda MC. Homozygosity for two mild glucocerebrosidase mutations of probable Iberian origin. Clin Genet. 1999;56: 100–102.

46. Montfort M, Chabás A, Vilageliu L, Grinberg D. Functional analysis of 13 GBA mutant alleles identified in Gaucher disease patients: Pathogenic changes and “modifier” polymorphisms. Hum Mutat. 2004;23: 567–575.

47. Alfonso P, Aznarez S, Giralt M, Pocovi M, Giraldo P, Spanish Gaucher’s Disease Registry. Mutation analysis and genotype/phenotype relationships of Gaucher disease patients in Spain. J Hum Genet. 2007;52: 391–396.

48. Sheth J, Bhavsar R, Mistri M, Pancholi D, Bavdekar A, Dalal A, et al. Gaucher disease: single gene molecular characterization of one-hundred Indian patients reveals novel variants and the most prevalent mutation. BMC Med Genet. 2019;20: 31.

49. Malini E, Grossi S, Deganuto M, Rosano C, Parini R, Dominisini S, et al. Functional analysis of 11 novel GBA alleles. Eur J Hum Genet. 2014;22: 511–516.

50. Kim JW, Liou BB, Lai MY, Ponce E, Grabowski GA. Gaucher disease: identification of three new mutations in the Korean and Chinese (Taiwanese) populations. Hum Mutat. 1996;7: 214–218.

51. Beutler E, Gelbart T, Demina A, Zimran A, LeCoutre P. Five new Gaucher disease mutations. Blood Cells Mol Dis. 1995;21: 20–24.

52. Zampieri S, Cattarossi S, Bembi B, Dardis A. GBA Analysis in Next-Generation Era: Pitfalls, Challenges, and Possible Solutions. J Mol Diagn. 2017;19: 733–741.

53. Baker M, Mackenzie IR, Pickering-Brown SM, Gass J, Rademakers R, Lindholm C, et al. Mutations in progranulin cause tau-negative frontotemporal dementia linked to chromosome 17. Nature. 2006;442: 916–919.

54. Bit-Ivan EN, Suh E, Shim H-S, Weintraub S, Hyman BT, Arnold SE, et al. A novel GRN mutation (GRN c.708+6_+9delTGAG) in frontotemporal lobar degeneration with TDP-43-positive inclusions: clinicopathologic report of 6 cases. J Neuropathol Exp Neurol. 2014;73: 467–473.

55. Chiang H-H, Forsell C, Lilius L, Öijerstedt L, Thordardottir S, Shanmugarajan K, et al. Novel progranulin mutations with reduced serum-progranulin levels in frontotemporal lobar degeneration. Eur J Hum Genet. 2013;21: 1260–1265.

56. Rademakers R, Baker M, Gass J, Adamson J, Huey ED, Momeni P, et al. Phenotypic variability associated with progranulin haploinsufficiency in patients with the common 1477C-->T (Arg493X) mutation: an international initiative. Lancet Neurol. 2007;6: 857–868.

57. Orme T, Hernandez D, Ross OA, Kun-Rodrigues C, Darwent L, Shepherd CE, et al. Analysis of neurodegenerative disease-causing genes in dementia with Lewy bodies. Acta Neuropathol Commun. 2020;8: 5.

58. McKeith IG, Dickson DW, Lowe J, Emre M, O'Brien JT, Feldman H, et al. Diagnosis and management of dementia with Lewy bodies: third report of the DLB Consortium. Neurology. 2005;65: 1863–1872.

59. Vega IE, Cabrera LY, Wygant CM, Velez-Ortiz D, Counts SE. Alzheimer's Disease in the Latino Community: Intersection of Genetics and Social Determinants of Health. J Alzheimers Dis. 2017;58: 979–992.

60. Lee JH, Kahn A, Cheng R, Reitz C, Vardarajan B, Lantigua R, et al. Disease-related mutations among Caribbean Hispanics with familial dementia. Mol Genet Genomic Med. 2014;2: 430–437.

61. Masellis M, Momeni P, Meschino W, Heffner R Jr, Elder J, Sato C, et al. Novel splicing mutation in the progranulin gene causing familial corticobasal syndrome. Brain. 2006;129: 3115–3123.

62. Steele NZ, Bright AR, Lee SE, Fong JC, Bonham LW, Karydas A, et al. Frequency of frontotemporal dementia gene variants in C9ORF72, MAPT, and GRN in academic versus commercial laboratory cohorts. Adv Genomics Genet. 2018;8: 23–33.

63. Gass J, Cannon A, Mackenzie IR, Boeve B, Baker M, Adamson J, et al. Mutations in progranulin are a major cause of ubiquitin-positive frontotemporal lobar degeneration. Hum Mol Genet. 2006;15: 2988–3001.

64. Gómez-Isla T, Wasco W, Pettingell WP, Gurubhagavatula S, Schmidt SD, Jondro PD, et al. A novel presenilin-1 mutation: increased beta-amyloid and neurofibrillary changes. Ann

Neurol. 1997;41: 809–813.

65. Jiao B, Liu H, Guo L, Xiao X, Liao X, Zhou Y, et al. The role of genetics in neurodegenerative dementia: a large cohort study in South China. NPJ Genom Med. 2021;6: 69.

66. Larner AJ, Ray PS, Doran M. The R269H mutation in presenilin-1 presenting as late-onset autosomal dominant Alzheimer's disease. J Neurol Sci. 2007;252: 173–176.

67. Cohn-Hokke PE, Wong TH, Rizzu P, Breedveld G, van der Flier WM, Scheltens P, et al. Mutation frequency of PRKAR1B and the major familial dementia genes in a Dutch early onset dementia cohort. J Neurol. 2014;261: 2085–2092.

68. Mann DM, Pickering-Brown SM, Takeuchi A, Iwatsubo T, Members of the Familial Alzheimer's Disease Pathology Study Group. Amyloid angiopathy and variability in amyloid beta deposition is determined by mutation position in presenilin-1-linked Alzheimer's disease. Am J Pathol. 2001;158: 2165–2175.

69. Ryan NS, Nicholas JM, Weston PSJ, Liang Y, Lashley T, Guerreiro R, et al. Clinical phenotype and genetic associations in autosomal dominant familial Alzheimer's disease: a case series. Lancet Neurol. 2016;15: 1326–1335.

70. Saito Y, Geyer A, Sasaki R, Kuzuhara S, Nanba E, Miyasaka T, et al. Early-onset, rapidly progressive familial tauopathy with R406W mutation. Neurology. 2002;58: 811–813.

71. Lindquist SG, Holm IE, Schwartz M, Law I, Stokholm J, Batbayli M, et al. Alzheimer disease-like clinical phenotype in a family with FTDP-17 caused by a MAPT R406W mutation. Eur J Neurol. 2008;15: 377–385.

72. Wojtas A, Heggeli KA, Finch N, Baker M, Dejesus-Hernandez M, Younkin SG, et al. C9ORF72 repeat expansions and other FTD gene mutations in a clinical AD patient series from Mayo Clinic. Am J Neurodegener Dis. 2012;1: 107–118.

73. van Duijn CM, Hendriks L, Cruts M, Hardy JA, Hofman A, Van Broeckhoven C. Amyloid precursor protein gene mutation in early-onset Alzheimer's disease. Lancet. 1991;337: 978.

74. Karlinsky H, Vaula G, Haines JL, Ridgley J, Bergeron C, Mortilla M, et al. Molecular and prospective phenotypic characterization of a pedigree with familial Alzheimer's disease and a missense mutation in codon 717 of the beta-amyloid precursor protein gene. Neurology. 1992;42: 1445–1453.

75. Alcalay RN, Caccappolo E, Mejia-Santana H, Tang MX, Rosado L, Ross BM, et al. Frequency of known mutations in early-onset Parkinson disease: implication for genetic counseling: the consortium on risk for early onset Parkinson disease study. Arch Neurol. 2010;67: 1116–1122.

76. Kapasi A, Brosch JR, Nudelman KN, Agrawal S, Foroud TM, Schneider JA. A novel SNCA E83Q mutation in a case of dementia with Lewy bodies and atypical frontotemporal

lobar degeneration. Neuropathology. 2020;40: 620–626.

77. Lake J, Warly Solsberg C, Kim JJ, Acosta-Uribe J, Makarious MB, Li Z, et al. Multi-ancestry meta-analysis and fine-mapping in Alzheimer's disease. Mol Psychiatry. 2023;28: 3121–3132.

78. Belloy ME, Andrews SJ, Le Guen Y, Cuccaro M, Farrer LA, Napolioni V, et al. APOE Genotype and Alzheimer Disease Risk Across Age, Sex, and Population Ancestry. JAMA Neurol. 2023;80: 1284–1294.

79. Lumsden AL, Mulugeta A, Zhou A, Hyppönen E. Apolipoprotein E (APOE) genotype-associated disease risks: a phenome-wide, registry-based, case-control study utilising the UK Biobank. EBioMedicine. 2020;59: 102954.

80. Sevigny J, Uspenskaya O, Heckman LD, Wong LC, Hatch DA, Tewari A, et al. Progranulin AAV gene therapy for frontotemporal dementia: translational studies and phase 1/2 trial interim results. Nat Med. 2024;30: 1406–1415.

Figure legends

Figure 1- Demographic and clinical characteristics of biobank-scale cohorts under study.

The figure illustrates distributions of age, sex, and the number of cases and controls per ancestry across five datasets in this study: All of Us (AoU), Alzheimer's Disease Sequencing Project (ADSP), 100,000 Genomes Project (100KGP), UK Biobank (UKB), and Accelerating Medicines Partnership in Parkinson's Disease (AMP PD). Ancestries represented include European (EUR), African (AFR), American Admixed (AMR), African Admixed (AAC), Ashkenazi Jewish (AJ), Central Asian (CAS), Eastern Asian (EAS), South Asian (SAS), Middle Eastern (MDE), Finnish (FIN), and Complex Admixture History (CAH).

Figure 2- Workflow.

Our workflow begins with creating cohorts within the datasets. We leverage short-read whole genome sequencing data to characterize genes of interest. Variant annotation focuses on missense, frameshift, start loss, stop loss, stop gain, and splicing variants. Next, we compare the frequency of identified variants in cases and controls. Pathogenicity assessment involves using ClinVar, Human Gene Mutation Database (HGMD), American College of Medical Genetics and Genomics (ACMG) guidelines, and Combined Annotation Dependent Depletion (CADD) scores. Finally, we prioritize variants that are present only in the case cohort and have a CADD score greater than 20.

Figure 3- Mutation sites from identified genetic variants mapped on the predicted protein structures encoded by genes associated with AD/ADRDs.

The predicted protein structures encoded by eight genes associated with AD/ADRDs (*APP*, *PSEN1*, *PSEN2*, *TREM2*, *GBA1*, *GRN*, *MAPT*, and *SNCA*) were obtained from the EMBL AlphaFold Protein Structure Database to ensure that all of the residues are present in each protein structure. PyMOL v. 2.6.0 was used to represent the protein structures and their associated mutation sites from identified genetic variants. The yellow color shows beta sheets, the red color shows alpha helices, and the green color shows connecting loops and turns.

Figure 4- Proportions of *APOE* ε4/ε4 across 11 genetic ancestries in Alzheimer's disease, related dementias, and controls in all datasets.

Figure 5- Proportions of individuals carrying both *APOE* ε4 and protective or disease-modifying variants across 11 genetic ancestries in Alzheimer's disease, related dementias, and controls in all datasets.

(A) SNP frequencies for each variant within each ancestry relative to the total number of ε4 carriers per ancestry across all datasets. (B) AD-to-control allele frequency ratios (left) and related dementia-to-control ratios (right)**.** Warmer colors represent higher frequencies in cases versus

controls, while cooler colors represent higher frequencies in controls versus cases, with dark blue (N/A) representing variants not present in either cases or controls.

Figure 6- Upset plot showing protective, conditional, and interaction models across multiple ancestries.

The Y-axis represents each ancestry population with a large enough sample size, and the X-axis represents the six protective/disease-modifying variants. The color bar shows the magnitude of effects as log of the odds ratio (beta value) and directionality, with red color denoting negative directionality, and blue colors denoting positive directionality.

Supplementary Figure 1- PCA plots in (A) All of Us, (B) UKB, (C) ADSP, (D) AMP PD, and (E) 100 KGP.

Supplementary Figure 2- Heatmaps showing the frequencies of all identified variants in the discovery and replication phases across all ancestries in each biobank.

Supplementary Figure 3- Proportion of *APOE* genotypes in (A) Alzheimer's disease, (B) related dementias, and (C) controls across 11 genetic ancestries.

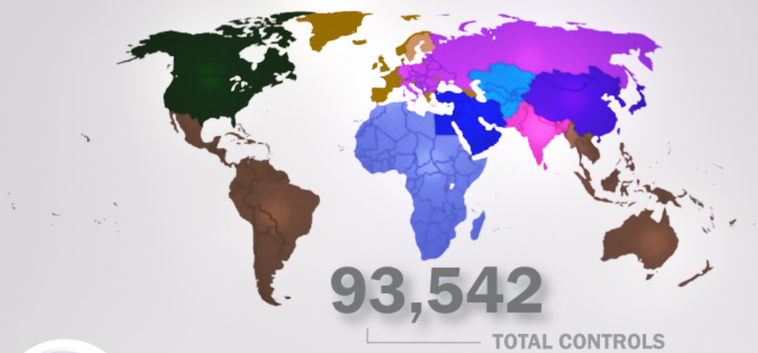
Unknown genotypes and those absent across all ancestries were excluded from the analysis. Genotypes and ancestries not available in the 100KGP were also excluded.

Supplementary Figure 4- Proportions of individuals carrying both *APOE* ε4/ε4 genotypes and protective or disease-modifying variants across 11 genetic ancestries in Alzheimer's disease, related dementias, and controls in all datasets.

Supplementary Figures 4A and 4C represent SNP distribution within each cohort, and Supplementary Figures 4B and 4D represent SNP distribution between cohorts. The total populations of each ancestry were used to generate 4A and 4B, while the total numbers of ε4/ε4 carriers for each ancestry were used to generate 4C and 4D. Supplementary Figures 4B and 4D show allele frequency ratios (AD-to-Control, left; Related dementias-to-Control, right) among *APOE* ε4/ε4 carriers for each of the candidate protective or disease-modifying variant, per ancestry. Warmer colors represent higher frequencies in cases versus controls, while cooler colors represent higher frequencies in controls versus cases, with dark blue (N/A) representing variants not present in either cases or controls.

25,001

TOTAL ALZHEIMER'S DISEASE AND RELATED DEMENTIA CASES



- AAC = African Admixed
- AMR= American Admixed
- FIN = Finnish
- CAS = Central Asian
- MDE = Middle Eastern
- AFR = African
- EAS = Eastern Asian
- AJ = Ashkenazi Jewish
- SAS = South Asian
- EUR = European
- CAH = Complex Admixture History

All of Us Research Program

AD 539
50 AJ, 0 CAH, 333 EUR, 71 AMR, 2 SAS, 7 EAS, 2 CAS, 1 MDE, 41 AAC, 32 AFR, 0 FIN

Dementia 1,655
103 AJ, 0 CAH, 975 EUR, 214 AMR, 10 SAS, 19 EAS, 7 CAS, 6 MDE, 144 AAC, 175 AFR, 2 FIN

Controls 13,835
873 AJ, 0 CAH, 8880 EUR, 1304 AMR, 73 SAS, 366 EAS, 42 CAS, 66 MDE, 898 AAC, 1305 AFR, 28 FIN

GENDER

	♂	♀
AD	253	274
Dementia	778	830
Controls	6,457	7,046

AGE

	♀	♂
AD	77.8±10.8 YRS.	79.0±9.6 YRS.
Dementia	75.1±12.2 YRS.	75.7±11.0 YRS.
Controls	77.0±5.6 YRS.	77.5±5.9 YRS.

adsp

AD 10,566
537 AJ, 1064 CAH, 5812 EUR, 1086 AMR, 176 SAS, 50 EAS, 20 CAS, 27 MDE, 1119 AAC, 664 AFR, 11 FIN

Controls 16,217
459 AJ, 1672 CAH, 4411 EUR, 3967 AMR, 2279 SAS, 29 EAS, 131 CAS, 9 MDE, 1563 AAC, 1685 AFR, 12 FIN

GENDER

	♂	♀
AD	4,200	6,366
Controls	6,044	10,173

AGE

	♀	♂
AD	70.7±17.1 YRS.	68.4±17.1 YRS.
Controls	70.7±17.1 YRS.	68.4±17.1 YRS.

100KGP

Dementia 180
143 EUR, 7 AFR, 3 AMR, 1 EAS, 16 SAS, 10 NA

Controls 3,479
3037 EUR, 62 AFR, 86 AMR, 22 EAS, 241 SAS, 31 NA

GENDER

	♂	♀
Dementia	90	90
Controls	1,669	1,810

AGE

	♀	♂
Dementia	68.6±12.1 YRS.	68.2±13.4 YRS.
Controls	75.1±7.9 YRS.	74.3±7.7 YRS.

biobank

AD 4,225
27 AJ, 8 CAH, 4051 EUR, 0 AMR, 49 SAS, 7 EAS, 10 CAS, 9 MDE, 12 AAC, 52 AFR, 0 FIN

Dementia 5,306
43 AJ, 27 CAH, 5028 EUR, 3 AMR, 70 SAS, 10 EAS, 23 CAS, 13 MDE, 24 AAC, 64 AFR, 1 FIN

Controls 56,741
416 AJ, 169 CAH, 54423 EUR, 48 AMR, 672 SAS, 154 EAS, 204 CAS, 62 MDE, 133 AAC, 445 AFR, 15 FIN

GENDER

	♂	♀
AD	1,999	2,226
Dementia	2,960	2,346
Controls	28,341	28,400

AGE

	♀	♂
AD	64.6±4.4 YRS.	64.7±4.4 YRS.
Dementia	63.7±5.2 YRS.	63.8±5.1 YRS.
Controls	66.9±1.5 YRS.	66.9±1.5 YRS.

AMP PD

DLB 2,530
113 AJ, 0 CAH, 2406 EUR, 0 AMR, 0 SAS, 0 EAS, 0 CAS, 6 MDE, 0 AAC, 0 AFR, 5 FIN

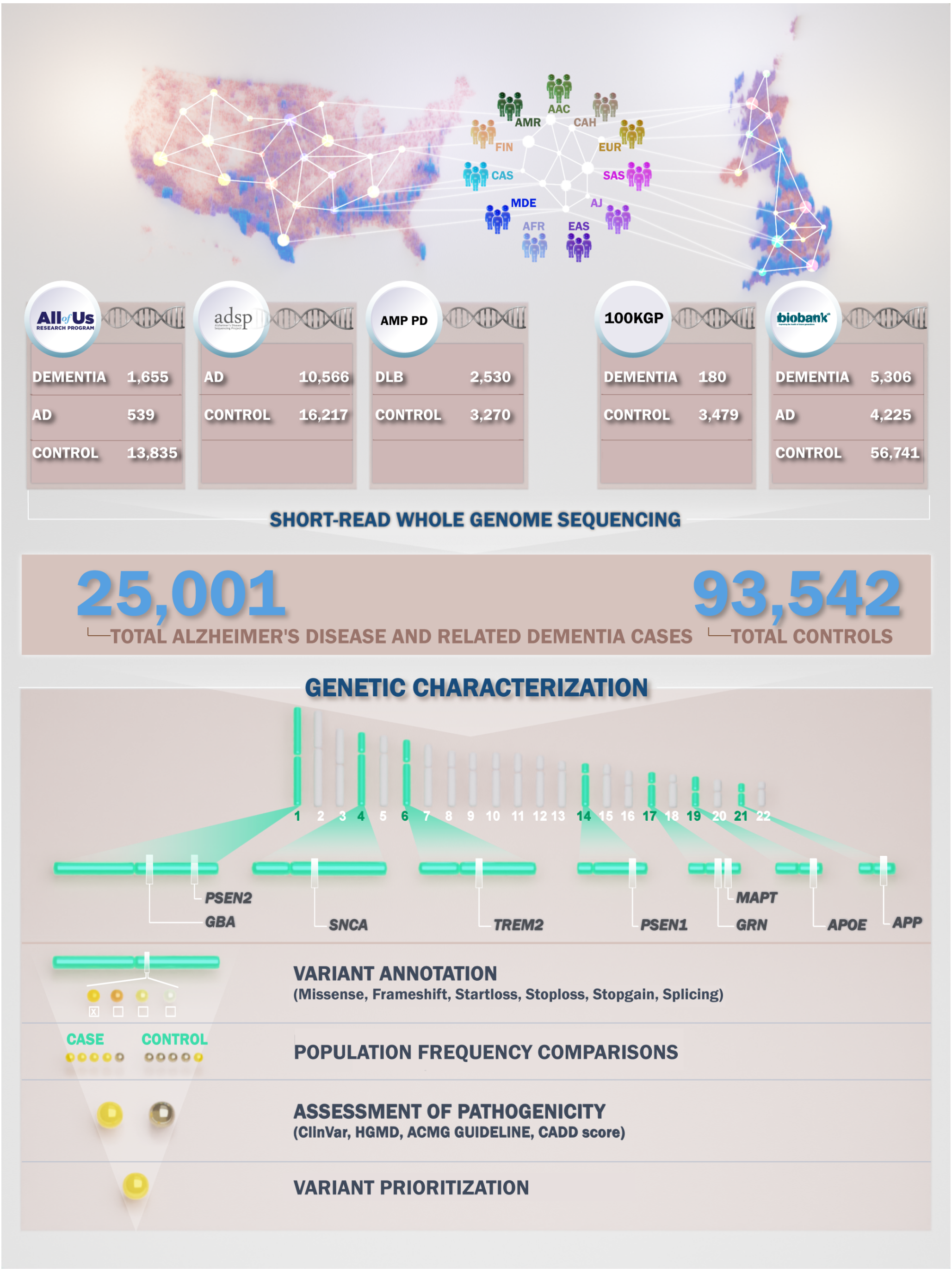
Controls 3,270
246 AJ, 4 CAH, 2919 EUR, 14 AMR, 1 SAS, 5 EAS, 3 CAS, 4 MDE, 40 AAC, 29 AFR, 5 FIN

GENDER

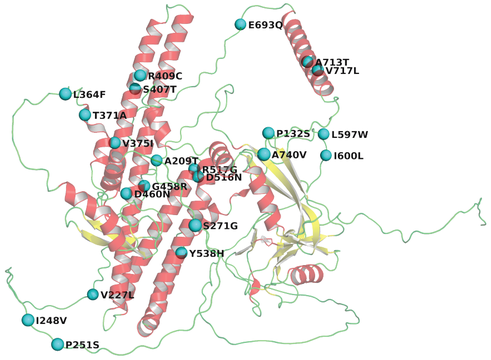
	♂	♀
DLB	1,618	912
Controls	1,563	1,707

AGE

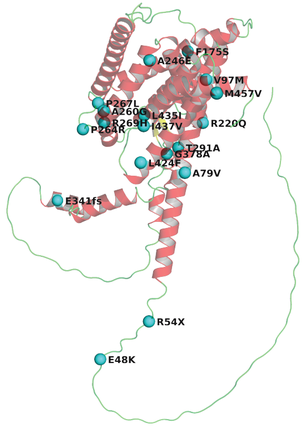
	♀	♂
DLB	79.0±8.9 YRS.	75.7±8.5 YRS.
Controls	68.8±13.1 YRS.	69.9±13.3 YRS.


