Check for updates

Received: 28 May 2024 | Revised: 8 August 2024 | Accepted: 5 September 2024

DOI: 10.1002/alz.14283

Alzheimer's & Dementia®
THE JOURNAL OF THE ALZHEIMER'S ASSOCIATION

RESEARCH ARTICLE

Association of common and rare variants with Alzheimer's disease in more than 13,000 diverse individuals with whole-genome sequencing from the Alzheimer's Disease Sequencing Project

Wan-Ping Lee[1,2] | Seung Hoan Choi[3] | Margaret G. Shea[4] | Po-Liang Cheng[1,2] | Beth A. Dombroski[1] | Achilleas N. Pitsillides[3] | Nancy L. Heard-Costa[5,6] | Hui Wang[1,2] | Katia Bulekova[7] | Amanda B. Kuzma[1,2] | Yuk Yee Leung[1,2] | John J. Farrell[8] | Honghuang Lin[9] | Brian W. Kunkle[10,11] | Adam Naj[12] | Elizabeth E. Blue[13,14] | Frederick Nusetor[3] | Dongyu Wang[3] | Eric Boerwinkle[15,16] | William S. Bush[17,18] | Xiaoling Zhang[8,3] | Philip L. De Jager[19] | Josée Dupuis[3,20] | Lindsay A. Farrer[8,5,3,6,21,22] | Myriam Fornage[23,24] | Eden Martin[10,11,25] | Margaret Pericak-Vance[10,11,25] | Sudha Seshadri[26] | Ellen M. Wijsman[13,27,28] | Li-San Wang[1,2] | The Alzheimer's Disease Sequencing Project | Gerard D. Schellenberg[1,2] | Anita L. Destefano[3,5] | Jonathan L. Haines[17,18] | Gina M. Peloso[3]

Correspondence
Gina M. Peloso, Department of Biostatistics, Boston University School of Public Health, 800 Massachusetts Ave, Boston, MA 02118, USA.
Email: gpeloso@bu.edu

Funding information
National Institute on Aging

Abstract

INTRODUCTION: Alzheimer's disease (AD) is a common disorder of the elderly that is both highly heritable and genetically heterogeneous.

METHODS: We investigated the association of AD with both common variants and aggregates of rare coding and non-coding variants in 13,371 individuals of diverse ancestry with whole genome sequencing (WGS) data.

RESULTS: Pooled-population analyses of all individuals identified genetic variants at apolipoprotein E (*APOE*) and *BIN1* associated with AD ($p < 5 \times 10^{-8}$). Subgroup-specific analyses identified a haplotype on chromosome 14 including *PSEN1* associated with AD in Hispanics, further supported by aggregate testing of rare coding and non-coding variants in the region. Common variants in *LINC00320* were observed associated with AD in Black individuals ($p = 1.9 \times 10^{-9}$). Finally, we observed rare non-coding variants in the promoter of *TOMM40* distinct of *APOE* in pooled-population analyses ($p = 7.2 \times 10^{-8}$).

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2024 The Author(s). *Alzheimer's & Dementia* published by Wiley Periodicals LLC on behalf of Alzheimer's Association.

DISCUSSION: We observed that complementary pooled-population and subgroup-specific analyses offered unique insights into the genetic architecture of AD.

KEYWORDS
Alzheimer's disease, genetics, rare genetic variants, whole genome sequencing

Highlights

- We determine the association of genetic variants with Alzheimer's disease (AD) using 13,371 individuals of diverse ancestry with whole genome sequencing (WGS) data.
- We identified genetic variants at apolipoprotein E (*APOE*), *BIN1*, *PSEN1*, and *LINC00320* associated with AD.
- We observed rare non-coding variants in the promoter of *TOMM40* distinct of *APOE*.

1 | BACKGROUND

Alzheimer's disease (AD) is a progressive neurological disorder characterized by a decline in cognitive and memory functions, which ultimately results in the inability to carry out daily activities. It is the most common cause of dementia, affecting > 50 million people worldwide. This number is projected to almost triple by 2050, reaching 152 million people, as the baby boom generation (born between 1946 and 1964) has already begun to reach age 65 years and beyond.[1] The disease is more common among older individuals, with the risk increasing significantly after the age of 65 years. In the United States, it is estimated that ≈ 6.5 million people currently have AD, with a projection of nearly 14 million by 2050.[1]

Although the underlying multidimensional causes of AD are not fully understood, evidence suggests that genetics plays a crucial role in the development of the disease. Rare coding changes in presenilin 1 (*PSEN1*),[2] presenilin 2 (*PSEN2*),[3] and amyloid precursor protein (*APP*)[4–7] underlie autosomal dominant early-onset AD, while other coding changes in these genes are associated with an increased risk of late-onset AD. The apolipoprotein E (*APOE*) gene is the strongest susceptibility gene associated with AD,[8,9] with isoforms defined by common missense variants associated with large effects on AD risk. Individuals with one copy of the *APOE* ε4 allele have approximately a 3-fold increased risk of developing AD, while those with two copies of the ε4 allele are at approximately a 12-fold increased risk.[10] The presence of the *APOE* ε4 allele is also associated with an earlier onset of AD.[11] Each of these loci was first identified by family-based studies.

Recent large-scale genome-wide association studies using array-based genotyping and imputation have identified > 80 common genetic loci associated with AD.[12–14] It is estimated that 25% of phenotypic variation in AD remains unexplained by known genetic variants associated with AD,[15,16] suggesting that additional risk loci are yet to be discovered. While genotype arrays are a useful tool for studying genetic variants associated with AD, they have limitations when it comes to discovering rare or novel genetic variant associations with disease. Array-based genotyping relies on a predesigned set of probes that target specific genetic loci across the genome. Computational imputation may mitigate this limitation and improve the accuracy and coverage of array-based genotype data. However, imputation accuracy is directly determined by the size and quality of the reference panel and observed array data used, as well as the underlying patterns of genetic variation in the populations being studied. In contrast, whole genome sequencing (WGS) enables a full-spectrum exploration of short insertion/deletions (INDELs) and single nucleotide variants (SNVs) across the genome and provides a comprehensive view of an individual's genetic information, allowing for the testing of both common and rare genetic variants that may be unique to individuals or populations not previously observed.

The Alzheimer's Disease Sequencing Project (ADSP) is a collaborative research effort that uses WGS to identify both protective and risk genetic factors for AD. Data are collected from diverse individuals, recognizing that genetic risk, incidence, and prevalence rates can vary across populations. For example, *APOE* ε4 is more common in White individuals (≈ 24%) and Black individuals (≈ 26%) compared to Asian individuals (≈ 12%) or Hispanic individuals (≈ 15%) and the effect of *APOE* ε4 varies among populations.[17,18] The odds ratios of ε3/ε4 are 2.49 and 3.83 for African ancestry and European ancestry, respectively, while ε4/ε4 are 8.17 and 14.35, respectively.[19] Leveraging large-scale WGS from the ADSP, we performed association testing of single variants (minor allele frequency [MAF] > 0.5%) as well as aggregates of rare (MAF < 1%) coding and non-coding variants in up to 13,371 individuals ($N_{cases} = 6519$ and $N_{control} = 6852$).

2 | METHODS

2.1 | Study participants

Data from the ADSP are available to qualified investigators via the National Institute on Aging Genetics of Alzheimer's Disease Data

15525279, 0, Downloaded from https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.14283 by Massey University Library, Wiley Online Library on [18/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

RESEARCH IN CONTEXT

1. **Systematic review**: We examined peer-reviewed (PubMed) and preprint articles using whole genome sequencing (WGS) data to determine genetic variants associated with Alzheimer's disease (AD). We performed association analyses using Alzheimer's Disease Sequencing Project WGS data to determine common variants and aggregates of rare variants with AD.
2. **Interpretation**: We observed genetic variants at APOE, *BIN1*, *PSEN1*, and *LINC00320* associated with AD. Furthermore, we observed an aggregate of rare coding variants in *PSEN1* and aggregated rare non-coding variants in the promoter of *TOMM40* associated with AD.
3. **Future directions**: Our study provides insights into rare variations associated with AD. Future studies will be needed to functionally validate these variants.