AAC
AMR
CAH
FIN
EUR
CAS
SAS
MDE
AJ
AFR
EAS
All of Us RESEARCH PROGRAM
DEMENTIA 1,655
AD 539
CONTROL 13,835
adsp
AD 10,566
CONTROL 16,217
AMP PD
DLB 2,530
CONTROL 3,270
100KGP
DEMENTIA 180
CONTROL 3,479
biobank
DEMENTIA 5,306
AD 4,225
CONTROL 56,741
SHORT-READ WHOLE GENOME SEQUENCING
25,001
TOTAL ALZHEIMER'S DISEASE AND RELATED DEMENTIA CASES
93,542
TOTAL CONTROLS
GENETIC CHARACTERIZATION
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
PSEN2
GBA
SNCA
TREM2
PSEN1
MAPT
GRN
APOE
APP
CASE
CONTROL
VARIANT ANNOTATION
(Missense, Frameshift, Startloss, Stoploss, Stopgain, Splicing)
POPULATION FREQUENCY COMPARISONS
ASSESSMENT OF PATHOGENICITY
(ClinVar, HGMD, ACMG GUIDELINE, CADD score)
VARIANT PRIORITIZATION

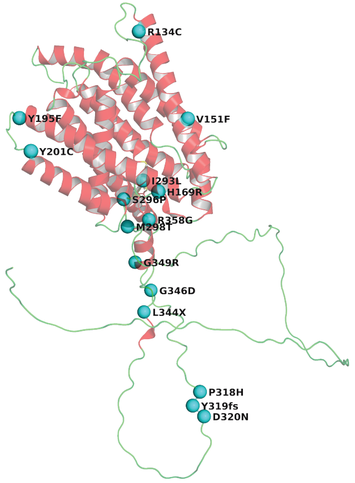
APP
Amyloid precursor protein



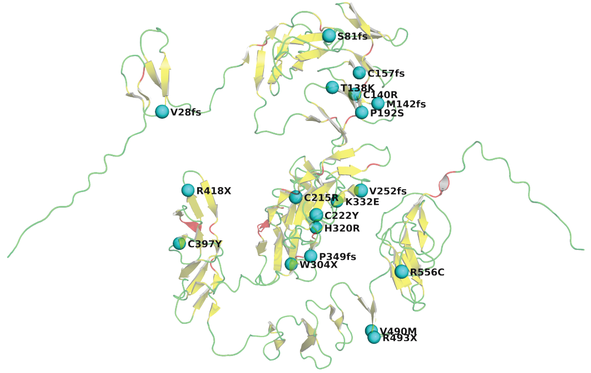
PSEN1
Presenilin-1



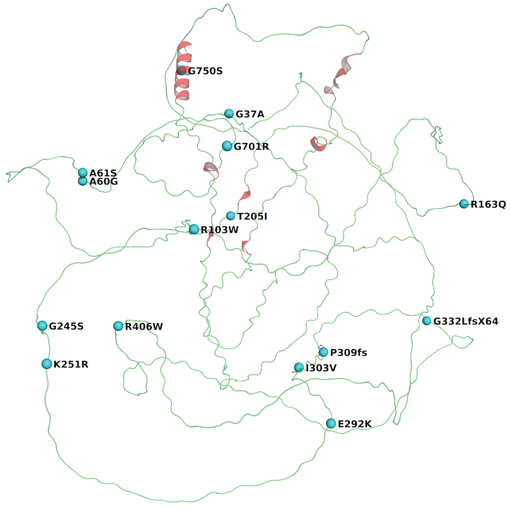
PSEN2
Presenilin-2



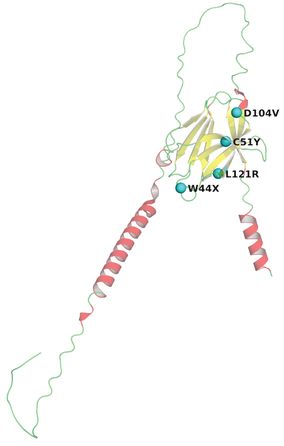
GRN
Progranulin



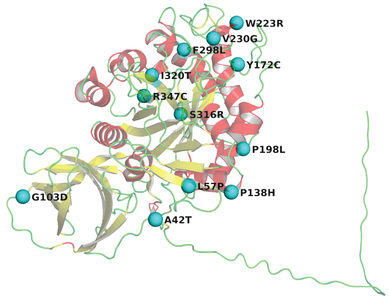
MAPT
Microtubule-associated protein tau



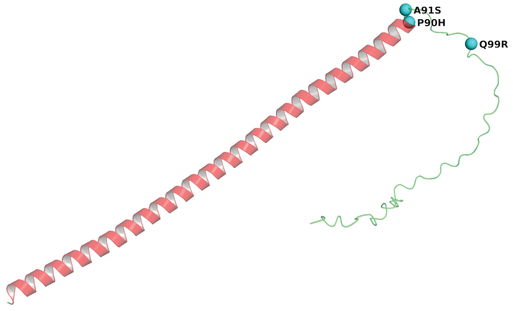
TREM2
Triggering receptor expressed on myeloid cells 2

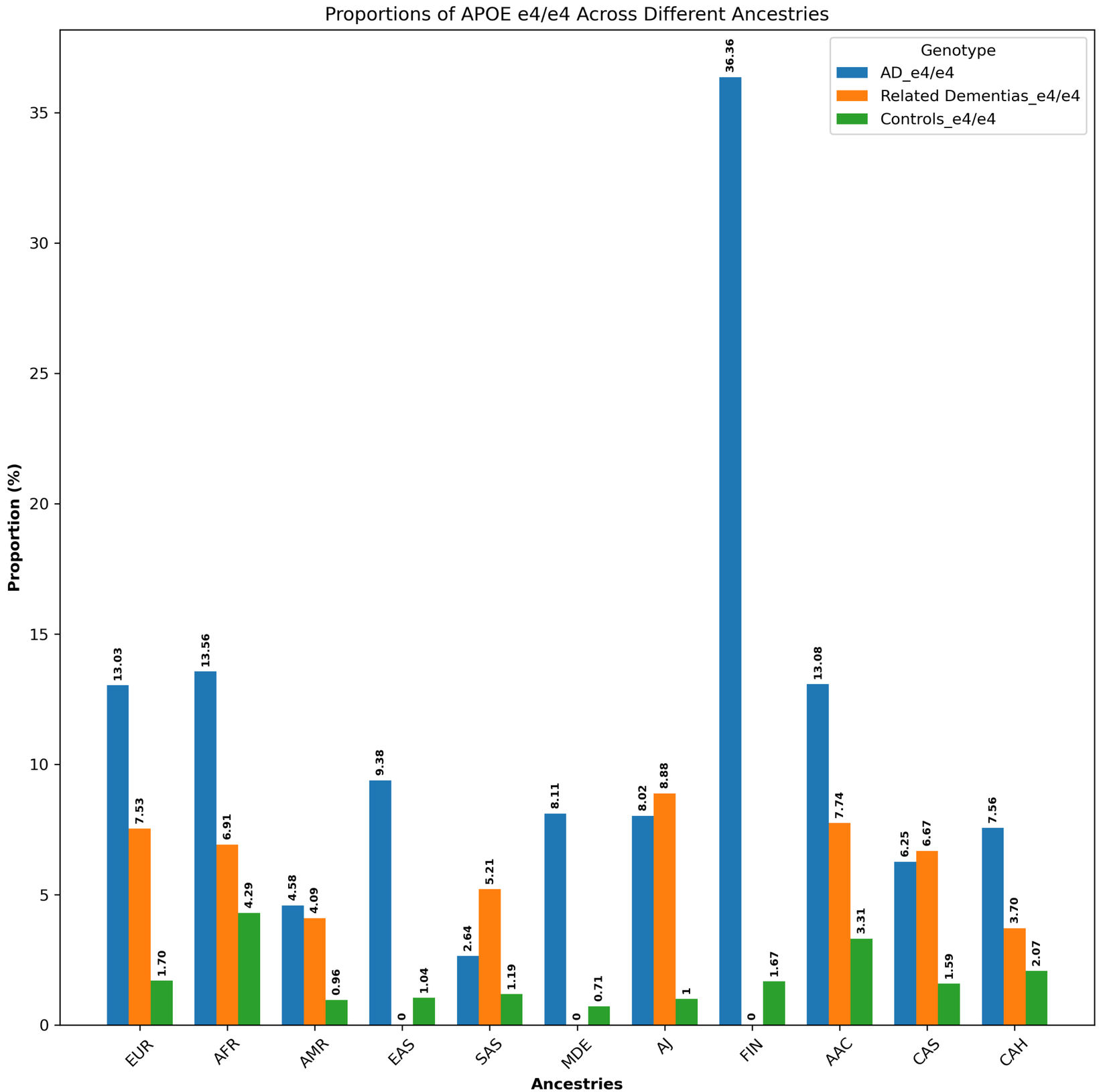


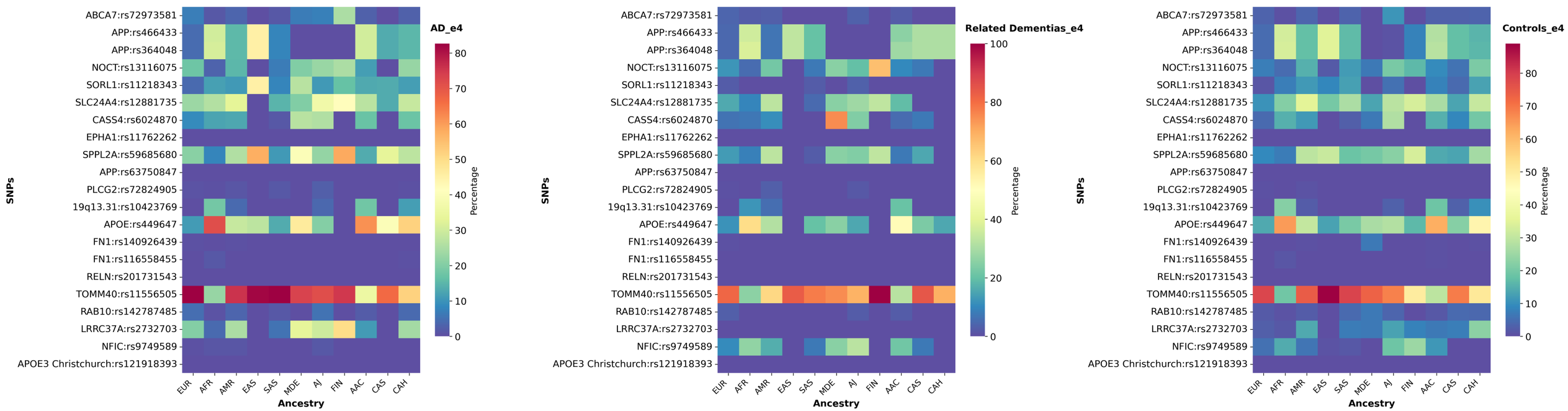
GBA1
Glucocerebrosidase

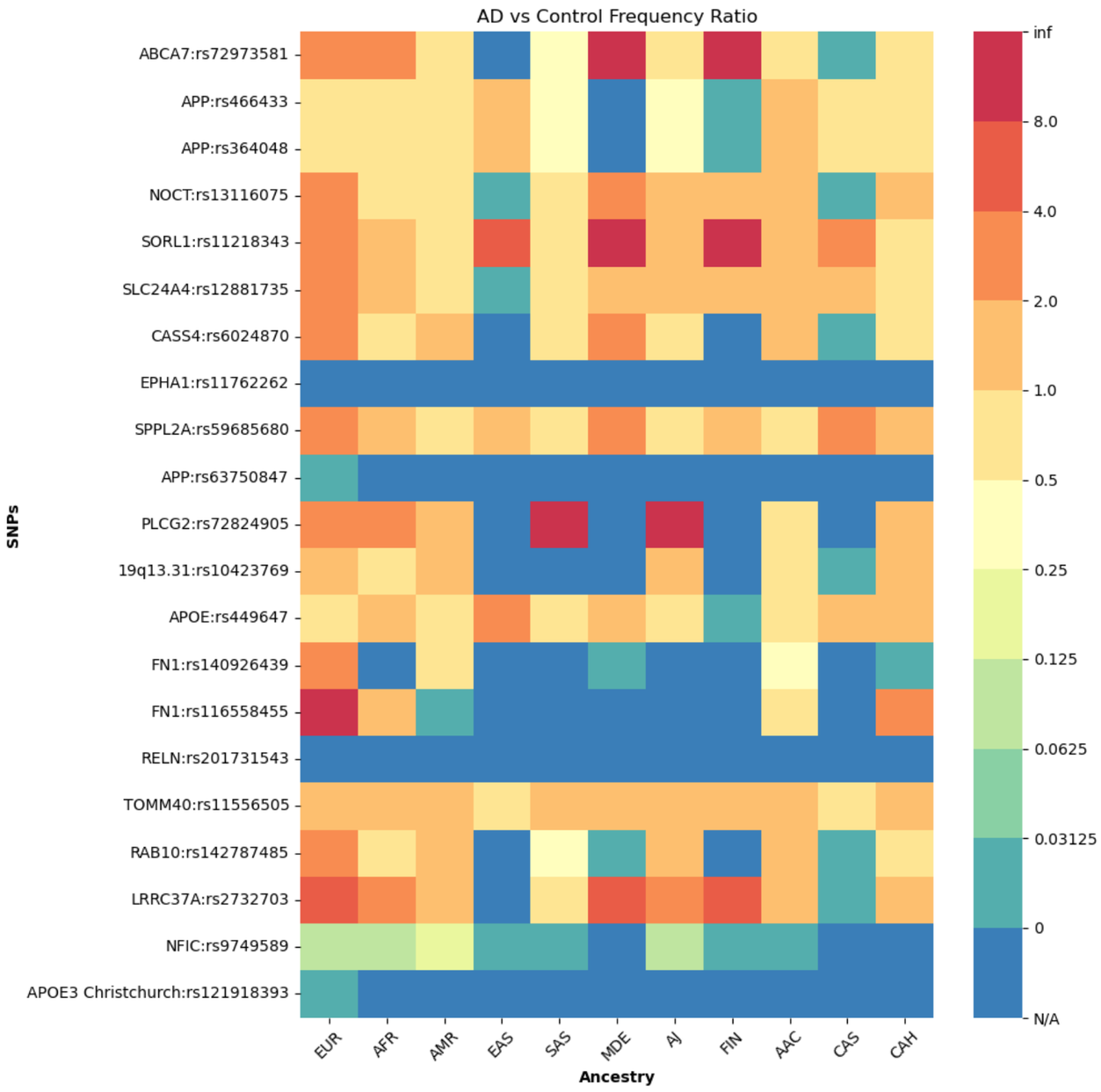


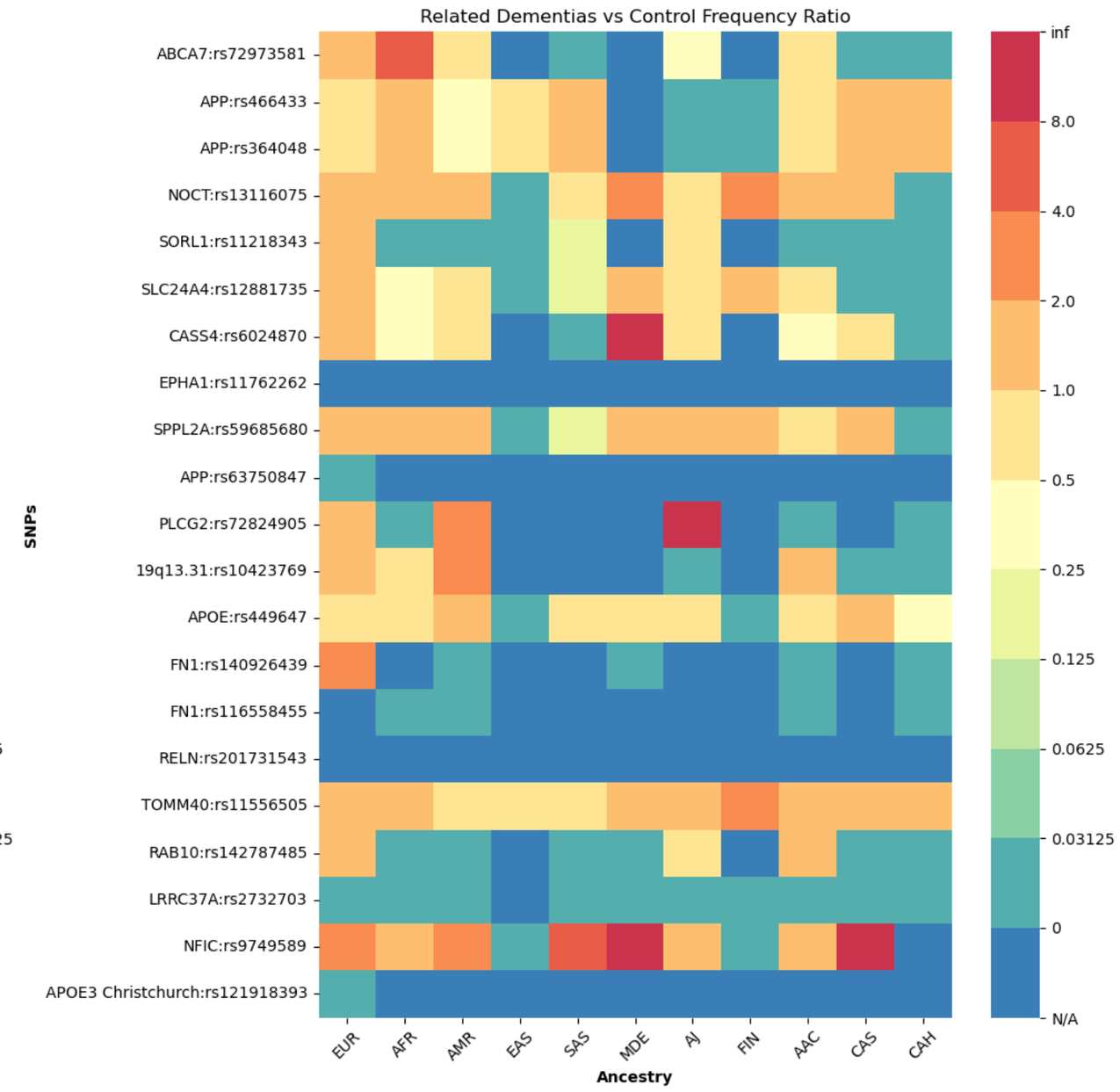
SNCA
Alpha-synuclein

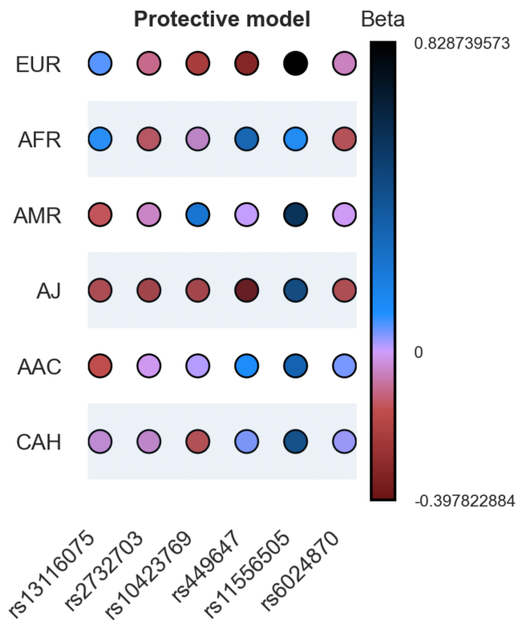


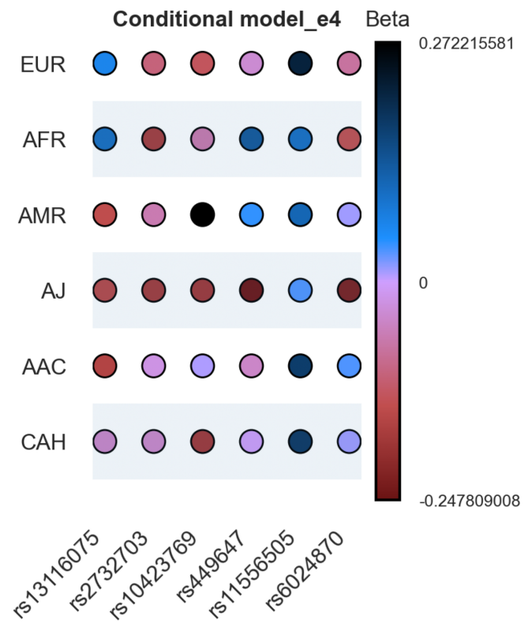

Proportions of APOE e4/e4 Across Different Ancestries
Genotype
AD_e4/e4
Related Dementias_e4/e4
Controls_e4/e4
Proportion (%)
0
5
10
15
20
25
30
35
13.03
7.53
1.70
13.56
6.91
4.29
4.58
4.09
0.96
9.38
0
1.04
2.64
5.21
1.19
8.11
0
0.71
8.02
8.88
1
36.36
0
1.67
13.08
7.74
3.31
6.25
6.67
1.59
7.56
3.70
2.07
EUR
AFR
AMR
EAS
SAS
MDE
AJ
FIN
AAC
CAS
CAH
Ancestries

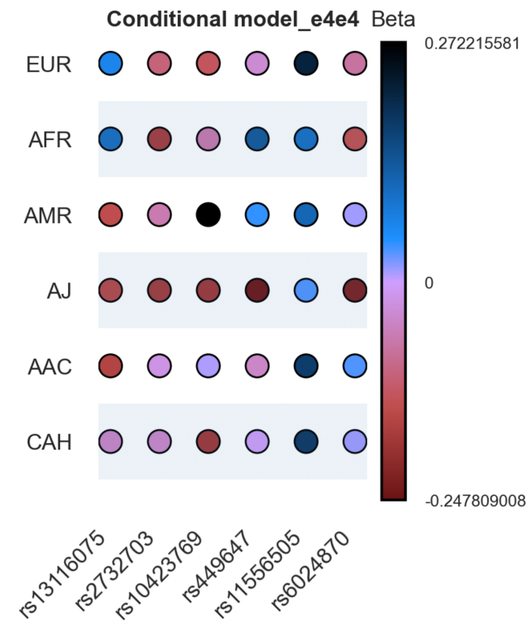

AD_e4
Related Dementias_e4
Controls_e4
SNPs
ABCA7:rs72973581
APP:rs466433
APP:rs364048
NOCT:rs13116075
SORL1:rs11218343
SLC24A4:rs12881735
CASS4:rs6024870
EPHA1:rs11762262
SPPL2A:rs59685680
APP:rs63750847
PLCG2:rs72824905
19q13.31:rs10423769
APOE:rs449647
FN1:rs140926439
FN1:rs116558455
RELN:rs201731543
TOMM40:rs11556505
RAB10:rs142787485
LRRC37A:rs2732703
NFIC:rs9749589
APOE3 Christchurch:rs121918393
EUR
AFR
AMR
EAS
SAS
MDE
AJ
FIN
AAC
CAS
CAH
Ancestry
Percentage
0
10
20
30
40
50
60
70
80
100

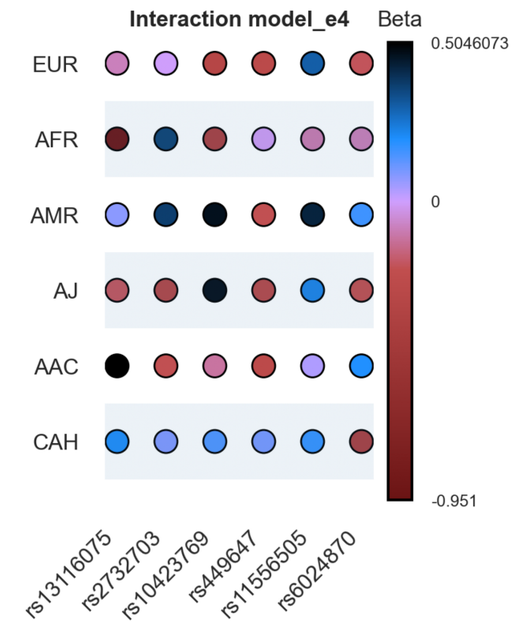

AD vs Control Frequency Ratio
ABCA7:rs72973581
APP:rs466433
APP:rs364048
NOCT:rs13116075
SORL1:rs11218343
SLC24A4:rs12881735
CASS4:rs6024870
EPHA1:rs11762262
SPPL2A:rs59685680
APP:rs63750847
PLCG2:rs72824905
19q13.31:rs10423769
APOE:rs449647
FN1:rs140926439
FN1:rs116558455
RELN:rs201731543
TOMM40:rs11556505
RAB10:rs142787485
LRRC37A:rs2732703
NFIC:rs9749589
APOE3 Christchurch:rs121918393
SNPs
EUR
AFR
AMR
EAS
SAS
MDE
AJ
FIN
AAC
CAS
CAH
Ancestry
inf
8.0
4.0
2.0
1.0
0.5
0.25
0.125
0.0625
0.03125
0
N/A

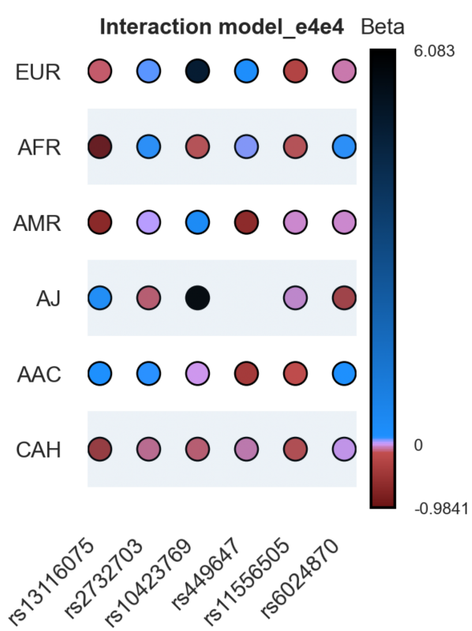

Related Dementias vs Control Frequency Ratio
ABCA7:rs72973581
APP:rs466433
APP:rs364048
NOCT:rs13116075
SORL1:rs11218343
SLC24A4:rs12881735
CASS4:rs6024870
EPHA1:rs11762262
SPPL2A:rs59685680
APP:rs63750847
PLCG2:rs72824905
19q13.31:rs10423769
APOE:rs449647
FN1:rs140926439
FN1:rs116558455
RELN:rs201731543
TOMM40:rs11556505
RAB10:rs142787485
LRRC37A:rs2732703
NFIC:rs9749589
APOE3 Christchurch:rs121918393
SNPs
EUR
AFR
AMR
EAS
SAS
MDE
AJ
FIN
AAC
CAS
CAH
Ancestry
inf
8.0
4.0
2.0
1.0
0.5
0.25
0.125
0.0625
0.03125
0
N/A

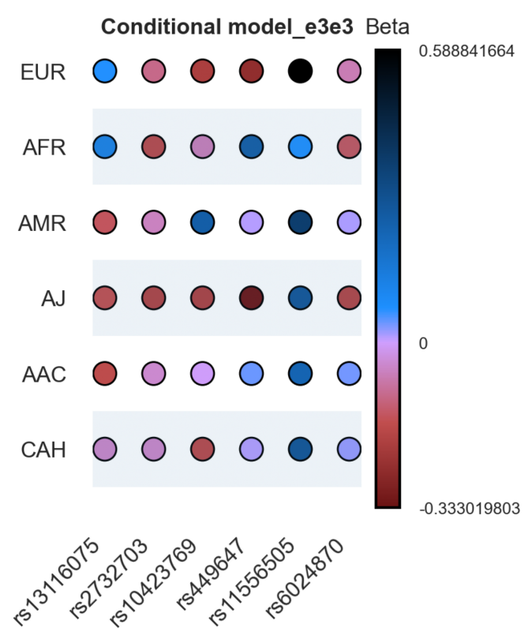

Protective model
Beta
0.828739573
0
-0.397822884
EUR
AFR
AMR
AJ
AAC
CAH
rs13116075
rs2732703
rs10423769
rs449647
rs11556505
rs6024870

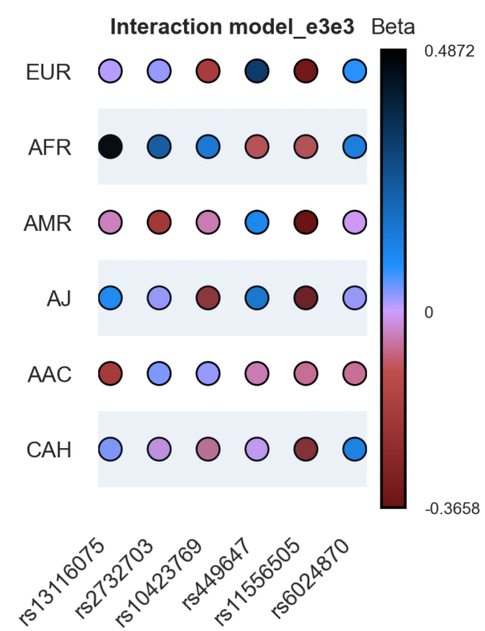

Conditional model_e4
Beta
0.272215581
0
-0.247809008
EUR
AFR
AMR
AJ
AAC
CAH
rs13116075
rs2732703
rs10423769
rs449647
rs11556505
rs6024870


Conditional model_e4e4
Beta
0.272215581
0
-0.247809008
EUR
AFR
AMR
AJ
AAC
CAH
rs13116075
rs2732703
rs10423769
rs449647
rs11556505
rs6024870


Interaction model_e4
Beta
0.5046073
0
-0.951
EUR
AFR
AMR
AJ
AAC
CAH
rs13116075
rs2732703
rs10423769
rs449647
rs11556505
rs6024870


Interaction model_e4e4
Beta
6.083
0
-0.9841
EUR
AFR
AMR
AJ
AAC
CAH
rs13116075
rs2732703
rs10423769
rs449647
rs11556505
rs6024870


Conditional model_e3e3
Beta
0.588841664
0
-0.333019803
EUR
AFR
AMR
AJ
AAC
CAH
rs13116075
rs2732703
rs10423769
rs449647
rs11556505
rs6024870


Interaction model_e3e3
Beta
0.4872
0
-0.3658
EUR
AFR
AMR
AJ
AAC
CAH
rs13116075
rs2732703
rs10423769
rs449647
rs11556505
rs6024870

Overview Tables

Table 1- Discovery phase: Multi-ancestry summary of known potential disease-causing variants only present in Alzheimer's disease and related dementia cases in AoU, 100KGP and UKB
Table 2- Discovery phase: Multi-ancestry summary of novel potential disease-causing variants only present in Alzheimer's disease and related dementia cases in AoU, 100KGP and UKB
Table 3- Replication phase: potential disease-causing variants only present in Alzheimer's disease and related dementia cases in ADSP
Table 4- Replication phase: potential disease-causing variants only present in Alzheimer's disease and related dementia cases in AMP PD
Table 5- Multi-ancestry summary of *APOE* genotypes in Alzheimer's disease and related dementia cases and controls in AoU, ADSP, UKB, AMP PD and 100KGP
Table 6- Summary of the analysis of protective/disease-modifying variants in Alzheimer's disease cases and controls in the ADSP
Supplementary Table 1- Discovery phase: Multi-ancestry summary of all variants identified in Alzheimer's disease and related dementia cases in AoU
Supplementary Table 2- Discovery phase: Multi-ancestry summary of all variants identified in Alzheimer's disease and related dementia cases in UKB
Supplementary Table 3- Discovery phase: Multi-ancestry summary of all variants identified in Alzheimer's disease and related dementia cases in 100KGP
Supplementary Table 4- Multi-ancestry summary of variants previously reported as potential disease-causing but identified in controls in multiple databases in this study
Supplementary Table 5- Phenotypic data for all individuals carrying known and novel potential disease-causing variants in the discovery phase
Supplementary Table 6- Phenotypic data for all individuals carrying known and novel potential disease-causing variants in the replication phase
Supplementary Table 7- Multi-ancestry summary of individuals carrying both *APOE* genotypes and protective or disease-modifying variants in patients and controls in AoU
Supplementary Table 8- Multi-ancestry summary of individuals carrying both *APOE* genotypes and protective or disease-modifying variants in patients and controls in ADSP
Supplementary Table 9- Multi-ancestry summary of individuals carrying both *APOE* genotypes and protective or disease-modifying variants in patients and controls in AMP PD
Supplementary Table 10- Multi-ancestry summary of individuals carrying both *APOE* genotypes and protective or disease-modifying variants in patients and controls in 100KGP
Supplementary Table 11- Multi-ancestry summary of individuals carrying both *APOE* genotypes and protective or disease-modifying variants in patients and controls in UKB
Supplementary Table 12- Combined results of data for individuals carrying both *APOE* genotypes and protective or disease-modifying variants in patients and controls across AoU, UKB, ADSP, 100KGP, and AMP PD
Supplementary Table 13- Assessment of the protective model in ADSP
Supplementary Table 14- Assessment of the conditional model for *APOE* ε4 carriers in ADSP
Supplementary Table 15- Assessment of the conditional model for *APOE* ε4/ε4 genotype in ADSP
Supplementary Table 16- Assessment of the correlation model for *APOE* ε4 carriers in ADSP
Supplementary Table 17- Assessment of the correlation model for *APOE* ε4/ε4 genotype in ADSP
Supplementary Table 18- Assessment of the interaction model for *APOE* ε4 carriers in ADSP
Supplementary Table 19- Assessment of the interaction model for *APOE* ε4/ε4 genotype in ADSP
Supplementary Table 20- Assessment of the conditional model for *APOE* ε3/ε3 genotype in ADSP
Supplementary Table 21- Assessment of the correlation model for *APOE* ε3/ε3 genotype in ADSP
Supplementary Table 22- Assessment of the interaction model for *APOE* ε3/ε3 genotype in ADSP

Table 1- Discovery phase: Multi-ancestry summary of known potential disease-causing variants only present in Alzheimer's disease and related dementia cases in AoU, 100KGP and UKB

	Gene	Position	rs ID/ClinVar ID	cDNA changes	Protein change/Splicing	Clinical significance	HGMD/Disease reported	CADD	Genetic ancestry	Zygosity	GnomAD
AoU											
	APP	chr21: 25891796*	rs63750066	C>T	p.A713T	Pathogenic, Likely pathogenic, VUS	CM930033, AD	26.9	EUR	Het	3.64E-05
	PSEN1	**chr14:73170945**	**rs63749824**	**C>T**	**p.A79V**	**Pathogenic, Likely pathogenic**	**CM981649, AD**	**26.1**	**AFR**	**Het**	**1.34E-05**
	MAPT	**chr17:46024061**	**rs63750424**	**C>T**	**p.R406W**	**Pathogenic,VUS**	**CM981237, FTD with parkinsonism**	**23.9**	**EUR**	**Het**	**1.98E-05**
	GRN	chr17:44350757	rs777704177	G>A	p.C222Y	VUS	CM149714, AD	28.9	AMR	Het	2.24E-05
100KGP											
	PSEN1	**chr14:73198067**	**rs63750900**	**G>A**	**p.R269H**	**Pathogenic, Likely pathogenic**	**CM971254, AD**	**29.9**	**EUR**	**Het**	**8.48E-06**
UKB											
	GBA1	chr1:155237446	rs1671825414	G>T	p.F298L	Likely pathogenic	CM000164, Gaucher disease 2	23.6	EUR	Het	6.78E-06
	GBA1	chr1:155238206	rs381427	A>C	p.V230G	Pathogenic/VUS	CM980833, Gaucher disease	22.5	EUR	Het	3.39E-06
	GBA1	chr1:155238228	rs61748906	A>G	p.W223R	Pathogenic/Likely pathogenic/VUS	CM001166, Gaucher disease 2	28	EUR	Het	1.19E-05
	GBA1	chr1:155238302	rs80222298	G>A	p.P198L	Likely pathogenic, VUS	CM980827, Gaucher disease	28.6	EUR	Het	8.62E-07
	PSEN1	**chr14:73170945**	**rs63749824**	**C>T**	**p.A79V**	**Pathogenic, Likely pathogenic**	**CM981649, AD**	**26.1**	**EUR**	**Het**	**1.10E-05**
	PSEN1	chr14:73192832	rs63750526	C>A	p.A246E	Pathogenic, Likely pathogenic	CM951075, AD	25.4	EUR	Het	8.48E-07
	PSEN1	chr14:73198061	rs63750779	C>T	p.P267L	Likely pathogenic	CM033803, AD	25.8	EUR	Het	8.49E-07
	PSEN1	**chr14:73198067**	**rs63750900**	**G>A**	**p.R269H**	**Pathogenic, Likely pathogenic**	**CM971254, AD**	**29.9**	**EUR**	**Het**	**8.48E-06**
	GRN	chr17:44350449	VCV001922048.3	->CTGTGAAGACAGGGTGCACTGCTGT	p.P166fs*	Pathogenic	Not reported, FTD	34	EUR	Het	8.48E-07
	GRN	chr17:44350801	rs63749817	G>A	c.708+1G>A	Pathogenic/Likely pathogenic	CS200794, FTD	34	EUR	Het	3.42E-06
	GRN	chr17:44351438	rs63751177	G>A	p.W304X	Pathogenic	CM064045, FTD - CM188618, FTLD	39	AJ	Het	0
	GRN	chr17:44352087	rs63751180	C>T	p.R418X	Pathogenic	CM062773, FTD	25.6	EUR	Het	4.24E-06
	GRN	chr17:44352404	rs63751294	C>T	p.R493X	Pathogenic	CM064044, FTD	36	EUR	Het	1.53E-05
	MAPT	**chr17:46024061**	**rs63750424**	**C>T**	**p.R406W**	**Pathogenic, VUS**	**CM981237, FTD with parkinsonism**	**23.9**	**EUR**	**Het**	**1.86E-05**
	APP	chr21:25891784	rs63750264	C>T	p.V717L	Pathogenic/Likely pathogenic	CM003587, AD	26.8	EUR	Het	1.70E-06
	APP	chr21:25891856	rs63750579	C>G	p.E693Q	Pathogenic/Likely pathogenic	CM920067, AD	27.1	EUR	Het	-

Key: cDNA, complementary DNA; VUS, variant uncertain significance; Het, Heterozygous; CADD,Combined Annotation Dependent Depletion; AD, Alzheimer's disease; FTD, frontotemporal dementia; Clinical significance based on dbSNP, ClinVar, and ACMG guideline. AoU, All of Us; 100KGP, 100,000 Genomes Project; UKB, UK Biobank; EUR, European; AFR, African; AMR, American Admixed; AAC, African Admixed; AJ, Ashkenazi Jewish; CAS, Central Asian; EAS, Eastern Asian; SAS, South Asian; MDE, Middle Eastern; FIN, Finnish; CAH, Complex Admixture History. HGMD; Human Gene Mutation Database, The frequency of gnomAD refers to the frequency in the ancestry where the variations were found. Bold variants were replicated in different databases. Disease reported refers to the disease for which the variants were previously reported.

*, This variant was not replicated only in cases across the diverse biobanks in the discovery phase. Position refers to GRCh 38.