Storage Site (NIAGADS: https://dss.niagads.org/). The current analyses focused on participants with WGS in the NIAGADS data named "R3 17K WGS Project Level VCF." WGS data have been generated in multiple cohorts as part of the ADSP. Samples were sequenced by multiple centers with different platforms and libraries (Table S1, Supplementary Methods in supporting information). The Genome Center for Alzheimer's Disease (GCAD) mapped short reads against the reference genome hg38 using BWA MEM,[20] called variants using the Genome Analysis Toolkit (GATK)[21] HaplotypeCaller for each sample, and then performed joint genotyping across all samples using GATK.[22] The GCAD quality control (QC) working group performed quality checks of variants and genotypes and assigned a quality annotation.[23] Details of studies included in the ADSP can be found at NIAGADS under dataset: NG00067 ADSP Umbrella Study (https://dss.niagads.org/datasets/ng00067/). Of the 16,905 individuals in the ADSP Release 3 data, individuals with unknown AD status and genetically identical individuals were excluded using GCAD recommendations for excluding genetically identical individuals. After removing these individuals, 13,371 individuals were available for association analysis with AD.

We created three groups for subgroup-specific analyses. Individuals who reported their race as White and their ethnicity as non-Hispanic or missing were classified as non-Hispanic White (NHW). Individuals who indicated any reported race and Hispanic ethnicity were classified as Hispanic (HIS). Individuals who reported their ethnicity as non-Hispanic or had missing ethnicity and their race as Black were classified as African American (AA). There were 179 individuals that were not classified into one of our subgroup-specific analyses. Using genetic similarity clustering did not substantially change the subgroups. The results of subgroup-specific analyses were also confirmed by groups clustered using genetic information.

2.2 | Sample clustering using genetic similarity

To cluster samples based on genetic information, we used the following approach. First, we calculated the Euclidean distance of PC1-2 between each sample and the three reference populations from the 1000 Genomes Project,[24,25] Europeans (EUR), Africans (AFR), and East Asians (EAS). Then, we used a second approach to create sample clusters, which involved conducting global ancestry inference analysis with GRAF-pop,[26] which uses 100,437 fingerprint single nucleotide polymorphisms (SNPs) and is a principal component analysis (PCA)-free method for ancestry inference. Based on the results of GRAF-pop, we clustered the samples into subgroups: African-, Latin-, and European-ancestry groups.

2.3 | AD phenotype definition

The ADSP provides different AD status variable definitions for participants included via case-control versus family-based studies. Additionally, distinct phenotype data are available for some augmentation studies. In the current analysis, for individuals in the ADSP case-control study, we defined AD cases as individuals with either prevalent or incident AD. Case-control individuals with no prevalent or incident AD were defined as controls and those with NA for status were defined as unknown. In the ADSP family phenotype file, possible values for the AD status variable include no dementia, definite AD, probable AD, possible AD, family-reported AD, other dementia, family reported no dementia, and unknown. For family-based individuals, we defined an AD case as either possible, probable, or definite AD. AD controls were defined as individuals coded as no dementia. We redefined individuals with family-reported AD, other dementia, or unknown status as missing AD status. The Alzheimer's Disease Neuroimaging Initiative phenotype data, which is part of the ADSP augmentation study, provides information on mild cognitive impairment (MCI) in addition to AD status. Individuals with a current diagnosis of MCI were excluded in the current study. Each study contributing data to ADSP (see supporting information) adjudicates their participants independently.

2.4 | Genotype calling and quality control

The data used in this study were processed by GCAD. GCAD processed all WGS samples received in FASTQ, BAM, or compressed BAM (CRAM) formats using the official GCAD/ADSP pipeline, VCPA1.1.[22] In brief, GCAD aligned the samples against GRCh38/hg38 with BWA-mem and called variants, including SNVs and short indels, using GATK 4.1.1. QC was performed on all samples and variants following the GCAD/ADSP QC pipeline.[23] Various filtering and quality-checking strategies were applied at both the genotype and variant levels. Each genotype was set to missing ("./.") if either the read depth (DP) was < 10 or the genotype quality (GQ) score was < 20. Variants of low quality were flagged but not excluded from the datasets. For variants, flags were applied in the following order: (1) variants in GATK low

15525279, 0, Downloaded from https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.14283 by Massey University Library, Wiley Online Library on [18/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

sequence quality tranches (variants without a FILTER value of "PASS" that are above the 99.8% VQSR Tranche), (2) monomorphic variants were flagged, (3) variants with high missing rate were flagged, and (4) variants with high DP were flagged. ABHet is a variant-level annotation estimating if biallelic variants match expected allelic ratios, with an ideal heterozygous variant having a value close to 0.5. Variants used for analysis were those without QC flags, with ABHet in the range of 0.25 to 0.75, and with a missing rate of < 5%.

2.5 | Principal component analysis

PCA is widely used for analyzing large datasets that have a high number of dimensions or features per observation. PCA is a statistical technique for reducing the dimensionality of a dataset, while still retaining as much of the original variation as possible. In genetic studies, PCA is commonly used to infer population structure in the data, because population structure is a major factor that affects sample genotypes. Typically, the top principal components (PCs) calculated from the genotype data reflect population structure among the individuals. To ensure accurate ancestry inference, we used PC-AiR in the GENESIS[27] package for a PCA, a tool that accounts for sample relatedness and thus provides accurate ancestry inference. We calculated PCs for the pooled samples using variants with MAF > 5%, call rate > 99%, GCAD provide variant flag (VFlag) = 0 (no exclusion),[23] RUTH[28] Hardy–Weinberg Equilibrium (HWE) p value > 10^{-4}, and excluded variants in high linkage disequilibrium (LD) regions (https://genome.sph.umich.edu/wiki/Regions_of_high_linkage_disequilibrium_(LD)) and variants in LD using an $r^2 < 0.1$. For the AA, HIS, and NHW subgroups, we calculated PCs using SNVs with MAF > 5%, VFlag = 0, and LD threshold r^2 (0.1).

To identify the locations of *PSEN1* p.G206A (rs63750082) carriers, we extended the PC calculation to include samples of the 1000 Genomes Project. The VCFs of the 1000 Genomes Project were downloaded and merged with ADSP R3 VCFs by bcftools. Population and super-population information of each sample was also downloaded from the 1000 Genomes Project. The MAF cutoff, >5%, was applied, and then PC calculation was performed.

2.6 | Covariates, analysis models, and single variant analysis

We included sex, technical sequencing variables (sequencing center, sequencing length, and polymerase chain reaction [PCR] status), and PCs (PC1–5 for subgroup-specific analysis and PCs associated with AD status for the pooled samples analysis) as covariates in all our models along with a genetic relationship matrix to adjust for relatedness among individuals. As a secondary model, we additionally adjusted for the number (0, 1, or 2) of *APOE* ε4 alleles and the number of *APOE* ε2 alleles. We tested each variant with a MAF > 0.5% for association with AD using the score test in the GENESIS[29] package to fit a penalized quasi-likelihood (PQL) approximation to the generalized linear mixed model. Variants that failed HWE by RUTH[28] (HWE_SLP_I < −4 or HWE_SLP_I > 4) in controls were excluded as well as variants in low complexity regions. We used a standard significance threshold of 5×10^{-8} for our single variant association analyses.

2.7 | Aggregates of rare variants analysis

To test aggregates of coding and non-coding rare variants, we implemented the STAAR pipeline[30] using both SNVs and INDELs. The STAAR pipeline is a set of routines for performing association analysis of large-scale WGS data using the STAAR framework[31] to aggregate rare variants using variant set analysis for both gene-centric coding and gene-centric non-coding analysis. We used the STAAR-O p value, which combines p values across multiple annotation-weighted variant set tests.[31]

The gene-centric coding analysis of the STAAR pipeline provides five genetic categories: putative loss of function (pLoF), missense, disruptive missense, pLoF and disruptive missense, and synonymous. The gene-centric non-coding analysis provides eight genetic categories: promoter or enhancer overlaid with Cap Analysis of Gene Expression (CAGE) or DNase Hypersensitivity (DHS) sites, untranslated region (UTR), upstream, downstream, and non-coding RNA genes. We set our significance threshold for our rare variant aggregation tests to be 1×10^{-7} (Bonferroni correction for testing ≈ 20,000 genes across five coding categories and eight non-coding categories). For gene-centric non-coding analysis, due to the known associations in *PSEN1* and *APOE* regions, we performed conditional analyses adjusting for *PSEN1* p.G206A (rs63750082) in chromosome 14 and *APOE* ε2 and ε4 alleles in chromosome 19 in the pooled samples analysis. To incorporate additional features of the STAAR pipeline, we created a GitHub repository that performs variant extraction and conditional analysis (see code availability).