Table 2- Discovery phase: Multi-ancestry summary of novel potential disease-causing variants only present in Alzheimer's disease and related dementia cases in AoU, 100KGP and UKB

	Gene	Position	rs ID/ClinVar ID	cDNA changes	Protein change/Splicing	Clinical significance	CADD	PP2 prediction	Genetic ancestry	Zygosity	GnomAD
AoU											
	APP	chr21:25881764	Novel	G>A	p.A740V	VUS	26.7	Probably Damaging	AJ	Het	0
	APP	chr21:25911860	rs765301301	A>C	p.L597W	VUS	25.7	Probably Damaging	AFR	Het	1.47E-04
	APP	chr21:26021894	Novel	T>C	p.S271G	VUS	24.5	Benign	EUR	Het	0
	APP	chr21:25982349	rs2042464223	A>T	p.S407T	VUS	24.2	Probably Damaging	EUR	Het	3.60E-06
	APP	chr21:25982445*	rs141331202	C>T	p.V375I	VUS	25.7	Probably Damaging	EUR	Het	3.22E-05
	APP	chr21:25982457	rs747438691	T>C	p.T371A	VUS	24.6	Possibly Damaging	AMR	Het	4.57E-05
	APP	chr21:25997360	rs749453173	G>A	p.L364F	VUS	23.9	Benign	AFR	Het	1.74E-04
	PSEN1	chr14:73170869	Novel	C>T	p.R54X	Pathogenic	36	-	AFR	Het	2.70E-06
	PSEN2	chr1:226890133	rs1410382029	T>C	p.S296P	VUS	33	Probably Damaging	EUR	Het	0
	PSEN2	chr1:226891797	Novel	CT>C	p.L344X	Pathogenic	23.9	-	AMR	Het	0
	PSEN2	chr1:226888098	rs1661490243	A>G	p.H169R	VUS	26.2	Probably Damaging	EUR	Het	8.48E-07
	PSEN2	chr1:226890124	rs199689738	A>T	p.I293L	VUS	27.1	Possibly Damaging	EUR	Het	4.16E-05
	PSEN2	**chr1:226891817**	**rs759669954**	**G>A**	**p.G349R**	**VUS**	**22.1**	**Benign**	**AAC**	**Het**	**8.45E-05**
	GRN	chr17:44350296	rs1248058567	T>C	p.C140R	VUS	24.2	Probably Damaging	EUR	Het	5.71E-06
	GRN	chr17:44350735	rs1201429668	T>C	p.C215R	VUS	29.5	Probably Damaging	EUR	Het	4.78E-06
	GRN	chr17:44351575	Novel	A>G	p.H320R	VUS	26.2	Probably Damaging	EUR	Het	2.86E-06
	GRN	chr17:44352682	rs63750116	C>T	p.R556C	VUS	24.7	Probably Damaging	EUR	Het	7.19E-06
	MAPT	chr17:45974472	rs747085337	G>A	g.45974472G>A	-	23.3	-	EUR	Het	6.33E-06
	MAPT	chr17:45983788	Novel	AGGGGCCCCTGGAGAGGGGCCAGAGGCCC>A	p.G332LfsX64	-	27.3	-	EUR	Het	0
	MAPT	chr17:46018716	rs948573449	G>A	p.G701R	VUS	34	Probably Damaging	EUR	Het	8.10E-06
	MAPT	chr17:46024088*	rs768841567	G>A	p.G750S	VUS	33	Probably Damaging	EUR	Het	2.20E-05
	TREM2	chr6:41161502	rs369181900	C>T	p.C51Y	VUS	27.9	Probably Damaging	EUR	Het	1.44E-05
	TREM2	chr6:41161523	Novel	C>T	p.W44X	Pathogenic	37	-	AAC	Het	0
	GBA1	chr1:155239657	rs759174705	G>T	p.P138H	VUS	22.3	Benign	EUR	Het	2.54E-06
	GBA1	chr1:155239762	rs748485792	C>T	p.G103D	VUS	20.4	Benign	EUR	Het	3.60E-06
	SNCA	chr4:89822256	rs757477802	T>C	p.Q99R	VUS	21.9	Benign	EUR	Het	1.36E-05
100KGP											
	APP	chr21:25954665*	rs779792929	A>G	p.Y407H	VUS	25.6	Probably Damaging	EUR	Het	8.31E-05
	PSEN2	chr1:226891349	rs565698726	G>A	p.D320N	VUS	21.8	Benign	EUR	Het	5.93E-06
UKB											
	GBA1	chr1:155235769	rs747284798	G>A	p.R347C	VUS	31	Probably damaging	EUR	Het	8.47E-06
	GBA1	chr1:155236249	rs1057519358	A>G	p.I320T	VUS	25.3	Probably damaging	EUR	Het	7.63E-06
	GBA1	chr1:155236262	Novel	T>G	p.S316R	VUS	27.7	Probably damaging	EUR	Het	2.54E-06
	GBA1	chr1:155236471	Novel	T>C	c.1000-2A>G	-	31	-	EUR	Het	-
	GBA1	chr1:155237564	Novel	T>C	p.Y172C	VUS	28	Probably damaging	EUR	Het	4.24E-06
	GBA1	chr1:155237579	Novel	G>A	c.762-1G>A	-	26.8	-	EUR	Het	-
	GBA1	chr1:155239639	Novel	A>G	p.L57P	Likely pathogenic	28.2	Probably damaging	EUR	Het	8.47E-07
	GBA1	chr1:155239685	rs1671971599	C>T	p.A42T	VUS	21.1	Possibly damaging	EUR	Het	0.00E+00
	PSEN2	chr1:226885581	rs1363866270	C>T	p.R134C	VUS	32	Probably damaging	EUR	Het	1.70E-06
	PSEN2	chr1:226885632	Novel	G>T	p.V151F	VUS	20.8	Benign	EUR	Het	8.48E-07
	PSEN2	chr1:226888846	Novel	A>T	p.Y195F	VUS	28.2	Possibly damaging	EUR	Het	7.63E-06

PSEN2	chr1:226888864	rs200410369	A>G	p.Y201C	VUS	28.5	Probably damaging	EUR	Het	1.27E-05
PSEN2	chr1:226891284	rs1482790603	T>C	p.M298T	VUS	25.3	Possibly damaging	EAS	Het	1.34E-04
PSEN2	chr1:226891344	Novel	C>A	p.P318H	VUS	23.7	Possibly damaging	EUR	Het	1.70E-06
PSEN2	chr1:226891347	Novel	A>-	p.Y319fs	-	27.9	-	EUR	Het	-
PSEN2	chr1:226891809	rs1365789341	G>A	p.G346D	VUS	20.3	Benign	EUR	Het	8.47E-07
PSEN2	**chr1:226891817**	**rs759669954**	**G>A**	**p.G349R**	**VUS**	**22.1**	**Benign**	**EUR**	**Het**	**0**
PSEN2	chr1:226891844	Novel	A>G	p.R358G	VUS	23.4	Benign	EUR	Het	8.48E-07
PSEN1	chr14:73170851	rs1377702483	G>A	p.E48K	VUS	23.1	Possibly damaging	EUR	Het	2.54E-06
PSEN1	chr14:73170998	rs63750852	G>A	p.V97M	VUS	28.3	Probably damaging	EUR	Het	7.63E-06
PSEN1	chr14:73170999	rs1356498068	T>C	p.V97A	VUS	25.6	Probably damaging	EUR	Het	8.47E-07
PSEN1	chr14:73186896	rs63750771	T>C	p.F175S	VUS	23.7	Probably damaging	EUR	Het	8.48E-07
PSEN1	chr14:73192754	rs763831389	G>A	p.R220Q	VUS	23.8	Possibly damaging	EUR	Het	5.93E-06
PSEN1	chr14:73198040	Novel	C>G	p.A260G	VUS	26.8	Probably damaging	EUR	Het	8.57E-07
PSEN1	chr14:73198052	Novel	C>G	p.P264R	Likely pathogenic	25.7	Probably damaging	EUR	Het	8.50E-07
PSEN1	chr14:73206388	rs63750298	A>G	p.T291A	VUS	27.8	Possibly damaging	EUR	Het	5.93E-06
PSEN1	chr14:73211836	Novel	->GCCC	p.E341fs	-	34	-	EUR	Het	-
PSEN1	chr14:73217129	rs63750323	G>C	p.G378A	Likely pathogenic	25.4	Probably damaging	EUR	Het	5.09E-06
PSEN1	chr14:73219155	rs1555358260	C>T	p.L424F	Likely pathogenic	25.1	Probably damaging	EUR	Het	-
PSEN1	chr14:73219188	Novel	C>A	p.L435I	Likely pathogenic	25.8	Probably damaging	EUR	Het	1.70E-06
PSEN1	chr14:73219194	rs764971634	A>G	p.I437V	Likely pathogenic, VUS	23	Benign	EUR	Het	6.78E-06
PSEN1	chr14:73219254	rs1430581353	A>G	p.M457V	VUS	23.5	Probably damaging	EUR	Het	1.70E-06
GRN	chr17:44349248	rs63751057	->GCCT	p.V28fs	-	33	-	EUR	Het	4.24E-06
GRN	chr17:44349529	rs63751193	C>-	p.S81fs	Pathogenic	32	-	EUR	Het	8.47E-07
GRN	chr17:44350291	rs146769257	C>A	p.T138K	VUS	24.6	Probably damaging	AFR	Het	4.01E-05
GRN	chr17:44350303	Novel	->GGTC	p.M142fs	-	26.1	-	EUR	Het	-
GRN	chr17:44350553	Novel	C>T	p.P192S	VUS	24.1	Probably damaging	EUR	Het	1.70E-06
GRN	chr17:44350801	Novel	G>T	c.708+1G>T	-	32	-	EUR	Het	1.71E-06
GRN	chr17:44351082	Novel	->TG	p.V252fs	-	25.8	-	EUR	Het	-
GRN	chr17:44351610	Novel	A>G	p.K332E	VUS	20.3	Benign	EUR	Het	3.39E-06
GRN	chr17:44351663	Novel	->G	p.P349fs	-	23	-	EUR	Het	8.48E-07
GRN	chr17:44352025	Novel	G>A	p.C397Y	VUS	26.4	Probably damaging	EUR	Het	8.48E-07
GRN	chr17:44352249	Novel	G>A	c.1413+1G>A	-	35	-	EUR	Het	8.49E-07
GRN	chr17:44352395	rs886053006	G>A	p.V490M	VUS	25.4	Probably damaging	EUR	Het	3.39E-06
MAPT	chr17:45962447	rs966689443	G>C	p.G37A	VUS	21.6	Probably damaging	EUR	Het	3.56E-05
MAPT	chr17:45978420	rs139796158	C>G	p.A60G	VUS	25.1	Probably damaging	AFR	Het	5.64E-04
MAPT	chr17:45978422	Novel	G>T	p.A61S	VUS	23.9	Probably damaging	EUR	Het	1.70E-06
MAPT	chr17:45982886	rs940936590	C>T	p.R103W	-	20.5	-	EUR	Het	1.93E-05
MAPT	chr17:45983453	rs2073193780	G>A	p.E292K	VUS	23.1	Possibly damaging	EUR	Het	8.50E-07
MAPT	chr17:45983504	Novel	C>-	p.P309fs	-	22.5	-	EUR	Het	-
MAPT	chr17:45996504	rs779901466	G>A	p.R163Q	VUS	29.3	Probably damaging	EUR	Het	8.83E-06
MAPT	chr17:45996630	Novel	C>T	p.T205I	VUS	27.5	Probably damaging	EUR	Het	8.48E-07
MAPT	chr17:46010394	Novel	G>A	p.G245S	VUS	33	Probably damaging	EUR	Het	3.46E-06
MAPT	chr17:46018621	Novel	G>A	p.G245D	VUS	33	Probably damaging	EUR	Het	8.51E-07
MAPT	chr17:46018639	Novel	A>G	p.K251R	VUS	24.7	Probably damaging	EUR	Het	-
MAPT	chr17:46024019	rs991713081	A>G	p.I303V	VUS	26.2	Probably damaging	EUR	Het	4.24E-06
APP	chr21:25891742	Novel	T>G	p.I600L	VUS	25.4	-	EUR	Het	5.09E-06
APP	chr21:25905045	rs768182065	G>C	p.R517G	VUS	25.8	Probably damaging	SAS	Het	3.30E-05
APP	chr21:25905048	rs368159818	C>T	p.D516N	VUS	27	Probably damaging	EUR	Het	5.09E-06
APP	chr21:25911879	rs201874897	C>T	p.D460N	VUS	26.9	Probably damaging	EUR	Het	5.09E-06
APP	chr21:25911885	rs755645885	C>T	p.G458R	VUS	27.8	Probably damaging	EUR	Het	1.19E-05
APP	chr21:25954659	rs200500889	G>A	p.R409C	VUS	32	Probably damaging	EUR	Het	5.93E-06

APP	chr21:25982424	rs752243493	G>A	p.P251S	VUS	26.4	Probably damaging	EUR	Het	1.10E-05
APP	chr21:26000138	rs200539466	T>C	p.I248V	VUS	23.1	Probably damaging	EUR	Het	1.27E-05
APP	chr21:26021858	rs772069024	C>G	p.V227L	VUS	25.4	Probably damaging	EUR	Het	8.48E-07
APP	chr21:26021912	rs754672142	C>T	p.A209T	VUS	20.9	Probably damaging	EUR	Het	4.24E-06
APP	chr21:26051100	rs199744129	G>A	p.P132S	VUS	27.9	Probably damaging	EUR	Het	1.70E-06
SNCA	chr4:89726638	rs746232417	G>T	p.P90H	Likely pathogenic	24.6	Probably damaging	EUR	Het	2.55E-06
SNCA	chr4:89822281	Novel	C>A	p.A91S	Likely pathogenic	23.8	Probably damaging	EUR	Het	1.70E-06
TREM2	chr6:41161292	Novel	A>C	p.L121R	VUS	27	Probably damaging	EUR	Het	6.78E-06
TREM2	chr6:41161343	Novel	T>A	p.D104V	VUS	23.7	Probably damaging	EUR	Het	8.47E-07

Key: cDNA, complementary DNA; VUS, variant uncertain significance; Het, Heterozygous; CADD,Combined Annotation Dependent Depletion; PP2,PolyPhen-2; Clinical significance based on dbSNP, ClinVar, and ACMG guideline. Position refers to GRCh 38.

AoU, All of Us; 100KGP, 100,000 Genomes Project; UKB, UK Biobank; EUR, European; AFR, African; AMR, American Admixed; AAC,African Admixed; AJ, Ashkenazi Jewish; CAS, Central Asian; EAS, Eastern Asian; SAS, South Asian; MDE, Middle Eastern; FIN, Finnish; CAH, Complex Admixture History. The frequency of gnomAD refers to the frequency in the ancestry where the variants were found. Bold variants were replicated in different databases. Novel in the title means it has not been reported for the disease.

*, These variants were not replicated only in cases across the diverse biobanks in the discovery phase.

Table 3- Replication phase: potential disease-causing variants only present in Alzheimer's disease and related dementia cases in ADSP

Cases

Variant status	Gene	Position	Protein change	rs ID	Number of AD Cases	AF in AD Cases (n=10566)	AF(AAC)-AD (n=1119)	AF(AMR)-AD (n=1086)	AF(FIN)-AD (n=11)	AF(CAS)-AD (n=20)	AF(MDE)-AD (n=27)	AF(AFR)-AD (n=664)	AF(EAS)-AD (n=50)	AF(AJ)-AD (n=537)	AF(SAS)-AD (n=176)	AF(EUR)-AD (n=5812)	AF(CAH)-AD (n=1064)	MISSING_CT	OBS_CT	F_MISS	ALT_FREQS	OBS_CT	Reported in other diseases
AoU																							
Known	*APP*	chr21: 25891796:C:T	p.A713T	rs63750066	1	4.73E-05	0	0	0	0	0	0	0	0	0	8.61E-05	0	11	36361	3.03E-04	8.25E-05	72700	1 DLB
Known	*PSEN1*	chr14:73170945:C:T	p.A79V	rs63749824	5	2.37E-04	0	0	0	0	0	0	0	0	0	4.30E-04	0	8	36361	2.20E-04	1.24E-04	72706	1 Possible AD
Novel	*APP*	chr21:25911860:A:C	p.L597W	rs765301301	3	1.42E-04	8.94E-04	0	0	0	0	0	0	0	0	0	4.70E-04	2	36361	5.50E-05	4.13E-05	72718	0
Novel	*MAPT*	chr17:46018716:G:A	p.G701R	rs948573449	1	4.73E-05	0	0	0	0	0	0	0	0	0	8.61E-05	0	6	36361	1.65E-04	1.38E-05	72710	0
Novel	*MAPT*	chr17:46024088:G:A	p.G750S	rs768841567	0	0	0	0	0	0	0	0	0	0	0	0	0	16	36361	4.40E-04	2.75E-05	72690	0
Novel	*SNCA*	chr4:89822256:T:C	p.Q99R	rs757477802	1	4.73E-05	0	0	0	0	0	0	0	0	0	8.61E-05	0	5	36361	1.38E-04	2.75E-05	72712	0
100KGP																							
Known	*PSEN1*	chr14:73198067:G:A	p.R269H	rs63750900	5	2.37E-04	8.94E-04	0	0	0	0	0	0	0	0	2.58E-04	0	9	36361	2.48E-04	1.10E-04	72704	1 MCI
Novel	*PSEN2*	chr1:226891349:G:A	p.D320N	rs565698726	0	0	0	0	0	0	0	0	0	0	0	0	0	15	36361	4.13E-04	6.88E-05	72692	0
Novel	*APP*	chr21:25954665:A:G	p.Y407H	rs779792929	2	9.46E-05	0	0	0	0	0	0	0	0	0	1.72E-04	0	7	36361	1.93E-04	5.50E-05	70708	0
UKB																							
Known	*GBA1*	chr1:155238228:A:G	p.W223R	rs61748906	1	4.73E-05	0	0	0	0	0	0	0	0	0	0	4.70E-04	7	36361	1.93E-04	5.50E-05	72708	1 PSP
Novel	*PSEN2*	chr1:226891817:G:A	p.G349R	rs759669954	0	0	0	0	0	0	0	0	0	0	0	0	0	15	36361	4.13E-04	1.38E-05	72692	0
Known	*PSEN1*	chr14:73170945:C:T	p.A79V	rs63749824	5	2.37E-04	0	0	0	0	0	0	0	0	0	4.30E-04	0	8	36361	2.20E-04	1.24E-04	72706	1 Possible AD
Novel	*PSEN1*	chr14:73192754:G:A	p.R220Q	rs763831389	3	1.42E-04	0	0	0	0	0	7.53E-04	0	0	0	1.72E-04	0	6	36361	1.65E-04	5.50E-05	72710	0
Known	*PSEN1*	chr14:73198067:G:A	p.R269H	rs63750900	5	2.37E-04	8.94E-04	0	0	0	0	0	0	0	0	2.58E-04	0	9	36361	2.48E-04	1.10E-04	72704	1 MCI
Novel	*PSEN1*	chr14:73206388:A:G	p.T291A	rs63750298	1	4.73E-05	4.47E-04	0	0	0	0	0	0	0	0	0	0	5	36361	1.38E-04	6.88E-05	72712	0
Novel	*PSEN1*	chr14:73219194:A:G	p.I437V	rs764971634	1	4.73E-05	0	0	0	0	0	0	0	0	0	8.60E-05	0	0	36361	0	1.38E-05	72722	0
Novel	*PSEN1*	chr14:73219254:A:G	p.M457V	rs1430581353	1	4.73E-05	0	0	0	0	0	0	0	0	0	8.60E-05	0	2	36361	5.50E-05	1.38E-05	72718	0
Novel	*GRN*	chr17:44352395:G:A	p.V490M	rs886053006	1	4.73E-05	4.47E-04	0	0	0	0	0	0	0	0	0	0	13	36361	3.58E-04	1.38E-05	72696	0
Known	*GRN*	chr17:44352404:C:T	p.R493X	rs63751294	3	1.42E-04	0	0	0	0	0	0	0	0	0	2.58E-04	0	20	36361	5.50E-04	6.88E-05	72682	0
Novel	*MAPT*	chr17:45978420:C:G	p.A60G	rs139796158	2	9.46E-05	8.98E-04	0	0	0	0	0	0	0	0	0	0	416	36361	1.14E-02	1.67E-04	71890	1 MCI
Novel	*MAPT*	chr17:45982886:C:T	p.R103W	rs940936590	0	0	0	0	0	0	0	0	0	0	0	0	0	8	36361	2.20E-04	2.75E-05	72706	0
Known	*MAPT*	chr17:46024061:C:T	p.R406W	rs63750424	4	1.89E-04	0	0	0	0	0	0	0	0	0	3.44E-04	0	9	36361	2.48E-04	9.63E-05	72704	0
Novel	*APP*	chr21:25905048:C:T	p.D516N	rs368159818	0	0	0	0	0	0	0	0	0	0	0	0	0	8	36361	2.20E-04	2.75E-05	72706	0
Novel	*APP*	chr21:25982424:G:A	p.P251S	rs752243493	0	0	0	0	0	0	0	0	0	0	0	0	0	12	36361	3.30E-04	1.38E-05	72698	0
Novel	*APP*	chr21:26021912:C:T	p.A209T	rs754672142	0	0	0	0	0	0	0	0	0	0	0	0	0	14	36361	3.85E-04	1.38E-05	72694	1 PSP