2.8 | Global and local ancestry inference analysis

The global ancestry inference was performed using the GRAF-pop[26] tool, which uses 100,437 fingerprint SNPs and is a PCA-free method for ancestry inference. GRAF-pop provides results of comparable quality to PCA-based methods such as EIGENSTRAT, FastPCA, and FlashPCA2, while offering an ultra-fast running time. Genotypes were provided to GRAF-pop in VCF format. The tool assumes that each individual is a mixture of three ancestries: European (E), African (F), and Asian (A), and estimates the ancestral proportions *Pe*, *Pf*, and *Pa* using barycentric coordinates.

To infer local ancestry, specifically for the analysis in the 14q24 region, we used RFMIX[32] with 2504 reference genomes from the 1000 Genomes Project. This tool outperforms other methods for estimating local ancestry in complex admixture scenarios.[33] Prior to using RFMIX, we phased variants using SHAPEIT4.[34] In addition, we used PICARD (http://broadinstitute.github.io/picard/) to LiftOver coordinates from hg38 to hg19 as the genetic map of reference samples

15525279, 0, Downloaded from https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.14283 by Massey University Library, Wiley Online Library on [18/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

we used was against HG19 (https://mathgen.stats.ox.ac.uk/impute/1000GP_Phase3.html). RFMIX allows for a comprehensive understanding of the local ancestry composition in the specific region of interest and provides insights into the complex genetic makeup of diverse populations.

2.9 | INDEL experimental validation

PCR Primer Design: The genomic sequence for the INDEL variants was determined by submitting the chromosomal location of the variants to the December 2013 (GRCh38/hg38) version of the Genome Browser[35] (http://genome.ucsc.edu). Sequence surrounding the variants was extracted and used to design PCR primers. Primers were designed outside of the breakpoints to amplify across the insertion/deletion sequence. All PCR primer sequences were submitted to the Blast-like alignment tool (BLAT) to check for the uniqueness of the sequences. When available, samples from three individuals reported as homozygous or heterozygous for the variant were used for sequence validation along with one control (or reference) sample. When possible, samples from multiple families were used for validation.

PCR and Sanger Sequencing: Genomic DNA (≈ 50 ng) was amplified using a SimpliAmp Thermal Cycler (Applied Biosystems) in a 20 µL reaction volume with HotStarTaq Master Mix (Qiagen) in the presence of 2 µM primers (IDT). The PCR conditions used were: 95°C 15 minutes followed by 30 cycles of 95°C 20 seconds, 55°C 30 seconds, 72°C 2 minutes with a final extension of 72°C 7 minutes. The amplified PCR products were prepared for Sanger sequencing by adding ExoSAP-IT (USB) and incubating at 37°C for 45 minutes followed by 80°C for 15 minutes. The PCR products were then Sanger sequenced using the BigDye Terminator v3.1 Cycle Sequencing kit (Part No. 4336917, Applied Biosystems). The sequencing reaction contained BigDye Terminator v3.1 Ready Reaction Mix, 5X Sequencing Buffer, 5 M Betaine solution (Part No. B0300 Sigma), and 0.64 µM sequencing primer (IDT) in a total volume of 5 µL. The sequencing reaction was performed in a SimpliAmp Thermal Cycler (Applied Biosystems) using the following program: 96°C 1 minute followed by 25 cycles of 96°C 10 seconds, 50°C 5 seconds, 60°C 1 minute 15 seconds. The products were cleaned using XTerminator and SAM Solution (Applied Biosystems) with 30 minutes of shaking at 1800 rpm (36 g) followed by centrifugation at 1000 rpm (200g) for 2 minutes. The sequencing products were analyzed on a SeqStudio Genetic Analyzer (Applied Biosystems) and the sequencing traces were analyzed using Sequencher 5.4 (Gene Code).

2.10 | INDEL in silico validation

Based on our findings in this study, we encountered 12 significant INDELs that were experimentally validated as false positives. These INDELs were all situated within sequences that exhibited discrepancies between the Telomere-to-Telomere Consortium (T2T)[36] and GRCh38 (the reference genome used in our study). Consequently, INDELs located in these discrepant sequences between T2T and GRCh38 were excluded.

To address this issue, we initially identified regions by considering the coordinates of the INDELs along with their respective lengths and then extended these regions by ±10 bp. Subsequently, we used LiftOver, using the hg38-to-chm13v2 chain (available at https://hgdownload.soe.ucsc.edu/goldenPath/hs1/liftOver/hg38-chm13v2.over.chain.gz) to convert the regions to the corresponding coordinates in the T2T (chm13v2) assembly. In cases in which the LiftOver process was unsuccessful, the INDELs were excluded from further analysis. For the regions that successfully underwent LiftOver, we compared the sequences obtained from GRCh38 and T2T. Only when the sequences were found to be identical, we included the corresponding INDELs in our report, ensuring accuracy and reliability in our findings.

2.11 | Replication dataset

The replication dataset was formed based on ADSP R4 36k samples by: (1) calculating identity-by-descent (IBD) for each pair of samples, (2) filtering samples with IBD > 0.4, and (3) excluding samples from R3 17k. This resulted in 17,266 samples (4696 cases and 9782 controls; Table S2 in supporting information). The dataset includes 2176 Black individuals (AA, $N_{cases} = 307$, $N_{controls} = 1366$, $N_{unkown} = 503$), 7452 Hispanics (HIS, $N_{cases} = 1879$, $N_{controls} = 4343$, $N_{unkown} = 1230$), and 4879 non-Hispanic Whites (NHW, $N_{cases} = 2243$, $N_{controls} = 1617$, $N_{unkown} = 1019$).

3 | RESULTS

3.1 | Overview

We performed association analyses using 13,371 of the 16,905 individuals in the ADSP Release 3 data to discover common and rare genetic variants associated with AD. The individuals who were excluded from the analysis represent technical replicates, unexpected duplicates, and individuals with unknown AD status. In the 13,371 individuals, 177.6 million bi-allelic SNVs and 12.9 million bi-allelic INDELs were observed.

Given the ADSP data are composed of diverse individuals, we performed association testing across all participants (pooled samples, $N_{cases} = 6519$ and $N_{control} = 6852$) and within the three subgroups: AA ($N_{cases} = 1137$ and $N_{control} = 1707$), HIS ($N_{cases} = 1021$ and $N_{control} = 1988$), and NHW ($N_{cases} = 4230$ and $N_{control} = 3109$) defined by reported race and ethnicity (Figure 1; Table 1). We performed single variant association testing (MAF > 0.5%) as well as association testing of aggregates of rare (MAF < 1%) coding and non-coding variants within the pooled samples and each of the three subgroups (AA, HIS, and NHW). The pooled samples analysis is most powerful to detect associations when there are similar frequencies and effect sizes across subgroups while the subgroup-specific analyses are beneficial to detect subgroup-specific effects. We limited the analyses to bi-allelic

15525279, 0, Downloaded from https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.14283 by Massey University Library, Wiley Online Library on [18/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

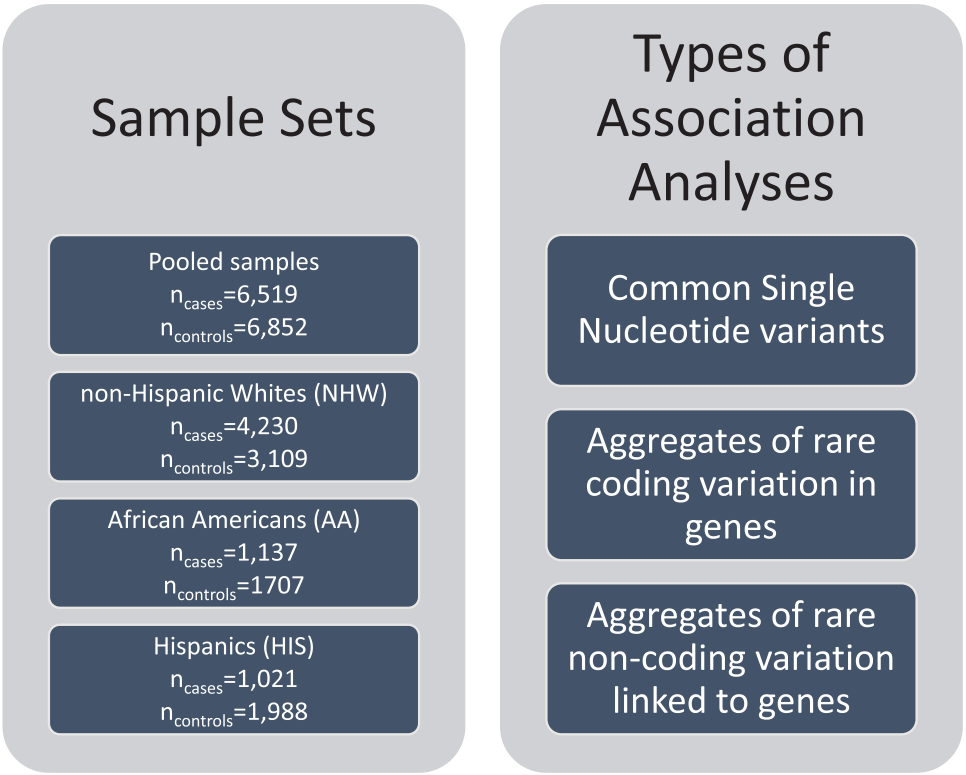

FIGURE 1 Study overview. Three types of association analyses in four sets of individuals were performed, pooled samples, non-Hispanic Whites (NHW), African Americans (AA), and Hispanics (HIS). The pooled samples set included all individuals in the NHW, AA, and HIS sets, plus individuals that were not defined to be in one of those subsets.

TABLE 1 Descriptive statistics of ADSP study samples used for analyses.

	AD				No AD			
	PS[b] (*N* = 6519; 100%)	AA (*N* = 1137; 17.4%)	Hispanic (*N* = 1021; 15.7%)	NHW (*N* = 4230; 64.9%)	PS[c] (*N* = 6852; 100%)	AA '(*N* = 1707; 24.9%)	Hispanic (*N* = 1988; 29.0%)	NHW (*N* = 3109; 45.4%)
Sex								
Female	3921 (60.1%)	798 (70.2%)	654 (64.1%)	2379 (56.2%)	4580 (66.8%)	1293 (75.8%)	1374 (69.1%)	1886 (60.7%)
Age								
Mean (SD)	71.6 (18.1)	74.6 (10.1)	74.4 (13.6)	70.1 (20.5)	74.5 (16.6)	75.1 (8.8)	65.1 (26.0)	80.2 (6.7)
APOE ε4 alleles[a]								
0	3022 (46.4%)	424 (37.3%)	677 (66.3%)	1876 (44.3%)	4608 (67.3%)	1112 (65.1%)	1550 (78.0%)	2005 (64.5%)
1	2917 (44.7%)	547 (48.1%)	288 (28.2%)	2016 (47.7%)	1993 (29.1%)	555 (32.5%)	415 (20.9%)	1004 (32.3%)
2	570 (8.7%)	166 (14.6%)	56 (5.5%)	338 (8.0%)	164 (2.4%)	40 (2.3%)	23 (1.2%)	100 (3.2%)
APOE ε2 alleles[a]								
0	6068 (93.1%)	995 (87.5%)	930 (91.1%)	4027 (95.2%)	5930 (86.5%)	1309 (76.7%)	1788 (89.9%)	2877 (92.5%)
1	427 (6.6%)	136 (12.0%)	89 (8.7%)	198 (4.7%)	800 (11.7%)	375 (22.0%)	196 (9.9%)	224 (7.2%)
2	14 (0.2%)	6 (0.5%)	2 (0.2%)	5 (0.1%)	35 (0.5%)	23 (1.3%)	4 (0.2%)	8 (0.3%)

Abbreviations: AA, Black African American; AD, Alzheimer's disease; *APOE*, apolipoprotein E; NHW, non-Hispanic White; PS, pooled samples; SD, standard deviation; WGS, whole genome sequencing.

[a] ε2 and ε4 are derived from *APOE* genotype not WGS.

[b] 131 (2.0%) samples are reported other ethnicities.

[c] 48 (0.7%) samples are reported other ethnicities.

15525279, 0, Downloaded from https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.14283 by Massey University Library, Wiley Online Library on [18/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

SNVs and INDELs after preliminary analyses showed false-positive associations across the genome for multi-allelic variants.

Sample clustering by using genetic similarity, based on the Euclidean distance of PC1-2 between each sample and the reference populations from the 1000 Genomes Project, indicated that 92.94% and 7.06% of reported AA samples were found to be closer to the AFR and EUR reference populations, respectively (Figure S1 in supporting information). Additionally, 8.28%, 0.003%, and 91.69% of reported HIS samples were closer to the AFR, EAS, and EUR reference populations, respectively. Similarly, 0.08%, 0.01%, and 99.90% of reported NHW samples were closer to the AFR, EAS, and EUR reference populations, respectively. This result aligns with the analysis from GRAF-pop, which revealed that 97.78% of reported AA samples were inferred to have a predominantly African ancestry, while 99.43% of reported NHW samples were inferred to have a predominantly European ancestry (Figure S2, Table S3 in supporting information). Regarding reported HIS samples, 74.74% were inferred to have Latin American 1 ancestry, 20.51% had African American ancestry, 2.43% had Latin American 2 ancestry, and 1.46% had European ancestry. Based on the results of GRAF-pop, we clustered the samples African and African American as African-ancestry group ($N = 3453$), Latin American 1 and Latin American 2 as Latin-ancestry group ($N = 2403$), and European as European-ancestry group ($N = 7440$). Subsequently, we conducted association analyses within these genetically defined groups. Nevertheless, the findings from these analyses revealed no notable differences compared to the previous subgroup analyses based on self-reported ethnicity (Figure S3 in supporting information).

3.2 | Association of single variants with AD

To determine individual SNVs and INDELs associated with AD status, we performed individual variant association testing. As expected, we found the strongest associations at the *APOE* locus (chr19:44,905,796-44,909,393), the major genetic risk factor for AD.[8,37] These associations were observed in the pooled samples analysis as well as the analysis within each of the three subgroups (Figures 2 and S4 in supporting information). We observed that the ε4 haplotype (the alternative allele at rs429358 and reference allele at rs7412) is more common in AD cases as well as most frequent in AA individuals and least frequent in HIS individuals (Table S4 in supporting information). While the odds of AD were higher in the AA individuals, the 95% confidence interval for the AA individuals overlaps with the confidence interval for the NHW individuals (Figure 2). We observed the ε2 haplotype (the reference allele at rs429358 and alternative allele at rs7412) was enriched in controls and had a lower frequency (frequency = 0.05) than the ε4 haplotype (frequency = 0.24). The ε2 haplotype is most frequent in AA individuals, followed by HIS individuals, and least frequent in NHW individuals. We confirmed the association with the *APOE* locus with the replication data (Table S5 in supporting information, pooled samples $p = 3.4 \times 10^{-148}$, AA $p = 2.2 \times 10^{-13}$, HIS $p = 1.7 \times 10^{-36}$, and NHW $p = 9.5 \times 10^{-85}$).