Controls

Variant status	Gene	Position	Protein change	rs ID	Number of Controls	AF in Controls (n=16217)	AF(AAC)-Controls (n=1563)	AF(AMR)-Controls (n=3967)	AF(FIN)-Controls (n=12)	AF(CAS)-Controls (n=131)	AF(MDE)-Controls (n=9)	AF(AFR)-Controls (n=1685)	AF(EAS)-Controls (n=29)	AF(AJ)-Controls (n=459)	AF(SAS)-Controls (n=2279)	AF(EUR)-Controls (n=4411)	AF(CAH)-Controls (n=1672)	MISSING_CT	OBS_CT	F_MISS	ALT_FREQS	OBS_CT
AoU																						
Known	*APP*	chr21: 25891796:C:T	p.A713T	rs63750066	2	6.17E-05	0	1.26E-04	0	0	0	0	0	0	0	1.13E-04	0	11	36361	3.03E-04	8.25E-05	72700
Known	*PSEN1*	chr14:73170945:C:T	p.A79V	rs63749824	1	3.08E-05	0	0	0	0	0	0	0	0	0	1.13E-04	0	8	36361	2.20E-04	1.24E-04	72706
Novel	*APP*	chr21:25911860:A:C	p.L597W	rs765301301	0	0	0	0	0	0	0	0	0	0	0	0	0	2	36361	5.50E-05	4.13E-05	72718
Novel	*MAPT*	chr17:46018716:G:A	p.G701R	rs948573449	0	0	0	0	0	0	0	0	0	0	0	0	0	6	36361	1.65E-04	1.38E-05	72710
Novel	*MAPT*	chr17:46024088:G:A	p.G750S	rs768841567	1	3.08E-05	0	1.26E-04	0	0	0	0	0	0	0	0	0	16	36361	4.40E-04	2.75E-05	72690
Novel	*SNCA*	chr4:89822256:T:C	p.Q99R	rs757477802	0	0	0	0	0	0	0	0	0	0	0	0	0	0				
100KGP																		5	36361	1.38E-04	2.75E-05	72712
Known	*PSEN1*	chr14:73198067:G:A	p.R269H	rs63750900	0	0	0	0	0	0	0	0	0	0	0	0	0	9	36361	2.48E-04	1.10E-04	72704
Novel	*PSEN2*	chr1:226891349:G:A	p.D320N	rs565698726	5	1.54E-04	3.20E-04	1.26E-04	0	0	0	0	0	0	0	1.13E-04	5.98E-04	15	36361	4.13E-04	6.88E-05	72692
Novel	*APP*	chr21:25954665:A:G	p.Y407H	rs779792929	1	3.08E-05	0	0	0	0	0	0	0	0	0	1.13E-04	0	7	36361	1.93E-04	5.50E-05	70708
UKB																						
Known	*GBA*	chr1:155238228:A:G	p.W223R	rs61748906	1	3.08E-05	0	0	0	0	0	2.97E-04	0	0	0	0	0	7	36361	1.93E-04	5.50E-05	72708
Novel	*PSEN2*	chr1:226891817:G:A	p.G349R	rs759669954	1	3.08E-05	0	1.26E-04	0	0	0	0	0	0	0	0	0	15	36361	4.13E-04	1.38E-05	72692
Known	*PSEN1*	chr14:73170945:C:T	p.A79V	rs63749824	1	3.08E-05	0	0	0	0	0	0	0	0	0	1.13E-04	0	8	36361	2.20E-04	1.24E-04	72706
Known	*PSEN1*	chr14:73198067:G:A	p.R269H	rs63750900	0	0	0	0	0	0	0	0	0	0	0	0	0	9	36361	2.48E-04	1.10E-04	72704
Novel	*PSEN1*	chr14:73206388:A:G	p.T291A	rs63750298	3	9.25E-05	0	1.26E-04	0	0	0	0	0	0	0	0	5.98E-04	5	36361	1.38E-04	6.88E-05	72712
Novel	*PSEN1*	chr14:73219194:A:G	p.I437V	rs764971634	0	0	0	0	0	0	0	0	0	0	0	0	0	0	36361	0	1.38E-05	72722
Novel	*PSEN1*	chr14:73219254:A:G	p.M457V	rs1430581353	0	0	0	0	0	0	0	0	0	0	0	0	0	2	36361	5.50E-05	1.38E-05	72718
Novel	*GRN*	chr17:44352395:G:A	p.V490M	rs886053006	0	0	0	0	0	0	0	0	0	0	0	0	0	13	36361	3.58E-04	1.38E-05	72696
Known	*GRN*	chr17:44352404:C:T	p.R493X	rs63751294	0	0	0	0	0	0	0	0	0	0	0	0	0	20	36361	5.50E-04	6.88E-05	72682
Novel	*MAPT*	chr17:45978420:C:G	p.A60G	rs139796158	7	2.16E-04	3.23E-04	0	0	0	0	6.00E-04	0	0	0	0	1.21E-03	416	36361	1.14E-02	1.67E-04	71890
Novel	*MAPT*	chr17:45982886:C:T	p.R103W	rs940936590	1	3.08E-05	0	1.26E-04	0	0	0	0	0	0	0	0	0	8	36361	2.20E-04	2.75E-05	72706
Known	*MAPT*	chr17:46024061:C:T	p.R406W	rs63750424	1	3.08E-05	0	0	0	0	0	0	0	0	0	1.13E-04	0	9	36361	2.48E-04	9.63E-05	72704
Novel	*APP*	chr21:25905048:C:T	p.D516N	rs368159818	0	0	0	0	0	0	0	0	0	0	0	0	0	8	36361	2.20E-04	2.75E-05	72706
Novel	*APP*	chr21:25982424:G:A	p.P251S	rs752243493	1	3.08E-05	0	0	0	0	0	0	0	0	2.20E-04	0	0	12	36361	3.30E-04	1.38E-05	72698
Novel	*APP*	chr21:26021912:C:T	p.A209T	rs754672142	0	0	0	0	0	0	0	0	0	0	0	0	0	14	36361	3.85E-04	1.38E-05	72694

Key: AF, allele frequency; AD, Alzheimer's disease; PSP, Progressive Supranuclear Palsy; MCI, Mild Cognitive Impairment; DLB, Dementia with Lewy bodies; n, Number of individuals; MISSING_CT, Missing Count; OBS_CT, Observed Count; F_MISS, Fraction Missing; ALT_FREQS, Alternate Allele Frequencies; Position refers to GRCh 38.

Table 4- Replication phase: potential disease-causing variants only present in Alzheimer's disease and related dementia cases in AMP PD

Cases

Variant status	Gene	Position	rs ID	Number of DLB Cases	AF in DLB Cases (n=2530)	AF(AAC)-DLB (n=0)	AF(AMR)-DLB (n=0)	AF(FIN)-DLB (n=5)	AF(CAS)-DLB (n=0)	AF(MDE)-DLB (n=6)	AF(AFR)-DLB (n=0)	AF(EAS)-DLB (n=0)	AF(AJ)-DLB (n=113)	AF(SAS)-DLB (n=0)	AF(EUR)-DLB (n=2406)	AF(CAH)-DLB (n=0)
AoU																
Known	*APP*	chr21: 25891796:C:T	rs63750066	1	1.98E-04	0	0	0	0	0	0	0	0	0	2.08E-04	0
Novel	*MAPT*	chr17:46024088:G:A	rs768841567	1	1.98E-04	0	0	0	0	0	0	0	0	0	2.08E-04	0
UKB																
Novel	*GRN*	chr17:44352395:G:A	rs886053006	0	0	0	0	0	0	0	0	0	0	0	0	0
Known	*GRN*	chr17:44352404:C:T	rs63751294	1	1.98E-04	0	0	0	0	0	0	0	0	0	2.08E-04	0
Novel	*APP*	chr21:25905048:C:T	rs368159818	1	1.98E-04	0	0	0	0	0	0	0	0	0	2.08E-04	0

Controls

Variant status	Gene	Position	rs ID	Number of Controls	AF in Controls (n=3270)	AF(AAC)-Controls (n=40)	AF(AMR)-Controls (n=14)	AF(FIN)-Controls (n=5)	AF(CAS)-Controls (n=3)	AF(MDE)-Controls (n=4)	AF(AFR)-Controls (n=29)	AF(EAS)-Controls (n=5)	AF(AJ)-Controls (n=246)	AF(SAS)-Controls (n=1)	AF(EUR)-Controls (n=2919)	AF(CAH)-Controls (n=4)
AoU																
Known	*APP*	chr21: 25891796:C:T	rs63750066	1	1.53E-04	0	0	0	0	0	0	0	0	0	1.71E-04	0
Novel	*MAPT*	chr17:46024088:G:A	rs768841567	1	1.53E-04	0	0	0	0	0	0	0	0	0	1.71E-04	0
UKB																
Novel	*GRN*	chr17:44352395:G:A	rs886053006	1	1.53E-04	0	0	0	0	0	0	0	0	0	1.71E-04	0
Known	*GRN*	chr17:44352404:C:T	rs63751294	0	0	0	0	0	0	0	0	0	0	0	0	0
Novel	*APP*	chr21:25905048:C:T	rs368159818	1	1.53E-04	0	0	0	0	0	0	0	0	0	1.71E-04	0

Key: UKB, UK Biobank; AMP PD, Accelerating Medicines Partnership in Parkinson's Disease ; AoU, All of Us; Position refers to GRCh 38.

Table 5- Multi-ancestry summary of *APOE* genotypes in Alzheimer's disease and related dementia cases and controls in AoU, ADSP, UKB, AMP PD and 100KGP

AoU													
AD													
	Genotypes	**Number of participants (n, %)**	**EUR (n, %)**	**AFR (n, %)**	**AMR (n, %)**	**EAS (n, %)**	**SAS (n, %)**	**MDE (n, %)**	**AJ (n, %)**	**FIN (n, %)**	**AAC (n, %)**	**CAS (n, %)**	**CAH (n, %)**
	ε1/ε1	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%
	ε1/ε2	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%
	ε1/ε4	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%
	ε2/ε2	3, 0.55%	2, 0.6%	1, 3.12%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%
	ε2/ε3	46, 8.53%	26, 7.80%	2, 6.25%	4, 5.63%	2, 28.57%	0, 0%	0, 0%	5, 10%	0, 0%	7, 17.07%	0, 0%	0, 0%
	ε2/ε4 or ε1/ε3	11, 2.04%	5, 1.50%	2, 6.25%	1, 1.40%	0, 0%	0, 0%	0, 0%	1, 2%	0, 0%	2, 4.87%	0, 0%	0, 0%
	ε3/ε3	281, 52.13%	171, 51.35%	13, 40.62%	46, 64.78%	2, 28.57%	1, 50%	1, 100%	26, 52%	0, 0%	20, 48.78%	1, 50%	0, 0%
	ε3/ε4	158, 29.31%	106, 31.83%	9, 28.12%	19, 26.76%	3, 42.85%	0, 0%	0, 0%	11, 22%	0, 0%	9, 21.95%	1, 50%	0, 0%
	ε4/ε4	40, 7.42%	23, 6.90%	5, 15.62%	1, 1.40%	0, 0%	1, 50%	0, 0%	7, 14%	0, 0%	3, 7.31%	0, 0%	0, 0%
	Total	539	333	32	71	7	2	1	50	0	41	2	0
Dementia													
	Genotypes	**Number of participants (n, %)**	**EUR (n, %)**	**AFR (n, %)**	**AMR (n, %)**	**EAS (n, %)**	**SAS (n, %)**	**MDE (n, %)**	**AJ (n, %)**	**FIN (n, %)**	**AAC (n, %)**	**CAS (n, %)**	**CAH (n, %)**
	00_CC, unknown	1, 0.06%	1, 0.09%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%
	ε1/ε1	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%
	ε1/ε2	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%
	ε1/ε4	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%
	ε2/ε2	12, 0.72%	7, 0.71%	1, 0.57%	2, 0.93%	0, 0%	0, 0%	0, 0%	1, 0.97%	0, 0%	1, 0.69%	0, 0%	0, 0%
	ε2/ε3	162, 9.78%	88, 9.02%	24, 13.71%	19, 8.87%	4, 21.05%	1, 10%	0, 0%	8, 7.76%	0, 0%	18, 12.5%	0, 0%	0, 0%
	ε2/ε4 or ε1/ε3	38, 2.29%	17, 1.74%	12, 6.85%	1, 0.46%	0, 0%	0, 0%	0, 0%	2, 1.94%	0, 0%	6, 4.16%	0, 0%	0, 0%
	ε3/ε3	912, 55.10%	551, 56.51%	77, 44%	130, 60.74%	10, 52.63%	7, 70%	6, 100%	60, 58.25%	2, 100%	66, 45.83%	3, 42.85%	0, 0%
	ε3/ε4	449, 27.12%	266, 27.28%	53, 30.28%	54, 25.23%	5, 26.31%	1, 10%	0, 0%	22, 21.35%	0, 0%	45, 31.25%	3, 42.85%	0, 0%
	ε4/ε4	81, 4.89%	45, 4.61%	8, 4.57%	8, 3.73%	0, 0%	1, 10%	0, 0%	10, 9.70%	0, 0%	8, 5.55%	1, 14.28%	0, 0%
	Total	1655	975	175	214	19	10	6	103	2	144	7	0
Controls													
	Genotypes	**Number of participants (n, %)**	**EUR (n, %)**	**AFR (n, %)**	**AMR (n, %)**	**EAS (n, %)**	**SAS (n, %)**	**MDE (n, %)**	**AJ (n, %)**	**FIN (n, %)**	**AAC (n, %)**	**CAS (n, %)**	**CAH (n, %)**
	CT_00, unknown	1, 0.007%	1, 0.01%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%
	00_CC, unknown	1, 0.007%	1, 0.01%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%
	ε1/ε1	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%
	ε1/ε2	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%
	ε1/ε4	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%
	ε2/ε2	90, 0.65%	51, 0.57%	18, 1.37%	4, 0.3%	3, 0.81%	0, 0%	0, 0%	6, 0.68%	0, 0%	8, 0.89%	0, 0%	0, 0%
	ε2/ε3	1734, 12.53%	1124, 12.65%	220, 16.85%	82, 6.28%	61, 16.66%	5, 6.84%	13, 19.69%	113, 12.94%	3, 10.71%	108, 12.02%	5, 11.90%	0, 0%
	ε2/ε4 or ε1/ε3	295, 2.13%	168, 1.89%	66, 5.05%	11, 0.84%	1, 0.27%	1, 1.36%	2, 3.03%	10, 1.14%	0, 0%	36, 4.00%	0, 0%	0, 0%
	ε3/ε3	8537, 61.70%	5597, 63.02%	587, 44.98%	927, 71.08%	252, 68.85%	51, 69.86%	46, 69.69%	575, 65.86%	20, 71.42%	453, 50.44%	29, 69.04%	0, 0%
	ε3/ε4	2930, 21.17%	1817, 20.46%	349, 26.74%	265, 20.32%	48, 13.11%	16, 21.91%	5, 7.57%	160, 18.32%	4, 14.28%	259, 28.84%	7, 16.66%	0, 0%
	ε4/ε4	247, 1.78%	121, 1.36%	65, 4.98%	15, 1.15%	1, 0.27%	0, 0%	0, 0%	9, 1.03%	1, 3.57%	34, 3.78%	1, 2.38%	0, 0%
	Total	13835	8880	1305	1304	366	73	66	873	28	898	42	0
ADSP													
AD													
	Genotypes	**Number of participants (n, %)**	**EUR (n, %)**	**AFR (n, %)**	**AMR (n, %)**	**EAS (n, %)**	**SAS (n, %)**	**MDE (n, %)**	**AJ (n, %)**	**FIN (n, %)**	**AAC (n, %)**	**CAS (n, %)**	**CAH (n, %)**

00_00, unknown	45, 0.42%	44, 0.75%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	1, 0.18%	0, 0%	0, 0%	0, 0%	0, 0%
00_CC, unknown	17, 0.16%	10, 0.17%	3, 0.45%	0, 0%	0, 0%	0, 0%	0, 0%	1, 0.18%	0, 0%	2, 0.17%	0, 0%	1, 0.09%
00_TC, unknown	1, 0.00000095%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	1, 0.18%	0, 0%	0, 0%	0, 0%	0, 0%
TT_00, unknown	7, 0.06%	7, 0.12%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%
CT_00, unknown	22, 0.20%	22, 0.37%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%
CC_00, unknown	9, 0.08%	9, 0.15%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%
ε1/ε1	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%
ε1/ε2	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%
ε1/ε4	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%
ε2/ε2	21, 0.19%	7, 0.12%	4, 0.60%	0, 0%	0, 0%	0, 0%	0, 0%	2, 0.37%	0, 0%	5, 0.44%	0, 0%	3, 0.28%
ε2/ε3	479, 4.53%	177, 3.04%	47, 7.07%	38, 3.49%	3, 6%	15, 8.52%	1, 3.70%	18, 3.35%	0, 0%	108, 9.65%	2, 10%	70, 6.57%
ε2/ε4 or ε1/ε3	251, 2.37%	129, 2.21%	32, 4.81%	9, 0.82%	1, 2%	4, 2.27%	0, 0%	11, 2.04%	0, 0%	41, 3.66%	1, 5%	23, 2.16%
ε3/ε3	4296, 40.65%	2070, 35.61%	210, 31.62%	686, 63.16%	24, 48%	112, 63.63%	16, 59.25%	241, 44.87%	3, 27.27%	377, 33.69%	9, 45%	548, 51.50%
ε3/ε4	4263, 40.34%	2596, 44.66%	285, 42.92%	301, 27.71%	16, 32%	43, 24.43%	8, 29.62%	222, 41.34%	4, 36.36%	443, 39.58%	7, 35%	338, 31.76%
ε4/ε4	1155, 10.93%	741, 12.74%	83, 12.5%	52, 4, 78%	6, 12%	2, 1.13%	2, 7.40%	40, 7.44%	4, 36.36%	143, 12.77%	1, 5%	81, 7.61%
Total	10566	5812	664	1086	50	176	27	537	11	1119	20	1064

Controls

Genotypes	Number of participants (n, %)	EUR (n, %)	AFR (n, %)	AMR (n, %)	EAS (n, %)	SAS (n, %)	MDE (n, %)	AJ (n, %)	FIN (n, %)	AAC (n, %)	CAS (n, %)	CAH (n, %)
00_00, unknown	124, 0.76%	120, 2.72%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	1, 0.06%	0, 0%	3, 0.17%
00_CC, unknown	21, 0.12%	13, 0.29%	2, 0.11%	1, 0.02%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	4, 0.25%	0, 0%	1, 0.05
00_TC, unknown	3, 0.01%	2, 0.04%	1, 0.05%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%
TT_00, unknown	5, 0.03%	3, 0.06%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	1, 0.21%	0, 0%	0, 0%	0, 0%	1, 0.05
CT_00, unknown	83, 0.51%	82, 1.85%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	1, 0.21%	0, 0%	0, 0%	0, 0%	0, 0%
CC_00, unknown	8, 0.04%	8, 0.18%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%
ε1/ε1	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%
ε1/ε2	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%
ε1/ε4	1, 0.006%	0, 0%	1, 0.05%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%
ε2/ε2	71, 0.43%	18, 0.40%	22, 1.30%	5, 0.12%	0, 0%	3, 0.13%	0, 0%	1, 0.21%	0, 0%	15, 0.95%	2, 1.52%	5, 0.29%
ε2/ε3	1528, 9.42%	418, 9.47%	244, 14.48%	217, 5.47%	5, 17.24%	175, 7.67%	1, 11.11%	47, 10.23%	3, 25%	235, 15.03%	18, 13.74%	165, 9.86%
ε2/ε4 or ε1/ε3	293, 1.80%	68, 1.54%	78, 4.62%	28, 0.70%	1, 3.44%	25, 1.09%	0, 0%	4, 0.87%	0, 0%	61, 3.90%	1, 0.76%	27, 1.61%
ε3/ε3	10197, 62.87%	2540, 57.58%	777, 46.11%	3008, 75.82%	12, 41.37%	1649, 72.35%	6, 66.66%	270, 58.82%	5, 41.66%	765, 48.94%	86, 65.64%	1079, 64.53%
ε3/ε4	3544, 21.85%	1026, 23.26%	494, 29.31%	672, 16.93%	8, 27.58%	398, 17.46%	2, 22.22%	128, 27.88%	4, 33.33%	435, 27.83%	20, 15.26%	357, 21.35%
ε4/ε4	339, 2.09%	113, 2.56%	66, 3.91%	36, 0.90%	3, 10.34%	29, 1.27%	0, 0%	7, 1.52%	0, 0%	47, 3.00%	4, 3.05%	34, 2.03%
Total	16217	4411	1685	3967	29	2279	9	459	12	1563	131	1672