We also observed a genome-wide (GW) significant ($p < 5 \times 10^{-8}$) association of *BIN1* (rs4663105, MAF = 0.47, odds ratio [OR] = 1.17, $p = 3.2 \times 10^{-9}$) with AD status in the pooled samples analysis. Previous studies have identified *BIN1* as an AD susceptibility gene after *APOE*.[12,14,38] After adjusting for *APOE* ε4 (rs429358) and ε2 (rs7412) alleles on chromosome 19, the association of *BIN1* on chromosome 2 remained largely the same (OR = 1.17, $p = 3.1 \times 10^{-9}$). *BIN1* variants were also associated with AD in the NHW subgroup (rs4663105, MAF = 0.43, OR = 1.22, $p = 1.2 \times 10^{-8}$, Tables 2 and S6 in supporting information). Although the association did not reach statistical significance, the *BIN1* variant showed a consistent direction of effect in the AA and HIS subgroups (AA: rs4663105, MAF = 0.41, OR = 1.16, $p = 0.0064$; HIS: rs4663105, MAF = 0.40, OR = 1.10, $p = 0.15$). We replicated the *BIN1* association in pooled samples and NHW samples (pooled samples $p = 8.3 \times 10\text{-}03$, NHW: $p = 2.2 \times 10^{-02}$).

In AA individuals, we observed an association between variants in *LINC00320* with AD (rs144204759, MAF = 0.018, OR = 3.4, $p = 1.9 \times 10^{-8}$; Tables 2 and S7, and Figure S5A in supporting information). This variant has a much lower frequency in NHW individuals (gnomAD frequency = 0.005) and *LINC00320* was previously implicated in AD[39] by a distinct variant in the International Genomics of Alzheimer's Project (IGAP) genome-wide association study (GWAS) of 25,170 AD cases and 41,052 cognitively normal controls. We determined the gene expression change using Agora (https://agora.adknowledgeportal.org/genes/ENSG00000224924/evidence/rna). Although it is not among the top genes, *LINC00320* appears to be differentially expressed in AD ($p = 5.8 \times 10^{-03}$ for the dorsolateral prefrontal cortex [DLPFC], $p = 1.1 \times 10^{-02}$ for the posterior cingulate cortex [PCC], and $p = 3.5 \times 10^{-05}$ for the parahippocampal cortex [PHC]). However, our genetic association was not replicated, which may be due to differences in AA sample structure or sample size between the discovery ($N_{case} = 1137$, $N_{control} = 1707$) and replication ($N_{case} = 307$, $N_{control} = 1366$) data. Additionally, three INDELs near *APOE* are associated with AD in the AA analysis (rs142042446, MAF = 0.040, OR = 2.27, $p = 5.7 \times 10^{-9}$; rs542555887, MAF = 0.12, OR = 1.77, $p = 2.0 \times 10^{-10}$; rs113492558, MAF = 0.07, OR = 2.56, $p = 1.7 \times 10^{-19}$); however, the reduction of these signals after conditioning *APOE* ε2 and ε4 alleles suggests that the impact of these INDELs is not distinct of these common *APOE* haplotypes.

We determined the association of all loci from a recent GWAS of AD[12] within our data. We observed that among of 143 SNVs reported, 75 have a $p < 0.05$ in our dataset (Table S8 in supporting information). However, spurious variants emerged from the association analysis, underlining the need for careful scrutiny of WGS data (Supplementary Methods, Table S9, Figure S6 in supporting information). Several SNVs in *ANK3* were identified as associated with AD but later proven to be false positives supported by low-quality sequencing reads. Similarly, numerous INDELs identified were not confirmed upon further scrutiny, particularly due to biases introduced by using a reference genome that may not accurately represent human genomic diversity. This underscores the necessity for rigorous verification and analysis methods to prevent misinterpretation of WGS data due to false positives.

15525279, 0, Downloaded from https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.14283 by Massey University Library, Wiley Online Library on [18/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

rs429358 (APOE e4 allele)

Group	MAF	OR	UCL	LCL	P-value
Pooled Samples	0.235	2.441	2.292	2.599	1.98e−170
NHW	0.258	2.398	2.208	2.605	1.04e−95
AA	0.267	2.785	2.468	3.144	6.67e−62
HIS	0.146	1.904	1.610	2.251	4.93e−14

rs7412 (APOE e2 allele)

Group	MAF	OR	UCL	LCL	P-value
Pooled Samples	0.051	0.546	0.484	0.616	6.56e−23
NHW	0.031	0.523	0.427	0.641	3.76e−10
AA	0.099	0.519	0.434	0.622	9.57e−13
HIS	0.052	0.767	0.584	1.008	0.0572

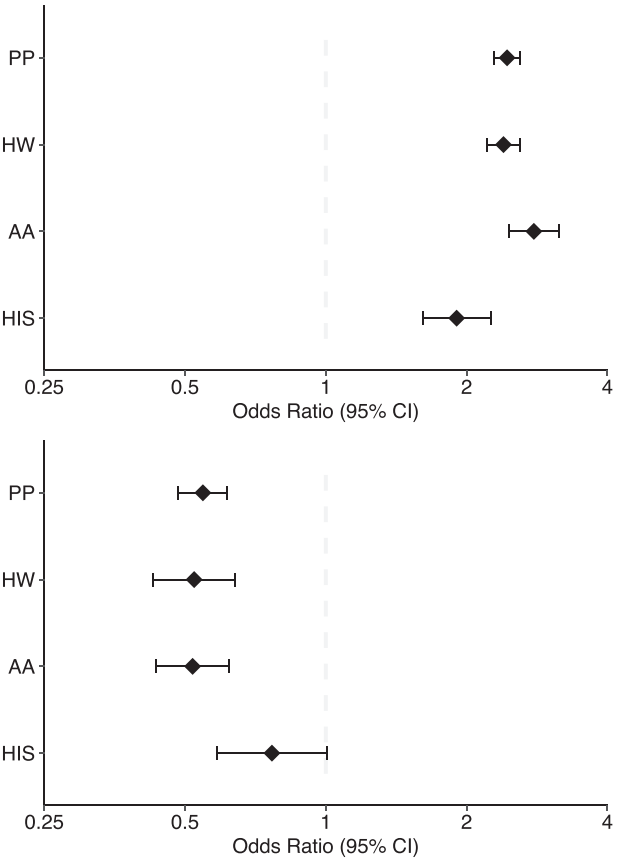

FIGURE 2 Association of *APOE* alleles with AD by subgroup. AA, African American; AD, Alzheimer's disease; *APOE*, apolipoprotein E; CI, confidence interval; HIS, Hispanic; LCL, lower confidence limit; MAF, minor allele frequency; NHW, non-Hispanic White; OR, odds ratio; PP, pooled population (samples); UCL, upper confidence limit.

TABLE 2 Genome-wide (P < 5 × 10−8) significant loci associated with AD.

Variants[a]	Gene	RSID	Group	MAF[b]	Odds ratio	*p*value[c]
2-127133851-A-C	*BIN1*	rs4663105	Pooled samples	0.470	1.16	3.2×10^{-9}
			NHW	0.427	1.22	1.2×10^{-8}
14-73615125-C-T	*(various)*	rs9671262	HIS	0.005	19.20	2.2×10^{-11}
19-44908684-T-C	*APOE*	rs429358	Pooled samples	0.230	2.44	2.0×10^{-170}
			AA	0.267	2.78	6.7×10^{-62}
			HIS	0.145	6.69	4.9×10^{-14}
			NHW	0.258	2.40	1.0×10^{-95}
21-20730315-G-A	*LINC00320*	rs144204759	AA	0.018	3.40	1.9×10^{-9}

Abbreviations: AA, Black or African American; *APOE*, apolipoprotein E; HIS, Hispanic; MAF, minor allele frequency; NHW, non-Hispanic White.
[a]Coordinates in GRCh38 for indexed variant.
[b]MAF for subgroup.
[c]Where more than one was significant for a linked gene, the most significant *p* value, either with or without *APOE* adjustment, is reported for each gene. The full lists are given in Tables S3–S5 in supporting information.

3.3 | 14q24 in Hispanics

From our single variant association testing, we observed a region from 14q24.2 to 14q24.3 (chr14:72,600,928-75,846,454), with 44 low-frequency variants (43 SNVs and one INDEL; MAF: 0.005–0.012; Tables 2 and S10 in supporting information) associated with AD across 13 genes in Hispanic individuals (Figures S4E, S4F, and S5B). These variants are not associated with AD in the AA subgroup despite a

15525279, 0, Downloaded from https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.14283 by Massey University Library, Wiley Online Library on [18/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

notably higher allele frequency (MAF: 0.007–0.041; Table S11 in supporting information). Single variant analyses were not conducted for these variants in the pooled samples and NHW subgroups because the MAF in these subgroups was below the 0.5% threshold (Table S11). We observed variants in 14q24.3 replicated ($p < 0.05$) in the Hispanic subgroup ($p = 2.4 \times 10^{-04}$).

This region contains *PSEN1* p.G206A (rs63750082 at 14:73192712), a known early-onset AD causal mutation.[40–43] After conditioning on *PSEN1* p.G206A, the association between AD and the other variants in the region was no longer significant ($p > 0.05$). *PSEN1* p.G206A was first identified in a few Caribbean Hispanic families,[40,41] and a follow-up study identified *PSEN1* p.G206A in 70 families of Caribbean Hispanic ancestry.[42] A more recent study of early-onset AD in Hispanics in Florida found that 13 out of 27 participants (48.1%) were *PSEN1* p.G206A carriers.[43] The allele frequency of this mutation is ultra-rare, only 1 in 1741 to 3790 individuals carries a *PSEN1* p.G206A in the Trans-Omics for Precision Medicine (TOPMed) Freeze 8[44] and Genome Aggregation Databases (gnomAD) v3.1.2[36] data, respectively. And while it is still ultra-rare, it is more common in Latino/Admixed Americans (1 in 423) within gnomAD v3.1.2. We observed 17 *PSEN1* p.G206A carriers in the entire ADSP R3 dataset (1 in 786), of which 16 individuals have AD, and all *PSEN1* p.G206A carriers are reported as Hispanic. Only one pair of *PSEN1* p.G206A carriers were inferred to be related (Figure S7 in supporting information). Among those with AD, we observed that *PSEN1* p.G206A carriers have an earlier age of diagnosis of AD (58.6 ± 7.6 years old) compared to non-carriers (74.7 ± 10.4 years old; $p = 1.1 \times 10^{-10}$).

Although previous studies have suggested that *PSEN1* p.G206A originated in Caribbean Hispanics,[40,41] we do not have detailed data on the geographic location of most ADSP participants. We applied PCA to decipher the origin of the *PSEN1* p.G206A allele. PCA captures the human population structure associated with ancestry.[45] Therefore, we placed *PSEN1* p.G206A carriers into a global context using PCA of carriers and reference samples in the 1000 Genomes Project[25] from 26 populations, including 139 Puerto Ricans. Expectedly, the 17 *PSEN1* p.G206A carriers are closer to Puerto Rican reference samples in the American group according to the Euclidean distance of the first five PCs (Figure S8 in supporting information); individuals with at least 21 of 43 haplotype-associated SNVs are similarly placed near Puerto Ricans in PC space with greater dispersion (Figure S9 in supporting information). A global ancestry analysis revealed that all *PSEN1* p.G206A carriers have a higher proportion of European (73.82% ± 6.85%) than African ancestry (16.05% ± 5.72%). However, we observed that the chr14 risk haplotype including *PSEN1* p.G206A is inherited on an African-derived haplotype (Figure S10 in supporting information). This conclusion is based on results from RFMIX of combining our dataset with 2504 reference genomes from the 1000 Genomes Project. As *PSEN1* p.G206A is not detected in any individual of the AA subgroup while the other chr14 haplotype-defining variants have higher allele frequencies in the AA subgroup; *PSEN1* p.G206A may be too rare to be detected on AA subgroup or may have arisen in Puerto Rico[46] locally based on a founder event of an African haplotype.

Of note, one female *PSEN1* p.G206A carrier, whose *APOE* genotype is ε3/ε3 and global ancestry is 76.94% European genomes, was diagnosed as an AD patient at the age of 74 years, 15 years older than the average age at onset of AD among *PSEN1* p.G206A carriers. This raises the question of whether any other protective variants countervail the impact of mutations in the haplotype. To the best of our knowledge, this is the first report of a significant association between AD and this 3Mbp haplotype, which was linked to onset age 20 years younger.[40,41]

3.4 | Association of aggregates of rare variants in genes

To determine whether aggregated of rare variants in genes associated with AD status, we performed gene-centric aggregates of coding variants for each protein-coding gene using five functional variant categories (pLoF, missense, disruptive missense, pLoF and disruptive missense, and synonymous; Figure S11 in supporting information). Unsurprisingly, we observed several significant gene-based associations ($p < 5 \times 10^{-7}$) in 14q24 in the HIS individuals (Table S12 in supporting information), which was also confirmed in the replication ($p < 0.15 \times 10^{-03}$, Table S5). Additionally, we observed that *PSEN1* was associated with AD ($p = 4.1 \times 10^{-8}$) in the pooled samples analysis when aggregating rare loss of function and disruptive missense variants (Figure 3). Several sensitivity analyses were performed in the pooled-population individuals. First, after excluding *PSEN1* p.G206A (rs63750082) from the gene-based tests, we observed a significant association between the aggregation of rare loss-of-function and disruptive missense variants in *PSEN1* and AD (OR 2.02, 95% confidence interval [CI] 1.3–3.0, $p = 7.8 \times 10^{-4}$). Second, we adjusted for *PSEN1* p.G206A, and observed almost identical results to the association without the adjustment of *PSEN1* p.G206A (OR 2.02, 95% CI 1.3–3.0, $p = 7.6 \times 10^{-4}$). Finally, none of the deleterious variants in *PSEN1* were in the list of significant chr14q24.2 variants reported in Table S11. These results suggest that there may be additional rare variants in *PSEN1* that contribute to the observed association with AD, particularly in the NHW subset.

We observed a few suggestive gene-based associations driven by rare coding variants ($5 \times 10^{-7} < p < 5 \times 10^{-5}$). Significant association signals between AD and *ABCA7* ($p = 5.4 \times 10^{-6}$) in the NHW analysis and *SPTLC2* ($p = 1.0 \times 10^{-5}$) in the HIS analysis were observed using multiple sets of rare coding variant aggregates (i.e., loss of function plus disruptive missense as well as missense, Table S12). There has been a growing interest in studying *ABCA7* due to accumulating in vitro and in vivo studies supporting the potential contribution to AD-related phenotypes.[29,47] In the NHW subgroup, there are 273 rare exonic variants in *ABCA7* (including 186 non-synonymous, 9 stop-gain, 7 frameshift, and 3 non-frameshift; Table S13 in supporting information), and 90 of them are not singletons (including 67 non-synonymous, 2 stop-gain, and 4 frameshift; MAF: 0.000068–0.0075). Two frameshift deletions, rs547447016 (chr19:1047508:AGGAGCAG:A; $N_{case} = 30$ and $N_{control} = 9$) and rs538591288 (chr19:1055908:CT:C; $N_{case} = 20$ and $N_{control} = 8$), have

15525279, 0, Downloaded from https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.14283 by Massey University Library, Wiley Online Library on [18/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Sample	N case	N control	N variant	cMAC	OR[95% CI]	Pvalue
Pooled	6519	6852	62	119	2.96[2,4.4]	4.07e–08
w/o rs63750082			61	102	2.02[1.3,3]	7.91e–04
Non–Hispanic White	4230	3109	40	58	2.01[1.2,3.4]	1.08e–02
w/o rs63750082			40	58	2.01[1.2,3.4]	1.08e–02
Hispanic	1021	1988	8	28	12.14[4.8,30.5]	1.12e–07
w/o rs63750082			7	11	0.56[0.1,2.1]	3.95e–01
African American	1137	1707	19	47	1.48[0.8,2.7]	1.97e–01
w/o rs63750082			19	47	1.48[0.8,2.7]	1.97e–01

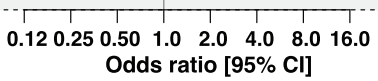

FIGURE 3 Association of *PSEN1* with AD by subgroup, with and without p.G206A (rs63750082) included. We performed aggregated-based association testing for variants within *PSEN1* in the pooled sample as well as in each subgroup. We then removed the p.G206A variant (rs63750082) from the aggregation-based test. We observed that the odds ratio and *p* value reduced significantly after excluding rs63750082 in the Hispanic subset but the results remained consistent in the NHW and AA subsets. AA, Black or African American; AD, Alzheimer's disease; CI, confidence interval; cMAC, cumulative minor allele counts across the variants contributing to the analysis; N variant, number of variants contributing to the analysis; NHW, non-Hispanic White.