UKB

AD

Genotypes	Number of participants (n, %)	EUR (n, %)	AFR (n, %)	AMR (n, %)	EAS (n, %)	SAS (n, %)	MDE (n, %)	AJ (n, %)	FIN (n, %)	AAC (n, %)	CAS (n, %)	CAH (n, %)
ε1/ε1	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%
ε1/ε2	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%
ε1/ε4	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%
ε2/ε2	9, 0.21%	8 , 0.20%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	1 , 11.11%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%
ε2/ε3	212, 5.02%	205 , 5.06%	5 , 9.62%	0 , 0%	0 , 0%	1 , 2.04%	0 , 0%	0 , 0%	0 , 0%	1 , 8.33%	0 , 0%	0 , 0%
ε2/ε4 or ε1/ε3	102, 0.02%	97 , 2.39%	5 , 9.62%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%
ε3/ε3	1391, 32.92%	1321 , 32.61%	11 , 21.15%	0 , 0%	5 , 71.43%	25 , 51.02%	6 , 66.67%	10 , 37.04%	0 , 0%	1 , 8.33%	6 , 60%	6 , 75%
ε3/ε4	1931, 45.70%	1867 , 46.09%	18 , 34.62%	0 , 0%	2 , 28.57%	20 , 40.82%	1 , 11.11%	15 , 55.56%	0 , 0%	3 , 25%	3 , 30%	2 , 25%
ε4/ε4	580, 13.73%	553 , 13.65%	13 , 25%	0 , 0%	0 , 0%	3 , 6.12%	1 , 11.11%	2 , 7.41%	0 , 0%	7 , 58.33%	1 , 10%	0 , 0%
total	4225	4051	52	0	7	49	9	27	0	12	10	8

Dementia

Genotypes	Number of participants (n, %)	EUR (n, %)	AFR (n, %)	AMR (n, %)	EAS (n, %)	SAS (n, %)	MDE (n, %)	AJ (n, %)	FIN (n, %)	AAC (n, %)	CAS (n, %)	CAH (n, %)

ε1/ε1	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%
ε1/ε2	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%
ε1/ε4	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%
ε2/ε2	24, 0.45%	23 , 0.46%	1 , 1.56%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%
ε2/ε3	447, 8.42%	427 , 8.49%	6 , 9.38%	0 , 0%	2 , 20%	6 , 8.57%	2 , 15.38%	2 , 4.65%	0 , 0%	1 , 4.17%	1 , 4.35%	0 , 0%
ε2/ε4 or ε1/ε3	165, 3.11%	153 , 3.04%	4 , 6.25%	0 , 0%	1 , 10%	1 , 1.43%	0 , 0%	1 , 2.33%	0 , 0%	3 , 12.50%	0 , 0%	2 , 7.41%
ε3/ε3	2432, 45.83%	2291 , 45.56%	23 , 35.94%	2 , 66.67%	6 , 60%	38 , 54.29%	10 , 76.92%	26 , 60.47%	0 , 0%	8 , 33.33%	15 , 65.22%	13 , 48.15%
ε3/ε4	1832, 34.53%	1748 , 34.77%	22 , 34.38%	1 , 33.33%	1 , 10%	21 , 30%	1 , 7.69%	13 , 30.23%	1 , 100%	7 , 29.17%	6 , 26.09%	11 , 40.74%
ε4/ε4	406, 7.65%	386 , 7.68%	8 , 12.50%	0 , 0%	0 , 0%	4 , 5.71%	0 , 0%	1 , 2.33%	0 , 0%	5 , 20.83%	1 , 4.35%	1 , 3.70%
total	5306	5028	64	3	10	70	13	43	1	24	23	27

Controls

Genotypes	Number of participants (n, %)	EUR (n, %)	AFR (n, %)	AMR (n, %)	EAS (n, %)	SAS (n, %)	MDE (n, %)	AJ (n, %)	FIN (n, %)	AAC (n, %)	CAS (n, %)	CAH (n, %)
ε1/ε1	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%
ε1/ε2	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%
ε1/ε4	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%
ε2/ε2	367, 0.65%	358 , 0.66%	3 , 0.67%	0 , 0%	1 , 0.65%	1 , 0.15%	0 , 0%	2 , 0.48%	0 , 0%	0 , 0%	1 , 0.49%	1 , 0.59%
ε2/ε3	7416, 13.07%	7167 , 13.17%	67 , 15.06%	4 , 8.33%	18 , 11.69%	51 , 7.59%	5 , 8.06%	38 , 9.13%	3 , 20%	29 , 21.80%	16 , 7.84%	18 , 10.65%
ε2/ε4 or ε1/ε3	1300, 2.29%	1231 , 2.26%	29 , 6.52%	0 , 0%	2 , 1.30%	8 , 1.19%	0 , 0%	11 , 2.64%	0 , 0%	9 , 6.77%	4 , 1.96%	6 , 3.55%
ε3/ε3	34770, 61.28%	33277 , 61.15%	193 , 43.37%	33 , 68.75%	103 , 66.88%	508 , 75.60%	50 , 80.65%	284 , 68.27%	10 , 66.67%	56 , 42.11%	153 , 75%	103 , 60.95%
ε3/ε4	11946, 21.05%	11491 , 21.11%	136 , 30.56%	11 , 22.92%	28 , 18.18%	95 , 14.14%	6 , 9.68%	78 , 18.75%	2 , 13.33%	33 , 24.81%	29 , 14.22%	37 , 21.89%
ε4/ε4	942, 1.66%	899 , 1.65%	17 , 3.82%	0 , 0%	2 , 1.30%	9 , 1.34%	1 , 1.61%	3 , 0.72%	0 , 0%	6 , 4.51%	1 , 0.49%	4 , 2.37%
total	56741	54423	445	48	154	672	62	416	15	133	204	169

AMP PD

DLB

Genotypes	Number of participants (n, %)	EUR (n, %)	AFR (n, %)	AMR (n, %)	EAS (n, %)	SAS (n, %)	MDE (n, %)	AJ (n, %)	FIN (n, %)	AAC (n, %)	CAS (n, %)	CAH (n, %)
unknown,unknown	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%
ε1/ε1	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%
ε1/ε2	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%
ε1/ε4	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%
ε2/ε2	5, 0.20%	5 , 0.21%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%
ε2/ε3	191, 7.55%	179 , 7.44%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	12 , 10.62%	0 , 0%	0 , 0%	0 , 0%	0 , 0%
ε2/ε4 or ε1/ε3	84, 3.32%	78 , 3.24%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	6 , 5.31%	0 , 0%	0 , 0%	0 , 0%	0 , 0%
ε3/ε3	1162, 45.93%	1108 , 46.05%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	3 , 50%	48 , 42.48%	3 , 60%	0 , 0%	0 , 0%	0 , 0%
ε3/ε4	882, 34.86%	842 , 35%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	3 , 50%	35 , 30.97%	2 , 40%	0 , 0%	0 , 0%	0 , 0%
ε4/ε4	206, 8.14%	194 , 8.06%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	12 , 10.62%	0 , 0%	0 , 0%	0 , 0%	0 , 0%
total	2530	2406	0	0	0	0	6	113	5	0	0	0

Controls

Genotypes	Number of participants (n, %)	EUR (n, %)	AFR (n, %)	AMR (n, %)	EAS (n, %)	SAS (n, %)	MDE (n, %)	AJ (n, %)	FIN (n, %)	AAC (n, %)	CAS (n, %)	CAH (n, %)
unknown,unknown	4, 0.12%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	1 , 100%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	3 , 100%	0 , 0%
ε1/ε1	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%
ε1/ε2	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%
ε1/ε4	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%
ε2/ε2	13, 0.40%	11 , 0.38%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	1 , 0.41%	0 , 0%	1 , 2.50%	0 , 0%	0 , 0%
ε2/ε3	352, 10.76%	310 , 10.62%	4 , 13.79%	1 , 7.14%	1 , 20%	0 , 0%	0 , 0%	31 , 12.60%	0 , 0%	4 , 10%	0 , 0%	1 , 25%
ε2/ε4 or ε1/ε3	49, 1.50%	40 , 1.37%	1 , 3.45%	1 , 7.14%	0 , 0%	0 , 0%	0 , 0%	7 , 2.85%	0 , 0%	0 , 0%	0 , 0%	0 , 0%
ε3/ε3	2100, 64.22%	1885 , 64.58%	19 , 65.52%	9 , 64.29%	3 , 60%	0 , 0%	4 , 100%	146 , 59.35%	5 , 100%	26 , 65%	0 , 0%	3 , 75%
ε3/ε4	698, 21.34%	620 , 21.24%	5 , 17.24%	3 , 21.43%	1 , 20%	0 , 0%	0 , 0%	60 , 24.39%	0 , 0%	9 , 22.50%	0 , 0%	0 , 0%
ε4/ε4	54, 1.65%	53 , 1.82%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	0 , 0%	1 , 0.41%	0 , 0%	0 , 0%	0 , 0%	0 , 0%
total	3270	2919	29	14	5	1	4	246	5	40	3	4

100KGP

Dementia

Genotypes	Number of participants (n, %)	EUR (n, %)	AFR (n, %)	AMR (n, %)	EAS (n, %)	SAS (n, %)	NA (n, %)
NA	10, 5.55%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	10, 100%
ε2/ε3	11, 6.11%	9, 6.29%	1, 14.28%	0, 0%	0, 0%	1, 6.25%	0, 0%
ε2/ε4 or ε1/ε3	3, 1.67%	3, 2.10%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%
ε3/ε3	69, 38.33%	58, 40.56%	1, 14.28%	2, 66.67%	1, 100%	7, 43.75%	0, 0%
ε3/ε4	66, 36.67%	54, 37.76%	4, 57.14%	0, 0%	0, 0%	8, 50%	0, 0%
ε4/ε4	21, 11.67%	19, 13.29%	1, 14.28%	1, 33.33%	0, 0%	0, 0%	0, 0%
Total	180	143	7	3	1	16	10

Controls

Genotypes	Number of participants (n, %)	EUR (n, %)	AFR (n, %)	AMR (n, %)	EAS (n, %)	SAS (n, %)	NA (n, %)
NA	31, 0.89%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	31, 100%
ε2/ε2	25, 0.71%	22, 0.72%	1, 1.61%	1, 1.16%	0, 0%	1, 0.41%	0, 0%
ε2/ε3	436, 12.53%	394, 12.97%	7, 11.29%	9, 10.46%	4, 18.18%	22, 9.12%	0, 0%
ε2/ε4 or ε1/ε3	72, 2.06%	61, 2.00%	6, 9.67%	1, 1.16%	1, 4.54%	3, 1.24%	0, 0%
ε3/ε3	2042, 58.69%	1765, 58.11%	25, 40.32%	61, 70.93%	14, 63.63%	177, 73.44%	0, 0%
ε3/ε4	802, 23.05%	729, 24.00%	20, 32.25%	13, 15.11%	3, 13.63%	37, 15.35%	0, 0%
ε4/ε4	71, 2.04%	66, 2.17%	3, 4.83%	1, 1.16%	0, 0%	1, 0.41%	0, 0%
Total	3479	3037	62	86	22	241	31

In 100KGP, genotypes that could not be detected are shown as NA. AD, Alzheimer's disease; DLB, Dementia with Lewy bodies; 100KGP, 100,000 Genomes Project; ADSP, Alzheimer's Disease Sequencing Project; UKB, UK Biobank; AMP PD, Accelerating Medicines Partnership in Parkinson's Disease; AoU, All of Us; EUR, European; AFR, African; AMR, American Admixed; AAC,African Admixed; AJ, Ashkenazi Jewish; CAS, Central Asian; EAS, Eastern Asian; SAS, South Asian; MDE, Middle Eastern; FIN, Finnish; CAH, Complex Admixture History.

Table 6- Summary of the analysis of protective/disease-modifying variants in Alzheimer's disease cases and controls in the ADSP

Protective_model <- glm(PHENO ~ Protective/disease-modifying variant + SEX + AGE + PC1 + PC2 + PC3 + PC4 + PC5, data = data, family = binomial)

***19q13.31*: rs10423769, chr19:43100929:G>A**

Ancestry	EUR	AFR	AMR	EAS	SAS	MDE	AJ	FIN	AAC	CAS	CAH
A1	A	A	A	A	A	A	A	A	A	A	A
P	0.460811	0.6021	0.215674	NA	NA	0.789232	0.46266	NA	0.842912	NA	0.201781
OR (L95_U95)	0.79711 (0.436315_1.45625)	0.949949 (0.783198_1.1522)	1.26959 (0.870087_1.85251)	NA	NA	1.89966 (0.0171931_209.893)	0.807628 (0.456669_1.4283)	NA	1.01806 (0.852894_1.21521)	NA	0.848645 (0.659614_1.09185)

***APOE*:rs449647, Chr19:44905307: A>T**

Ancestry	EUR	AFR	AMR	EAS	SAS	MDE	AJ	FIN	AAC	CAS	CAH
A1	T	T	T	T	T	T	T	T	T	T	T
P	2.19e-14	6.62e-06	0.921403	0.410879	0.71756	0.126724	0.000246962	0.657878	0.0596323	0.935904	0.490397
OR (L95_U95)	0.723607 (0.66597_0.786231)	1.37695 (1.19808_1.58252)	1.00596 (0.894033_1.13189)	2.47276 (0.285812_21.3935)	0.940468 (0.67439_1.31153)	0.25441 (0.0439125_1.47394)	0.671781 (0.543053_0.831022)	0.43746 (0.0112664_16.986)	1.12355 (0.995282_1.26834)	1.03813 (0.417042_2.58418)	1.04671 (0.91935_1.19172)

***TOMM40*:rs11556505, Chr19:44892887: C>T**

Ancestry	EUR	AFR	AMR	EAS	SAS	MDE	AJ	FIN	AAC	CAS	CAH
A1	T	T	T	T	T	T	T	T	T	T	T
P	4.86e-108	0.280115	5.78e-15	0.999971	0.290208	0.390511	0.00013708	0.726914	0.00011157	0.504467	2.85e-07
OR (L95_U95)	2.29043 (2.12799_2.46527)	1.12056 (0.911434_1.37768)	1.76617 (1.53118_2.03722)	0.999981 (0.36233_2.75981)	1.18727 (0.863778_1.63191)	3.89111 (0.175077_86.4808)	1.61043 (1.26061_2.05732)	1.62127 (0.107672_24.4124)	1.38513 (1.17414_1.63403)	1.41907 (0.50777_3.96587)	1.56334 (1.31814_1.85415)

***NOCT*:rs13116075 , Chr4:139008878: A>G**

Ancestry	EUR	AFR	AMR	EAS	SAS	MDE	AJ	FIN	AAC	CAS	CAH
A1	G	G	G	G	G	G	G	G	G	G	G
P	0.0874581	0.658188	0.113149	0.66702	0.9558	0.661009	0.142146	0.231494	0.180217	0.107993	0.705767
OR (L95_U95)	1.07114 (0.989957_1.15899)	1.10373 (0.712821_1.709)	0.865449 (0.723765_1.03487)	0.343613 (0.00264628_44.6173)	0.988192 (0.649261_1.50405)	1.89182 (0.109489_32.688)	0.831151 (0.649279_1.06397)	188.49 (0.0353377_1005400.0)	0.853469 (0.676933_1.07604)	0.10942 (0.00736793_1.62499)	0.965653 (0.80542_1.15776)

***CASS4*:rs6024870 , Chr20:56422512: G>A**

Ancestry	EUR	AFR	AMR	EAS	SAS	MDE	AJ	FIN	AAC	CAS	CAH
A1	A	A	A	A	A	A	A	A	A	A	A
P	0.317519	0.19334	0.947158	NA	0.743636	0.399733	0.159203	NA	0.604016	0.512028	0.79705
OR (L95_U95)	0.948467 (0.855001_1.05215)	0.851492 (0.668311_1.08488)	0.993588 (0.82148_1.20176)	NA	1.17134 (0.454001_3.02213)	3.55752 (0.185509_68.2227)	0.836323 (0.652077_1.07263)	NA	1.05138 (0.870015_1.27055)	0.458625 (0.0446159_4.7144)	1.02512 (0.848483_1.23854)

***LRRC37A*: rs2732703 , Chr17:46275856: T>G**

Ancestry	EUR	AFR	AMR	EAS	SAS	MDE	AJ	FIN	AAC	CAS	CAH
A1	G	G	G	G	G	G	G	G	G	G	G
P	0.0147474	0.545618	0.580112	NA	0.80132	0.877837	0.0979924	0.78562	0.931261	0.717238	0.639068
OR (L95_U95)	0.900167 (0.8272_0.979571)	0.862876 (0.534833_1.39212)	0.95293 (0.803301_1.13043)	NA	0.928456 (0.52079_1.65524)	0.827929 (0.0745303_9.19716)	0.803044 (0.619302_1.0413)	1.46559 (0.0932775_23.0276)	0.989343 (0.775568_1.26204)	0.737015 (0.141339_3.84319)	0.956805 (0.795579_1.1507)

Conditional_model <- glm(PHENO ~ Protective/disease-modifying variant + APOE_STATUS (ε4 carriers) +SEX + AGE + PC1 + PC2 + PC3 + PC4 + PC5, data = data, family = binomial)

***19q13.31*: rs10423769, chr19:43100929:G>A**

Ancestry	EUR	AFR	AMR	EAS	SAS	MDE	AJ	FIN	AAC	CAS	CAH
A1	A	A	A	A	A	A	A	A	A	A	A
P	0.690443	0.517056	0.161042	NA	NA	0.70344	0.535244	NA	0.922067	NA	0.157721
OR (L95_U95)	0.879192 (0.466549_1.6568)	0.936955 (0.769422_1.14097)	1.31287 (0.897222_1.92108)	NA	NA	2.51389 (0.0218339_289.443)	0.832044 (0.465285_1.4879)	NA	1.00909 (0.841704_1.20978)	NA	0.831481 (0.643669_1.07409)

***APOE*:rs449647, Chr19:44905307: A>T**

Ancestry	EUR	AFR	AMR	EAS	SAS	MDE	AJ	FIN	AAC	CAS	CAH
A1	T	T	T	T	T	T	T	T	T	T	T
P	0.438088	0.0624138	0.472261	0.43965	0.865612	0.139162	0.0285317	0.864837	0.496478	0.851361	0.941387
OR (L95_U95)	0.965782 (0.884422_1.05463)	1.15042 (0.99276_1.33312)	1.04487 (0.927005_1.17773)	2.38768 (0.262626_21.7078)	0.971482 (0.694882_1.35818)	0.291295 (0.0568107_1.49361)	0.780509 (0.625253_0.974316)	1.6768 (0.00436235_644.531)	0.956649 (0.841935_1.08699)	1.09711 (0.416133_2.89248)	0.995056 (0.871901_1.13561)

***TOMM40*:rs11556505, Chr19:44892887: C>T**

Ancestry	EUR	AFR	AMR	EAS	SAS	MDE	AJ	FIN	AAC	CAS	CAH
A1	T	T	T	T	T	T	T	T	T	T	T
P	3.05e-06	0.349246	0.258	0.305313	0.925043	0.306123	0.825232	0.864985	0.0409785	0.517952	0.0463152
OR (L95_U95)	1.24509 (1.1356_1.36513)	1.10709 (0.894683_1.36993)	1.12255 (0.918784_1.3715)	0.386149 (0.0626103_2.38157)	0.980508 (0.650674_1.47754)	12.6053 (0.0983582_1615.46)	1.03582 (0.757949_1.41555)	1.35469 (0.0409377_44.8285)	1.19548 (1.00735_1.41876)	0.637351 (0.162678_2.49707)	1.20744 (1.00308_1.45342)

***NOCT*:rs13116075 , Chr4:139008878: A>G**

Ancestry	EUR	AFR	AMR	EAS	SAS	MDE	AJ	FIN	AAC	CAS	CAH
A1	G	G	G	G	G	G	G	G	G	G	G
P	0.11744	0.645244	0.119401	0.551474	0.995776	0.682833	0.228236	0.415815	0.190788	0.123526	0.623371
OR (L95_U95)	1.06849 (0.983453_1.16087)	1.11107 (0.709591_1.7397)	0.866307 (0.723148_1.03781)	0.222352 (0.00157819_31.3274)	1.00114 (0.657627_1.52408)	0.664864 (0.0938476_4.71024)	0.856635 (0.666008_1.10183)	3.39754 (0.178546_64.6512)	0.853379 (0.672949_1.08219)	0.12043 (0.00814495_1.78066)	0.954866 (0.794128_1.14814)

***CASS4*:rs6024870 , Chr20:56422512: G>A**

Ancestry	EUR	AFR	AMR	EAS	SAS	MDE	AJ	FIN	AAC	CAS	CAH
A1	A	A	A	A	A	A	A	A	A	A	A
P	0.166437	0.266129	0.885204	NA	0.722628	0.396827	0.0769097	NA	0.725645	0.491759	0.891828
OR (L95_U95)	0.925783 (0.829988_1.03263)	0.868781 (0.678012_1.11322)	1.01427 (0.836751_1.22946)	NA	1.18792 (0.459014_3.07431)	4.06585 (0.158516_104.287)	0.795013 (0.616578_1.02509)	NA	1.0353 (0.852927_1.25668)	0.434108 (0.0402219_4.68526)	1.0134 (0.836458_1.22778)