been reported in previous studies[48–50] and experimentally validated. Another two, rs779501556 (chr19:1046404:CGT:C; $N_{case} = 2$ and $N_{control} = 0$) and rs745871063 (chr19:1054250:AG:A; $N_{case} = 0$ and $N_{control} = 2$), are novel and ultra-rare. Nine of the stop-gain SNVs in our *ABCA7* test (Table S13) were previously identified associated with AD[48,51–55] or autism.[56]

SPTLC2 (chr14:77,505,997-77,616,637) encodes a protein involved in the biosynthesis of sphingolipids, and there is some evidence to suggest that changes in sphingolipid metabolism may be associated with AD.[57,58] Specifically, previous studies have suggested that alterations in the levels of specific sphingolipids may contribute to the development or progression of the disease.[57,59] However, the role of *SPTLC2* in AD is not yet fully understood, and research in this area is ongoing. Our WGS study identified a suggestive significant association between the aggregate of rare disruptive missense and loss-of-function variants in *SPTLC2* (Table S14 in supporting information) with AD in the HIS subgroup ($p = 1.0 \times 10^{-5}$ and 1.0×10^{-5}), respectively.

3.5 | Association of aggregates of rare variants in non-coding sets

We next performed gene-centric aggregates of rare non-coding variants using eight functional variant categories (promoter or enhancer overlaid with CAGE or DHS sites, UTR, upstream, downstream, and non-coding RNA genes) to determine whether aggregated of rare non-coding variants associated with AD (Figure S11). We observed rare non-coding variant aggregates associated with AD near *TOMM40* ($p = 7.2 \times 10^{-8}$) and *PSEN1* ($p = 2.4 \times 10^{-11}$ to 3.2×10^{-8}) regions in the pooled samples and HIS individuals (Table 3). After conditioning on the number of *APOE* ε2 and ε4 alleles, there was an attenuation of the results for *TOMM40* in the pooled samples analysis, but the association remained ($p_{adj} < 7.1 \times 10^{-6}$). In our replication data, we observed *TOMM40* associated with AD ($p = 2.8 \times 10^{-03}$). Compared to the pooled samples analysis, we observed that none of subgroup-specific associations reached a GW significant level for our rare variant aggregation tests ($p < 1 \times 10^{-7}$); however, we observed suggestive association signals in the AA and NHW individuals with variants in the promoter of *TOMM40*, $p = 2.8 \times 10^{-4}$ ($p_{adj} = 0.010$) and 4.4×10^{-3} ($p_{adj} = 0.018$), respectively, whereas there was not an association in HIS individuals ($p = 0.41$, $p_{adj} = 0.45$). After adjusting for *PSEN1* p.G206A, we observed that the association of rare non-coding variants in the promoter of *PSEN1* was no longer significant ($p > 0.05$) in pooled samples and HIS analyses. This suggests these rare non-coding variants are on the same haplotype as the rare coding variants in *PSEN1*; this was confirmed by local ancestry analysis (Figure S10).

4 | DISCUSSION

WGS data allow for the testing of both common and rare genetic variations that may be unique to individuals or populations and provide a powerful tool for identifying genetic variation that may be missed by traditional genotyping methods. Because the ADSP includes the largest sample of diverse participants with WGS and AD status, we designed our study to perform association testing across all participants and within subgroups defined by reported race/ethnicity. The pooled samples analysis is most powerful for detecting associations when there are similar effects across ancestries, while the subgroup-specific analyses are able to detect subgroup-specific effects. Through our analysis framework, we were able to adequately control for the diversity within ADSP and leveraged ADSP WGS to learn more about known loci for AD, including population-specific genetic signals.

15525279, 0, Downloaded from https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.14283 by Massey University Library, Wiley Online Library on [18/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 Aggregates of rare variants in non-coding sets with AD.

Group[a]	Gene name	Chr	Category	# Variants	STAAR-O *p* value[a]
Pooled samples	*TOMM40*	19	Promoter (DHS)	134	7.2×10^{-8}
Pooled samples	*ELMSAN1*	14	Enhancer (DHS)	1133	1.8×10^{-9}
Pooled samples	*EIF2B2*	14	Enhancer (DHS)	1240	3.2×10^{-8}
Pooled samples	*MIR4505*	14	ncRNA	7	2.4×10^{-11}
HIS	*PTGR2*	14	Promoter (CAGE and DHS)	7	8.85×10^{-12}
HIS	*ELMSAN1*	14	Enhancer (DHS)	366	3.1×10^{-11}
HIS	*PTGR2*	14	Enhancer (CAGE)	153	5.9×10^{-11}
HIS	*ACOT6*	14	Enhancer (DHS)	33	4.2×10^{-10}
HIS	*ELMSAN1*	14	Promoter (DHS)	55	8.8×10^{-10}
HIS	*ACOT4*	14	Promoter (CAGE)	6	9.3×10^{-10}
HIS	*ACOT4*	14	Enhancer (CAGE)	9	9.8×10^{-10}

Abbreviations: CAGE, Cap Analysis of Gene Expression; DHS, DNase Hypersensitivity; HIS, Hispanic.
[a]Where more than one rare non-coding aggregate was significant for a linked gene, the most significant *p* value is reported for each category type.

As anticipated, our single variant GWAS showed the strongest associations with AD in *APOE* across all groups as well as *BIN1* for the pooled samples and NHW. We observed that genetic variation in *LINC00320* was associated with AD in AA. We identified 44 genetic variants on 14q24 in HIS (MAF: 0.005–0.012), and aggregates of rare variants analysis in HIS and pooled samples confirmed the region, which includes p.G206A in *PSEN1*, a well-known early-onset AD mutation.[40–43] Our PCA analysis indicates that *PSEN1* p.G206A carriers are closer to Puerto Ricans, consistent with previous studies,[46] while local ancestry analysis, however, pointed out that the local haplotype is derived from African genomes.

The analysis of coding rare variants identified suggestive associations in *ABCA7* in NHW and *SPTLC2* in HIS with AD. Rare coding variants in *ABCA7* have been found associated with AD risk.[60–62] From our analysis, a deeper investigation of *ABCA7* revealed two frameshift deletions, rs547447016 and rs538591288, that were reported in previous studies[48–50] and validated here, while rs779501556 and rs745871063 are variants newly associated with AD in the current analysis, indicating distinct *ABCA7* variants in multiple subgroups associated with AD risk. The non-coding rare variants in the promoter of *TOMM40* were identified as significant in the pooled samples analysis and confirmed to be distinct from the *APOE* haplotypes.

There are some limitations to our study. First, defining subgroups based on reported race and ethnicity, which was used in this study, may not be recommended as the best practice;[63] we observed a high consistency in reported and genetically defined subgroups, particularly for the AA and NHW subgroups (Table S3). Furthermore, humans cannot be grouped into discrete categories, and we acknowledge heterogeneity in our subgroup-specific analyses, and that the heterogeneity varies among the subgroups. We clustered samples using the Euclidean distance of PCA between each individual and the three 1000 Genomes Project reference populations, EUR, AFR, and EAS (Figure S1). The PCA-based clusters are similar to the results using global ancestry analysis, 97.78% reported AA are inferred as having a majority of African ancestry, and 99.43 reported NHW are inferred as having a majority of European ancestry. As to the reported HIS, 74.74%, 20.51%, 2.43%, and 1.46% are inferred by GRAF-pop, a global ancestry inference, as having ancestries of Latin American 1, African American, Latin American 2, and European, respectively from GRAF-pop.[26] Association analyses on genetically defined subgroups gave similar results. Second, despite being the largest sample with WGS ascertained on AD status, we still have limited power to detect associations with AD, particularly with rare non-coding variants in which the aggregation unit is not as well defined as with rare coding variants. We acknowledge there are much larger sample sizes using GWAS for AD to assess the contribution of single variants.[12] In a prior study, we restricted to variants previously identified by GWAS loci and implicated genes using WGS data.[64] Third, each study contributing to the ADSP adjudicates their participants, which could lead to differences in the definition of AD across the contributing studies and a diminished power to detect associations with AD status.