***LRRC37A*: rs2732703 , Chr17:46275856: T>G**

Ancestry	EUR	AFR	AMR	EAS	SAS	MDE	AJ	FIN	AAC	CAS	CAH
A1	G	G	G	G	G	G	G	G	G	G	G
P	0.0206232	0.478694	0.481538	NA	0.761611	0.861743	0.181446	0.831894	0.872138	0.833465	0.648529
OR (L95_U95)	0.900245 (0.823603_0.98402)	0.837245 (0.512161_1.36867)	0.939807 (0.790605_1.11717)	NA	0.91409 (0.511621_1.63316)	0.806351 (0.0715386_9.08883)	0.834828 (0.640625_1.0879)	1.38359 (0.0690329_27.7305)	0.979684 (0.763012_1.25788)	0.838877 (0.163097_4.31471)	0.957399 (0.793949_1.1545)

Conditional_model <- glm(PHENO ~ Protective/disease-modifying variant + APOE_STATUS (ε4/ε4) +SEX + AGE + PC1 + PC2 + PC3 + PC4 + PC5, data = data, family = binomial)

***19q13.31*: rs10423769, chr19:43100929:G>A**

Ancestry	EUR	AFR	AMR	EAS	SAS	MDE	AJ	FIN	AAC	CAS	CAH
A1	A	A	A	A	A	A	A	A	A	A	A
P	0.520432	0.487737	0.242755	NA	NA	0.723333	0.464294	NA	0.7821	NA	0.116951
OR (L95_U95)	0.818862 (0.445176_1.50622)	0.93305 (0.767187_1.13477)	1.25453 (0.857515_1.83537)	NA	NA	2.34504 (0.0209322_262.714)	0.806588 (0.453563_1.43439)	NA	1.02572 (0.8568_1.22794)	NA	0.814618 (0.630416_1.05264)

***APOE*:rs449647, Chr19:44905307: A>T**

Ancestry	EUR	AFR	AMR	EAS	SAS	MDE	AJ	FIN	AAC	CAS	CAH
A1	T	T	T	T	T	T	T	T	T	T	T
P	5.13e-07	0.00116208	0.65243	0.406828	0.713205	0.204203	0.00311608	0.937989	0.506202	0.940189	0.843606
OR (L95_U95)	0.80682 (0.741977_0.87733)	1.26706 (1.09841_1.4616)	1.02765 (0.912626_1.15718)	2.44129 (0.296298_20.1146)	0.939528 (0.673676_1.31029)	0.362537 (0.0757061_1.73609)	0.723028 (0.583137_0.896479)	1.17291 (0.021098_65.2066)	1.04321 (0.920878_1.1818)	1.03564 (0.414924_2.58492)	1.01332 (0.888485_1.1557)

***TOMM40*:rs11556505, Chr19:44892887: C>T**

Ancestry	EUR	AFR	AMR	EAS	SAS	MDE	AJ	FIN	AAC	CAS	CAH
A1	T	T	T	T	T	T	T	T	T	T	T
P	1.68e-58	0.375778	1.22e-07	0.649911	0.24013	0.847088	0.00962134	0.823812	0.00964987	0.494246	0.000521882
OR (L95_U95)	1.95842 (1.80479_2.12513)	1.09968 (0.891141_1.35703)	1.5297 (1.30682_1.79058)	1.38741 (0.337396_5.70517)	1.21776 (0.876594_1.69171)	1.26953 (0.112259_14.357)	1.41037 (1.08715_1.82968)	1.25464 (0.17031_9.24275)	1.25348 (1.05637_1.48738)	1.46486 (0.490248_4.37702)	1.37305 (1.14789_1.64237)

***NOCT*:rs13116075 , Chr4:139008878: A>G**

Ancestry	EUR	AFR	AMR	EAS	SAS	MDE	AJ	FIN	AAC	CAS	CAH
A1	G	G	G	G	G	G	G	G	G	G	G
P	0.0886405	0.577482	0.126453	0.748873	0.954283	0.742716	0.138021	0.529144	0.183083	0.104656	0.648668
OR (L95_U95)	1.07213 (0.989528_1.16163)	1.13356 (0.729266_1.762)	0.869347 (0.726455_1.04035)	0.424654 (0.00224227_80.4235)	0.987788 (0.648967_1.5035)	0.713686 (0.0952421_5.34792)	0.828076 (0.645363_1.06252)	2.58464 (0.134278_49.7502)	0.851783 (0.672605_1.07869)	0.105682 (0.00700063_1.59537)	0.958323 (0.797957_1.15092)

***CASS4* :rs6024870 , Chr20:56422512: G>A**

Ancestry	EUR	AFR	AMR	EAS	SAS	MDE	AJ	FIN	AAC	CAS	CAH
A1	A	A	A	A	A	A	A	A	A	A	A
P	0.315877	0.203848	0.932446	NA	0.74473	0.985781	0.152267	NA	0.500034	0.509982	0.826824
OR (L95_U95)	0.947416 (0.852505_1.05289)	0.852854 (0.667192_1.09018)	0.991742 (0.818715_1.20134)	NA	1.17055 (0.45363_3.02051)	0.979617 (0.101731_9.43317)	0.8319 (0.646608_1.07029)	NA	1.06851 (0.881347_1.29542)	0.456796 (0.04441_4.69855)	1.02149 (0.844336_1.23581)

***LRRC37A*: rs2732703 , Chr17:46275856: T>G**

Ancestry	EUR	AFR	AMR	EAS	SAS	MDE	AJ	FIN	AAC	CAS	CAH
A1	G	G	G	G	G	G	G	G	G	G	G
P	0.0170534	0.526575	0.657118	NA	0.800316	0.80885	0.108589	0.842417	0.84495	0.720352	0.691408
OR (L95_U95)	0.90071 (0.826556_0.981517)	0.854763 (0.525895_1.38929)	0.961802 (0.809848_1.14227)	NA	0.928096 (0.52057_1.65465)	0.767101 (0.0895205_6.57328)	0.806879 (0.620834_1.04868)	1.38422 (0.0561144_34.1458)	0.975534 (0.761079_1.25042)	0.739306 (0.141463_3.86371)	0.963032 (0.799585_1.15989)

r2_model <- lm(Protective/disease-modifying variant ~ APOE_STATUS (ε4 carriers) +SEX + AGE + PC1 + PC2 + PC3 + PC4 + PC5, data = data)

***19q13.31*: rs10423769, chr19:43100929:G>A**

Ancestry	EUR	AFR	AMR	EAS	SAS	MDE	AJ	FIN	AAC	CAS	CAH
Multiple R-squared	0.005848	0.006813	0.01403	NA	NA	0.06651	0.003517	NA	0.008922	NA	0.01368
Adjusted R-squared	0.00507	0.003417	0.01247	NA	NA	-0.2101	-0.00456	NA	0.005956	NA	0.01079
F-statistic	7.511	2.006	8.972	NA	NA	0.2404	0.4354	NA	3.008	NA	4.728
p-value (F-statistic)	4.81e-10	0.04213	2.69e-12	NA	NA	0.9792	0.9002	NA	0.002308	NA	8.90e-06

***APOE*:rs449647, Chr19:44905307: A>T**

Ancestry	EUR	AFR	AMR	EAS	SAS	MDE	AJ	FIN	AAC	CAS	CAH
Multiple R-squared	0.05969	0.09666	0.008391	0.1156	0.02943	0.1789	0.07881	0.6671	0.0556	0.07086	0.01471
Adjusted R-squared	5.90e-02	0.09357	0.006819	0.01455	0.02626	-0.06436	0.07134	0.4769	0.05278	0.01851	0.01182
F-statistic	81.04	31.3	5.336	1.144	9.272	0.7354	10.55	3.507	19.67	1.354	5.091
p-value (F-statistic)	< 2.2e-16	< 2.2e-16	1.08e-06	0.3455	1.15e-12	0.6598	2.72e-14	0.01951	< 2.2e-16	0.2221	2.62e-06

***TOMM40*:rs11556505, Chr19:44892887: C>T**

Ancestry	EUR	AFR	AMR	EAS	SAS	MDE	AJ	FIN	AAC	CAS	CAH
Multiple R-squared	0.4047	0.005166	0.5032	0.5528	0.3681	0.4932	0.4053	0.2905	0.03025	0.4191	0.1588
Adjusted R-squared	4.04e-01	0.001765	5.02e-01	0.5017	0.366	0.3431	0.4005	-0.1149	0.02735	0.3863	0.1563
F-statistic	867.9	1.519	638.5	10.82	178.1	3.285	84.07	0.7166	10.42	12.8	64.34
p-value (F-statistic)	< 2.2e-16	0.1453	< 2.2e-16	8.93e-10	< 2.2e-16	0.009537	< 2.2e-16	0.6747	1.76e-14	9.05e-14	< 2.2e-16

***NOCT*:rs13116075 , Chr4:139008878: A>G**

Ancestry	EUR	AFR	AMR	EAS	SAS	MDE	AJ	FIN	AAC	CAS	CAH
Multiple R-squared	0.001905	0.00331	0.04212	0.1196	0.01192	0.2339	0.01113	0.4501	0.02384	0.1287	0.007236
Adjusted R-squared	0.001124	-9.73E-05	0.0406	0.01903	0.008685	0.006878	0.003112	0.1359	0.02091	0.07962	0.004324
F-statistic	2.437	0.9714	27.72	1.189	3.688	1.03	1.388	1.432	8.159	2.622	2.485
p-value (F-statistic)	0.01247	0.4564	< 2.2e-16	0.3181	0.0002741	0.4381	0.1974	0.2659	5.90e-11	0.01045	0.01101

***CASS4* :rs6024870 , Chr20:56422512: G>A**

Ancestry	EUR	AFR	AMR	EAS	SAS	MDE	AJ	FIN	AAC	CAS	CAH
Multiple R-squared	0.005311	0.008006	0.04053	NA	0.003495	0.2169	0.009504	NA	0.001664	0.09364	0.006718
Adjusted R-squared	0.004532	0.004614	0.03901	NA	0.0002362	-0.01509	0.001476	NA	-0.001323	0.04258	0.003805
F-statistic	6.818	2.361	26.63	NA	1.072	0.935	1.184	NA	0.5571	1.834	2.306
p-value (F-statistic)	5.75e-09	0.01578	< 2.2e-16	NA	0.3794	0.5046	0.3055	NA	0.8136	0.07543	0.0184

***LRRC37A*: rs2732703 , Chr17:46275856: T>G**

Ancestry	EUR	AFR	AMR	EAS	SAS	MDE	AJ	FIN	AAC	CAS	CAH
Multiple R-squared	0.007921	0.03169	0.07651	NA	0.01552	0.3253	0.01275	0.3684	0.03173	0.0346	0.008147
Adjusted R-squared	0.007144	0.02838	0.07505	NA	0.0123	0.1254	0.004751	0.007558	0.02883	-0.01979	0.005237
F-statistic	10.19	9.574	52.24	NA	4.821	1.627	1.594	1.021	10.95	0.6361	2.8
p-value (F-statistic)	2.74e-14	3.99e-13	< 2.2e-16	NA	6.59e-06	0.1635	0.1223	0.4641	2.65e-15	0.7463	0.004337

r2_model <- lm(Protective/disease-modifying variant ~ APOE_STATUS (ε4/ε4) +SEX + AGE + PC1 + PC2 + PC3 + PC4 + PC5, data = data)

***19q13.31*: rs10423769, chr19:43100929:G>A**

Ancestry	EUR	AFR	AMR	EAS	SAS	MDE	AJ	FIN	AAC	CAS	CAH
Multiple R-squared	0.005736	0.007183	0.0141	NA	NA	0.05772	0.00307	NA	0.0089	NA	0.01519
Adjusted R-squared	0.004957	0.003789	0.01254	NA	NA	-0.2215	-0.00501	NA	0.005934	NA	0.0123
F-statistic	7.366	2.116	9.019	NA	NA	0.2067	0.38	NA	3.001	NA	5.258
p-value (F-statistic)	8.10e-10	0.03127	2.27e-12	NA	NA	0.9871	0.9315	NA	0.00236	NA	1.49e-06

***APOE*:rs449647, Chr19:44905307: A>T**

Ancestry	EUR	AFR	AMR	EAS	SAS	MDE	AJ	FIN	AAC	CAS	CAH
Multiple R-squared	0.0319	0.04362	0.005569	0.1584	0.02047	0.1899	0.03783	0.445	0.02097	0.04018	0.01183
Adjusted R-squared	3.11e-02	0.04035	0.003992	0.0622	0.01727	-0.05015	0.03003	0.1279	0.01804	-0.01389	0.008934
F-statistic	42.07	13.34	3.531	1.647	6.391	0.7911	4.85	1.403	7.155	0.7431	4.082
p-value (F-statistic)	< 2.2e-16	< 2.2e-16	0.0004401	0.1274	3.05e-08	0.6149	6.99e-06	0.2767	2.06e-09	0.6533	7.59e-05

***TOMM40*:rs11556505, Chr19:44892887: C>T**

Ancestry	EUR	AFR	AMR	EAS	SAS	MDE	AJ	FIN	AAC	CAS	CAH
Multiple R-squared	0.2721	0.005496	0.1374	0.4969	0.07396	0.5143	0.1281	0.3759	0.02289	0.2014	0.08108

Adjusted R-squared	2.72e-01	0.002096	1.36e-01	0.4394	0.07093	0.3704	0.1211	0.01934	0.01996	0.1564	0.07838
F-statistic	477.2	1.617	100.4	8.642	24.42	3.573	18.13	1.054	7.826	4.476	30.08
p-value (F-statistic)	< 2.2e-16	0.1149	< 2.2e-16	4.07e-08	< 2.2e-16	0.00599	< 2.2e-16	0.4443	1.93e-10	7.29e-05	< 2.2e-16
***NOCT*:rs13116075 , Chr4:139008878: A>G**											
Ancestry	**EUR**	**AFR**	**AMR**	**EAS**	**SAS**	**MDE**	**AJ**	**FIN**	**AAC**	**CAS**	**CAH**
Multiple R-squared	0.001848	0.003495	0.04217	0.1926	0.0117	0.2291	0.00869	0.4587	0.02382	0.1338	0.00723
Adjusted R-squared	0.001066	8.79E-05	0.04065	0.1004	0.008469	0.0007037	0.0006552	0.1493	0.02089	0.08501	0.004317
F-statistic	2.363	1.026	27.76	2.088	3.62	1.003	1.082	1.483	8.152	2.742	2.482
p-value (F-statistic)	0.01543	0.4139	< 2.2e-16	0.04845	0.0003401	0.4565	0.3735	0.248	6.05e-11	0.007637	0.01109
***CASS4* :rs6024870 , Chr20:56422512: G>A**											
Ancestry	**EUR**	**AFR**	**AMR**	**EAS**	**SAS**	**MDE**	**AJ**	**FIN**	**AAC**	**CAS**	**CAH**
Multiple R-squared	0.005269	0.007486	0.04021	NA	0.003431	0.1763	0.006623	NA	0.001632	0.09397	0.006616
Adjusted R-squared	0.00449	0.004093	0.03868	NA	0.0001714	-0.06771	-0.001429	NA	-0.001357	0.04292	0.003701
F-statistic	6.763	2.206	26.41	NA	1.053	0.7226	0.8225	NA	0.546	1.841	2.27
p-value (F-statistic)	6.97e-09	0.02438	< 2.2e-16	NA	0.3939	0.6703	0.5827	NA	0.8223	0.07417	0.02034
***LRRC37A*: rs2732703 , Chr17:46275856: T>G**											
Ancestry	**EUR**	**AFR**	**AMR**	**EAS**	**SAS**	**MDE**	**AJ**	**FIN**	**AAC**	**CAS**	**CAH**
Multiple R-squared	0.007924	0.03166	0.0766	NA	0.01559	0.3186	0.009751	0.2984	0.03179	0.03813	0.008379
Adjusted R-squared	0.007147	0.02835	0.07514	NA	0.01237	0.1167	0.001725	-0.1026	0.02889	-0.01606	0.00547
F-statistic	10.2	9.563	52.3	NA	4.843	1.578	1.215	0.7442	10.97	0.7036	2.88
p-value (F-statistic)	2.70e-14	4.14e-13	< 2.2e-16	NA	6.11e-06	0.1781	0.2866	0.6539	2.44e-15	0.688	0.0034
Interaction_model <- glm(PHENO ~ Protective/disease-modifying variant * APOE_STATUS (ε4 carriers) +SEX + AGE + PC1 + PC2 + PC3 + PC4 + PC5, data = data, family = binomial)											
***19q13.31* : rs10423769, chr19:43100929:G>A**											
Ancestry	**EUR**	**AFR**	**AMR**	**EAS**	**SAS**	**MDE**	**AJ**	**FIN**	**AAC**	**CAS**	**CAH**
β	-0.318826	-4.166e-01	4.690e-01	NA	NA	NA	0.4752150	NA	-0.1165126	NA	1.433e-01
Std. Error	0.708762	1.989e-01	4.654e-01	NA	NA	NA	0.6321330	NA	0.1847811	NA	2.718e-01
z value	-0.450	-2.095	1.008	NA	NA	NA	0.752	NA	-0.631	NA	0.527
p-value	0.652829	0.03619	0.313568	NA	NA	NA	0.4522	NA	0.52834	NA	0.598
***APOE* :rs449647, Chr19:44905307: A>T**											
Ancestry	**EUR**	**AFR**	**AMR**	**EAS**	**SAS**	**MDE**	**AJ**	**FIN**	**AAC**	**CAS**	**CAH**
β	-0.270222	-1.446e-02	-2.146e-01	-1.31410	-0.41516	-17.92462	-0.3188204	NA	-0.2674458	-0.28649	9.553e-02
Std. Error	0.105633	1.494e-01	1.423e-01	2.47308	0.50059	2971.41668	0.2675842	NA	0.1298759	1.13455	1.405e-01
z value	-2.558	-0.097	-1.508	-0.531	-0.829	-0.006	-1.191	NA	-2.059	-0.253	0.680
p-value	0.010524	0.92290	0.131445	0.595167	0.406911	0.9952	0.2335	NA	0.03947	0.800642	0.497
***TOMM40* :rs11556505, Chr19:44892887: C>T**											
Ancestry	**EUR**	**AFR**	**AMR**	**EAS**	**SAS**	**MDE**	**AJ**	**FIN**	**AAC**	**CAS**	**CAH**
β	3.048e-01	-1.037e-01	4.267e-01	-1.84506	-4.491e-02	-10.51189	0.2308497	50.468	0.0356318	-0.77961	1.709e-01
Std. Error	9.651e-02	2.173e-01	2.268e-01	1.98319	4.240e-01	2444.76386	0.3406798	3499.371	0.1757837	1.42636	1.915e-01
z value	3.158	-0.477	1.882	-0.930	-0.106	-0.004	0.678	0.014	0.203	-0.547	0.892
p-value	0.001586	0.63323	0.059868	0.35219	0.915644	0.9966	0.49802	0.988	0.83937	0.584671	0.372
***NOCT*:rs13116075 , Chr4:139008878: A>G**											
Ancestry	**EUR**	**AFR**	**AMR**	**EAS**	**SAS**	**MDE**	**AJ**	**FIN**	**AAC**	**CAS**	**CAH**
β	-8.385e-02	-9.510e-01	7.260e-02	NA	0.26446	-3.81041	-0.2010145	-5.534e+02	0.5046073	-1.385e+00	2.081e-01
Std. Error	8.649e-02	4.567e-01	2.000e-01	NA	0.49224	2.97982	0.2679418	2.485e+06	0.2500546	3.982e+03	2.023e-01
z value	-0.969	-2.082	0.363	NA	0.537	-1.279	-0.750	0.000	2.018	0.000	1.029
p-value	0.33232	0.03732	0.716605	NA	0.591085	0.2010	0.4531	1.000	0.043592	0.99972	0.304
***CASS4* :rs6024870 , Chr20:56422512: G>A**											
Ancestry	**EUR**	**AFR**	**AMR**	**EAS**	**SAS**	**MDE**	**AJ**	**FIN**	**AAC**	**CAS**	**CAH**
β	-2.061e-01	-9.165e-02	1.605e-01	NA	-0.14240	-17.90442	-0.2340929	NA	0.1964231	-1.592e+01	-4.237e-01
Std. Error	1.121e-01	2.531e-01	2.191e-01	NA	1.25526	2623.75448	0.2607464	NA	0.2011266	1.264e+03	2.070e-01
z value	-1.839	-0.362	0.732	NA	-0.113	-0.007	-0.898	NA	0.977	-0.013	-2.047
p-value	0.065858	0.71722	0.463960	NA	0.909678	0.9946	0.3693	NA	0.32876	0.989947	0.0406
***LRRC37A*: rs2732703 , Chr17:46275856: T>G**											
Ancestry	**EUR**	**AFR**	**AMR**	**EAS**	**SAS**	**MDE**	**AJ**	**FIN**	**AAC**	**CAS**	**CAH**
β	-6.999e-03	3.639e-01	3.705e-01	NA	-0.14487	-0.20542	-0.3455109	NA	-0.2151586	-15.95199	8.660e-02
Std. Error	9.252e-02	5.076e-01	1.866e-01	NA	0.67397	2.71021	0.2772725	NA	0.2525163	1689.31304	2.043e-01
z value	-0.076	0.717	1.985	NA	-0.215	-0.076	-1.246	NA	-0.852	-0.009	0.424
p-value	0.939698	0.47344	0.047098	NA	0.829805	0.9396	0.2127	NA	0.39418	0.992466	0.672
Interaction_model <- glm(PHENO ~ Protective/disease-modifying variant * APOE_STATUS (ε4/ε4) +SEX + AGE + PC1 + PC2 + PC3 + PC4 + PC5, data = data, family = binomial)											
***19q13.31* : rs10423769, chr19:43100929:G>A**											
Ancestry	**EUR**	**AFR**	**AMR**	**EAS**	**SAS**	**MDE**	**AJ**	**FIN**	**AAC**	**CAS**	**CAH**
β	4.911e+00	-1.297e-01	3.386e-01	NA	NA	NA	6.083e+00	NA	-0.0109888	NA	-1.266e-01
Std. Error	6.962e+01	1.673e-01	6.141e-01	NA	NA	NA	1.881e+02	NA	0.2050836	NA	2.496e-01
z value	0.071	-0.775	0.551	NA	NA	NA	0.032	NA	-0.054	NA	-0.507
p-value	0.9438	0.43845	0.581417	NA	NA	NA	0.974203	NA	0.95727	NA	0.611873
***APOE* :rs449647, Chr19:44905307: A>T**											
Ancestry	**EUR**	**AFR**	**AMR**	**EAS**	**SAS**	**MDE**	**AJ**	**FIN**	**AAC**	**CAS**	**CAH**
β	1.985e-01	3.898e-02	-6.337e-01	7.06937	-6.024637	NA	NA	NA	-0.4090145	7.737451	-6.361e-02
Std. Error	3.794e-01	1.254e-01	2.572e-01	872.99690	187.324036	NA	NA	NA	0.1331053	688.428638	1.580e-01
z value	0.523	0.311	-2.464	0.008	-0.032	NA	NA	NA	-3.073	0.011	-0.403
p-value	0.6008	0.755964	0.013753	0.993539	0.974343	NA	NA	NA	0.00212	0.991033	0.687273
***TOMM40* :rs11556505, Chr19:44892887: C>T**											
Ancestry	**EUR**	**AFR**	**AMR**	**EAS**	**SAS**	**MDE**	**AJ**	**FIN**	**AAC**	**CAS**	**CAH**
β	-2.868e-01	-1.377e-01	-3.712e-02	-7.69280	0.007408	5.14876	-0.0208789	-1.042e+02	-0.1621401	-7.062849	-1.911e-01
Std. Error	8.295e-02	1.929e-01	1.600e-01	868.83988	0.678194	2797.44372	0.2745880	4.104e+05	0.1470801	727.699410	1.585e-01
z value	-3.457	-0.714	-0.232	-0.009	0.011	0.002	-0.076	0.000	-1.102	-0.010	-1.206
p-value	0.000546	0.475415	0.816578	0.992936	0.991284	0.9985	0.9394	1.000	0.27029	0.992256	0.227812
***NOCT*:rs13116075 , Chr4:139008878: A>G**											

Ancestry	EUR	AFR	AMR	EAS	SAS	MDE	AJ	FIN	AAC	CAS	CAH
β	-1.127e-01	-9.841e-01	-6.797e-01	NA	-5.758776	NA	0.1252292	-1.879e+02	0.1242855	NA	-4.846e-01
Std. Error	9.604e-02	5.861e-01	3.291e-01	NA	194.924414	NA	0.4583244	2.720e+05	0.2773288	NA	2.277e-01
z value	-1.173	-1.679	-2.065	NA	-0.030	NA	0.273	-0.001	0.448	NA	-2.128
p-value	0.2407	0.093119	0.038876	NA	0.976431	NA	0.78467	0.999	0.6540	NA	0.033307

***CASS4*:rs6024870 , Chr20:56422512: G>A**

Ancestry	EUR	AFR	AMR	EAS	SAS	MDE	AJ	FIN	AAC	CAS	CAH
β	-6.057e-02	1.027e-01	-2.230e-02	NA	-5.08657	-6.10931	-0.3588910	NA	0.1269480	-6.715306	-1.595e-02
Std. Error	1.280e-01	2.424e-01	2.954e-01	NA	267.70598	2797.44428	0.3317201	NA	0.2388870	727.699378	2.476e-01
z value	-0.473	0.424	-0.075	NA	-0.019	-0.002	-1.082	NA	0.531	-0.009	-0.064
p-value	0.6361	0.671842	0.939822	NA	0.984841	0.9983	0.27929	NA	0.59513	0.992637	0.949

***LRRC37A*: rs2732703 , Chr17:46275856: T>G**

Ancestry	EUR	AFR	AMR	EAS	SAS	MDE	AJ	FIN	AAC	CAS	CAH
β	8.827e-02	1.163e-01	2.941e-02	NA	-5.49623	5.63281	-0.1043969	-1.440e+02	0.1053866	NA	-9.881e-02
Std. Error	1.098e-01	4.620e-01	2.958e-01	NA	218.06889	2797.44375	0.4237288	2.458e+05	0.2752225	NA	2.549e-01
z value	0.804	0.252	0.099	NA	-0.025	0.002	-0.246	-0.001	0.383	NA	-0.388
p-value	0.4215	0.801198	0.920809	NA	0.979892	0.9984	0.80539	1.000	0.70178	NA	0.698256

Conditional_model <- glm(PHENO ~ Protective/disease-modifying variant + APOE_STATUS (ε3/ε3) +SEX + AGE + PC1 + PC2 + PC3 + PC4 + PC5, data = data, family = binomial)

***19q13.31*: rs10423769, chr19:43100929:G>A**

Ancestry	EUR	AFR	AMR	EAS	SAS	MDE	AJ	FIN	AAC	CAS	CAH
A1	A	A	A	A	A	A	A	A	A	A	A
P	0.519217	0.508413	0.199474	NA	NA	0.720057	0.490052	NA	0.964739	NA	0.164826
OR (L95_U95)	0.8154 (0.438395_1.51662)	0.936386 (0.770645_1.13777)	1.28189 (0.877204_1.87326)	NA	NA	2.4021 (0.0199145_289.744)	0.816622 (0.459406_1.45159)	NA	0.995971 (0.832748_1.19119)	NA	0.835427 (0.648215_1.07671)

***APOE*:rs449647, Chr19:44905307: A>T**

Ancestry	EUR	AFR	AMR	EAS	SAS	MDE	AJ	FIN	AAC	CAS	CAH
A1	T	T	T	T	T	T	T	T	T	T	T
P	3.76e-09	0.000378349	0.843162	0.409142	0.744106	0.139232	0.00257546	0.220398	0.536594	0.790636	0.861296
OR (L95_U95)	0.774637 (0.71158_0.843283)	1.29317 (1.12226_1.49011)	1.01206 (0.898723_1.13969)	2.50033 (0.283747_22.0325)	0.946117 (0.678434_1.31942)	0.284837 (0.0539092_1.50498)	0.716756 (0.577206_0.890043)	.0239989 (6.15937e-05_9.3507	1.03989 (0.918565_1.17724)	0.880239 (0.34325_2.25731)	1.01175 (0.887525_1.15335)

***TOMM40*:rs11556505, Chr19:44892887: C>T**

Ancestry	EUR	AFR	AMR	EAS	SAS	MDE	AJ	FIN	AAC	CAS	CAH
A1	T	T	T	T	T	T	T	T	T	T	T
P	1.28e-38	0.50568	0.000100648	0.931317	0.915364	0.49694	0.0510141	0.649767	0.00876273	0.688142	0.00210663
OR (L95_U95)	1.8019 (1.64879_1.96923)	1.07366 (0.870902_1.32363)	1.43907 (1.19788_1.72882)	0.946922 (0.273961_3.27296)	1.02006 (0.707227_1.47127)	2.39398 (0.192874_29.7145)	1.33147 (0.998739_1.77504)	1.46764 (0.280186_7.68758)	1.25301 (1.05855_1.4832)	0.779007 (0.230106_2.63727)	1.33379 (1.11007_1.60258)

***NOCT*:rs13116075 , Chr4:139008878: A>G**

Ancestry	EUR	AFR	AMR	EAS	SAS	MDE	AJ	FIN	AAC	CAS	CAH
A1	G	G	G	G	G	G	G	G	G	G	G
P	0.0921729	0.579616	0.111308	0.658318	0.983852	0.690334	0.205043	0.492526	0.159087	0.150941	0.597635
OR (L95_U95)	1.07181 (0.988702_1.1619)	1.13268 (0.728891_1.76017)	0.863958 (0.721646_1.03434)	0.332386 (0.00252112_43.822)	0.99567 (0.654064_1.51569)	1.79433 (0.101114_31.8415)	0.850932 (0.66294_1.09223)	3.02023 (0.128563_70.9521)	0.844765 (0.667978_1.06834)	0.139696 (0.00952012_2.04987)	0.951919 (0.792742_1.14306)

***CASS4*:rs6024870 , Chr20:56422512: G>A**

Ancestry	EUR	AFR	AMR	EAS	SAS	MDE	AJ	FIN	AAC	CAS	CAH
A1	A	A	A	A	A	A	A	A	A	A	A
P	0.198142	0.2355	0.895017	NA	0.733941	0.317651	0.132828	NA	0.699167	0.495562	0.816941
OR (L95_U95)	0.93265 (0.838685_1.03714)	0.862441 (0.675375_1.10132)	1.01298 (0.83637_1.22689)	NA	1.17936 (0.455509_3.05349)	5.57921 (0.191552_162.501)	0.824916 (0.641812_1.06026)	NA	1.0385 (0.857433_1.2578)	0.442584 (0.0424502_4.61436)	1.02274 (0.845421_1.23726)

***LRRC37A*: rs2732703 , Chr17:46275856: T>G**

Ancestry	EUR	AFR	AMR	EAS	SAS	MDE	AJ	FIN	AAC	CAS	CAH
A1	G	G	G	G	G	G	G	G	G	G	G
P	0.0159467	0.469114	0.499232	NA	0.765328	0.888089	0.141762	0.630329	0.722082	0.859343	0.603084
OR (L95_U95)	0.899139 (0.824667_0.980338)	0.836562 (0.516022_1.35621)	0.942435 (0.793531_1.11928)	NA	0.915569 (0.513034_1.63394)	0.841499 (0.0760695_9.30887)	0.821482 (0.631937_1.06788)	2.02936 (0.113665_36.2319)	0.956198 (0.747062_1.22388)	0.86174 (0.166199_4.4681)	0.951874 (0.790384_1.14636)

r2_model <- lm(Protective/disease-modifying variant ~ APOE_STATUS (ε3/ε3) +SEX + AGE + PC1 + PC2 + PC3 + PC4 + PC5, data = data)

***19q13.31*: rs10423769, chr19:43100929:G>A**

Ancestry	EUR	AFR	AMR	EAS	SAS	MDE	AJ	FIN	AAC	CAS	CAH
Multiple R-squared	0.005739	0.007209	0.01396	NA	NA	0.08435	0.003222	NA	0.009795	NA	0.01382
Adjusted R-squared	0.00496	0.003815	0.01239	NA	NA	-0.187	-0.004857	NA	0.006832	NA	0.01093
F-statistic	7.369	2.124	8.924	NA	NA	0.3109	0.3988	NA	3.305	NA	4.776
p-value (F-statistic)	8.001e-10	0.03061	3.205e-12	NA	NA	0.9552	0.9215	NA	0.000918	NA	7.57e-06

***APOE*:rs449647, Chr19:44905307: A>T**

Ancestry	EUR	AFR	AMR	EAS	SAS	MDE	AJ	FIN	AAC	CAS	CAH
Multiple R-squared	0.01708	0.0417	0.004391	0.09561	0.01963	0.1763	0.03854	0.557	0.03105	0.05191	0.01413
Adjusted R-squared	0.01631	0.03843	0.002812	-0.007747	0.01642	-0.06776	0.03074	0.3038	0.02815	-0.001505	0.01124
F-statistic	22.18	12.73	2.781	0.925	6.122	0.7224	4.945	2.2	10.71	0.9718	4.885
p-value (F-statistic)	< 2.2e-16	< 2.2e-16	0.004543	0.5015	7.75e-08	0.6704	5.12e-06	0.09414	6.324e-15	0.4605	5.251e-06

***TOMM40*:rs11556505, Chr19:44892887: C>T**

Ancestry	EUR	AFR	AMR	EAS	SAS	MDE	AJ	FIN	AAC	CAS	CAH
Multiple R-squared	0.3606	0.007793	0.4079	0.3913	0.2386	0.542	0.3094	0.2094	0.02844	0.2733	0.1477
Adjusted R-squared	0.3601	0.004401	0.4069	0.3217	0.2361	0.4063	0.3038	-0.2424	0.02554	0.2324	0.1452
F-statistic	720.1	2.297	434.3	5.625	95.83	3.994	55.26	0.4635	9.782	6.676	59.09
p-value (F-statistic)	< 2.2e-16	0.01888	< 2.2e-16	1.641e-05	< 2.2e-16	0.003104	< 2.2e-16	0.862	1.769e-13	2.122e-07	< 2.2e-16

***NOCT*:rs13116075 , Chr4:139008878: A>G**

Ancestry	EUR	AFR	AMR	EAS	SAS	MDE	AJ	FIN	AAC	CAS	CAH
Multiple R-squared	0.001859	0.003644	0.04211	0.111	0.01174	0.2214	0.01144	0.4719	0.02385	0.1566	0.007509
Adjusted R-squared	0.001077	0.0002378	0.04059	0.009412	0.008509	-0.009359	0.003432	0.1701	0.02093	0.1091	0.004597
F-statistic	2.378	1.07	27.72	1.093	3.632	0.9594	1.428	1.564	8.162	3.297	2.579
p-value (F-statistic)	0.0148	0.3813	< 2.2e-16	0.3787	0.000327	0.4869	0.1803	0.2219	5.824e-11	0.001752	0.008361

***CASS4*:rs6024870 , Chr20:56422512: G>A**

Ancestry	EUR	AFR	AMR	EAS	SAS	MDE	AJ	FIN	AAC	CAS	CAH
Multiple R-squared	0.005369	0.008078	0.04074	NA	0.003443	0.1515	0.00698	NA	0.001719	0.09891	0.006579
Adjusted R-squared	0.00459	0.004687	0.03922	NA	0.0001839	-0.09995	-0.001069	NA	-0.001269	0.04814	0.003665

F-statistic	6.891	2.382	26.78	NA	1.056	0.6025	0.8672	NA	0.5753	1.948	2.258
p-value (F-statistic)	4.416e-09	0.01484	< 2.2e-16	NA	0.391	0.7675	0.5438	NA	0.7989	0.05729	0.02107
***LRRC37A*: rs2732703 , Chr17:46275856: T>G**											
Ancestry	**EUR**	**AFR**	**AMR**	**EAS**	**SAS**	**MDE**	**AJ**	**FIN**	**AAC**	**CAS**	**CAH**
Multiple R-squared	0.007884	0.03186	0.07655	NA	0.01582	0.3186	0.01161	0.3118	0.03262	0.03457	0.008172
Adjusted R-squared	0.007107	0.02855	0.07509	NA	0.0126	0.1167	0.003595	-0.08138	0.02972	-0.01982	0.005262
F-statistic	10.15	9.626	52.27	NA	4.915	1.578	1.449	0.793	11.27	0.6356	2.808
p-value (F-statistic)	3.267e-14	3.314e-13	< 2.2e-16	NA	4.801e-06	0.178	0.172	0.6178	8.397e-16	0.7468	0.004226
Interaction_model <- glm(PHENO ~ Protective/disease-modifying variant * APOE_STATUS (ε3/ε3) +SEX + AGE + PC1 + PC2 + PC3 + PC4 + PC5, data = data, family = binomial)											
***19q13.31*: rs10423769, chr19:43100929:G>A**											
Ancestry	**EUR**	**AFR**	**AMR**	**EAS**	**SAS**	**MDE**	**AJ**	**FIN**	**AAC**	**CAS**	**CAH**
β	-1.891e-01	1.515e-01	-4.637e-02	NA	NA	NA	-0.2460774	NA	0.0276379	NA	-6.915e-02
Std. Error	3.283e-01	1.026e-01	2.072e-01	NA	NA	NA	0.3012184	NA	0.0948777	NA	1.294e-01
z value	-0.576	1.476	-0.224	NA	NA	NA	-0.817	NA	0.291	NA	-0.535
p-value	0.5647	0.139875	0.822920	NA	NA	NA	0.413963	NA	0.77082	NA	0.593
***APOE*:rs449647, Chr19:44905307: A>T**											
Ancestry	**EUR**	**AFR**	**AMR**	**EAS**	**SAS**	**MDE**	**AJ**	**FIN**	**AAC**	**CAS**	**CAH**
β	3.175e-01	-1.109e-01	1.077e-01	0.32804	0.14264	9.32091	0.1583801	-1.113e+01	-0.0482250	0.237946	-3.086e-03
Std. Error	4.275e-02	7.697e-02	6.468e-02	1.21693	0.18349	1470.20000	0.1127026	1.886e+05	0.0665037	0.473634	6.682e-02
z value	7.426	-1.441	1.665	0.270	0.777	0.006	1.405	0.000	-0.725	0.502	-0.046
p-value	1.12e-13	0.149703	0.095932	0.787494	0.436955	0.9949	0.159935	1.000	0.46836	0.615398	0.963
***TOMM40*:rs11556505, Chr19:44892887: C>T**											
Ancestry	**EUR**	**AFR**	**AMR**	**EAS**	**SAS**	**MDE**	**AJ**	**FIN**	**AAC**	**CAS**	**CAH**
β	-3.407e-01	-1.131e-01	-3.658e-01	-8.11183	-0.1715	4.27063	-0.3470659	-1.898e+02	-0.0667558	-7.288818	-2.756e-01
Std. Error	6.200e-02	1.187e-01	1.209e-01	986.15329	0.2004	1199.77469	0.1815237	1.617e+05	0.0960520	807.636366	1.044e-01
z value	-5.495	-0.953	-3.026	-0.008	-0.856	0.004	-1.912	-0.001	-0.695	-0.009	-2.639
p-value	3.92e-08	0.34073	0.002482	0.99344	0.392163	0.9972	0.05588	0.999	0.48706	0.992799	0.00831
***NOCT*:rs13116075 , Chr4:139008878: A>G**											
Ancestry	**EUR**	**AFR**	**AMR**	**EAS**	**SAS**	**MDE**	**AJ**	**FIN**	**AAC**	**CAS**	**CAH**
β	0.013728	4.872e-01	-4.247e-02	NA	-0.27401	1.38743	0.0965789	-3033.5	-0.1944697	4.893e-01	3.525e-02
Std. Error	0.041199	2.252e-01	9.446e-02	NA	0.21894	1.33996	0.1279155	435634.5	0.1258033	1.749e+03	9.380e-02
z value	0.333	2.164	-0.450	NA	-1.252	1.035	0.755	-0.007	-1.546	0.000	0.376
p-value	0.7390	0.030476	0.653013	NA	0.210747	0.3005	0.45024	0.994	0.122147	0.999777	0.707
***CASS4* :rs6024870 , Chr20:56422512: G>A**											
Ancestry	**EUR**	**AFR**	**AMR**	**EAS**	**SAS**	**MDE**	**AJ**	**FIN**	**AAC**	**CAS**	**CAH**
β	8.275e-02	1.323e-01	-6.875e-03	NA	0.47141	8.55974	0.0218776	NA	-0.0635796	8.357e+00	1.233e-01
Std. Error	5.391e-02	1.266e-01	1.032e-01	NA	0.60405	1309.66064	0.1280412	NA	0.1009443	7.479e+02	9.777e-02
z value	1.535	1.045	-0.067	NA	0.780	0.007	0.171	NA	-0.630	0.011	1.261
p-value	0.1248	0.296074	0.946866	NA	0.435140	0.9948	0.864331	NA	0.52879	0.991085	0.207
***LRRC37A*: rs2732703 , Chr17:46275856: T>G**											
Ancestry	**EUR**	**AFR**	**AMR**	**EAS**	**SAS**	**MDE**	**AJ**	**FIN**	**AAC**	**CAS**	**CAH**
β	2.905e-02	2.406e-01	-2.053e-01	NA	0.1459	-0.35985	0.0223262	NA	0.0386594	-0.418667	-1.656e-02
Std. Error	4.379e-02	2.542e-01	8.825e-02	NA	0.3104	1.29242	0.1328074	NA	0.1275317	0.865756	9.532e-02
z value	0.663	0.946	-2.327	NA	0.470	-0.278	0.168	NA	0.303	-0.484	-0.174
p-value	0.5071	0.34390	0.019977	NA	0.638221	0.7807	0.86650	NA	0.7618	0.628680	0.862

ADSP, Alzheimer's Disease Sequencing Project; EUR, European; AFR, African; AMR, American Admixed; AAC,African Admixed; AJ, Ashkenazi Jewish; CAS, Central Asian; EAS, Eastern Asian; SAS, South Asian; MDE, Middle Eastern; FIN, Finnish; CAH, Complex Admixture History; NA, Not available