In conclusion, we have comprehensively analyzed up to 13,371 diverse individuals with WGS for AD and observed common and rare variants associated with AD.

AFFILIATIONS

[1]Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA

[2]Penn Neurodegeneration Genomics Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA

[3]Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA

[4]Biostatistics and Epidemiology Data Analytics Center, Boston University School of Public Health, Boston, Massachusetts, USA

15525279, 0, Downloaded from https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.14283 by Massey University Library, Wiley Online Library on [18/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

[5]Department of Neurology, Boston University Chobanian & Avedisian School of Medicine, Boston, Massachusetts, USA

[6]Framingham Heart Study, Framingham, Massachusetts, USA

[7]Research Computing Services, Information Services & Technology, Boston University, Boston, Massachusetts, USA

[8]Biomedical Genetics, Department of Medicine, Boston University Medical School, Boston, Massachusetts, USA

[9]Department of Medicine, University of Massachusetts Chan Medical School, Worcester, Massachusetts, USA

[10]John P. Hussman Institute for Human Genomics, Miami, Florida, USA

[11]John T. Macdonald Department of Human Genetics, Miami, Florida, USA

[12]Department of Biostatistics, Epidemiology, and Informatics, Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, USA

[13]Department of Medicine, Division of Medical Genetics, University of Washington, Seattle, Washington, USA

[14]Brotman Baty Institute for Precision Medicine, Seattle, Washington, USA

[15]Department of Epidemiology, Human Genetics and Environmental Sciences, School of Public Health, University of Texas Health Science Center at Houston, Houston, Texas, USA

[16]Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas, USA

[17]Cleveland Institute for Computational Biology, Cleveland, Ohio, USA

[18]Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, Ohio, USA

[19]Center for Translational and Computational Neuroimmunology, Columbia University Medical Center, New York, New York, USA

[20]Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, Quebec, Canada

[21]Department of Ophthalmology, Department of Medicine, Boston University Medical School, Boston, Massachusetts, USA

[22]Department of Epidemiology, Boston University School of Public Health, Boston, Massachusetts, USA

[23]Brown Foundation Institute of Molecular Medicine, McGovern Medical School, University of Texas Health Science Center at Houston, Houston, Texas, USA

[24]Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, Texas, USA

[25]University of Miami Miller School of Medicine, Miami, Florida, USA

[26]Glenn Biggs Institute for Alzheimer's & Neurodegenerative Diseases, University of Texas Health Science Center, San Antonio, Texas, USA

[27]Department of Biostatistics, University of Washington, Seattle, Washington, USA

[28]Department of Genome Sciences, University of Washington, Seattle, Washington, USA

AUTHOR CONTRIBUTIONS

Wan-Ping Lee, Seung Hoan Choi, Margaret G. Shea, Beth A. Dombroski, Frederick Nusetor, and Gina M. Peloso performed statistical analyses. Nancy L. Heard-Costa, Amanda B. Kuzma, Lindsay A. Farrer, Margaret Pericak-Vance, Li-San Wang, Jonathan L. Haines, Gina M. Peloso, and Xiaoling Zhang performed phenotype acquisition and/or harmonization. Wan-Ping Lee, Achilleas N. Pitsillides, A.M.B., Yuk Yee Leung, Honghuang Lin, William S. Bush, Eden Martin, Elizabeth E. Blue, Xiaoling Zhang, Margaret Pericak-Vance, Li-San Wang, Anita L. Destefano, Jonathan L. Haines, and Gina M. Peloso performed genotype acquisition and/or QC. Wan-Ping Lee, Seung Hoan Choi, Po-Liang Cheng, Hui Wang, Honghuang Lin, Josée Dupuis, Lindsay A. Farrer, Elizabeth E. Blue, Xiaoling Zhang, Myriam Fornage, Eden Martin, Ellen M. Wijsman, Li-San Wang, Anita L. Destefano, Jonathan L. Haines, and Gina M. Peloso interpretated results. Wan-Ping Lee and Gina M. Peloso wrote the first draft of the manuscript. All authors read, critically revised, and approved the manuscript.

ACKNOWLEDGMENTS

See Supplementary Text. Wan-Ping Lee reports grant support from RF1-AG074328 and P30-AG072979. Honghuang Lin reports grant support from U01AG068221. William S. Bush, Eden Martin, and Jonathan L. Haines report grant support from U01 AG058654. Josée Dupuis, Myriam Fornage, Elizabeth E. Blue, Anita L. Destefano, and Gina M. Peloso report grant support from U01 AG058589. Lindsay A. Farrer reports grant support from U01-AG058654, R01-AG048927, U19-AG068753, U01-AG062602, U01-AG032984, U54-AG058654, and P30-AG072978. Margaret Pericak-Vance reports grant support from U01 AG072547 and U01 AG070864. Ellen M. Wijsman reports grant support from U01AG058589 and P30AG066509. Li-San Wang reports grant support from U24-AG041689, U54-AG052427, U01-AG032984, U01-AG058654, and P30AG072979. Xiaoling Zhang reports grant support from U01AG072577 and R01AG080810. This work was supported by National Institute on Aging.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest. Author disclosures are available in the supporting information.

CONSENT STATEMENT

All human subjects included in this study provided informed consent.

DATA AVAILABILITY STATEMENT

ADSP R3 VCFs: https://dss.niagads.org/datasets/ng00067/1000 Genomes Project VCFs: http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/1000G_2504_high_coverage/working/20201028_3202_raw_GT_with_annot/

Full summary statistics will be made available through NIAGADS upon publication.

CODE AVAILABILITY

https://github.com/wanpinglee/CADRE_CHARGE_ADSP17K
https://github.com/seuchoi/STAARpipeline_plugin

REFERENCES

1. 2023 Alzheimer's disease facts and figures. *Alzheimers Dement*. 2023;19(4):1598-1695.
2. Sherrington R, Rogaev EI, Liang Y, et al. Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. *Nature*. 1995;375:754-760.
3. Sherrington R, Froelich S, Sorbi S, et al. Alzheimer's disease associated with mutations in presenilin 2 is rare and variably penetrant. *Hum Mol Genet*. 1996;5:985-958.
4. Selkoe DJ, Podlisny MB. Deciphering the genetic basis of Alzheimer's disease. *Annu Rev Genomics Hum Genet*. 2002;3:67-99.

15525279, 0, Downloaded from https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.14283 by Massey University Library, Wiley Online Library on [18/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

5. Goate A, Chartier-Harlin MC, Mullan M, et al. Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. *Nature*. 1991;349:704-706.
6. Chartier-Harlin MC, Crawford F, Houlden H, et al. Early-onset Alzheimer's disease caused by mutations at codon 717 of the beta-amyloid precursor protein gene. *Nature*. 1991;353:844-846.
7. Cruchaga C, Haller G, Chakraverty S, et al. Rare variants in APP, PSEN1 and PSEN2 increase risk for AD in late-onset Alzheimer's disease families. *PLoS One*. 2012;7:e31039.
8. Pericak-Vance MA, Bebout JL. Linkage studies in familial Alzheimer disease: evidence for chromosome 19 linkage. *Am J Hum Genet*. 1991;48:1034-1050.
9. Corder EH, Saunders AM, Strittmatter WJ, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science*. 1993;261:921-923.
10. Bertram L, McQueen MB, Mullin K, Blacker D, Tanzi RE. Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. *Nat Genet*. 2007;39:17-23.
11. Raber J, Huang Y, Ashford JW. ApoE genotype accounts for the vast majority of AD risk and AD pathology. *Neurobiol Aging*. 2004;25:641-650.
12. Bellenguez C, Kucukali F, Jansen IE, et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat Genet*. 2022;54:412-436.
13. Wightman DP, Jansen IE, Savage JE, et al. A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. *Nat Genet*. 2021;53:1276-1282.
14. Kunkle BW, Grenier-Boley B, Sims R, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. *Nat Genet*. 2019;51:414-430.
15. Ridge PG, Mukherjee S, Crane PK, Kauwe JS, Alzheimer's Disease Genetics C. Alzheimer's disease: analyzing the missing heritability. *PLoS One*. 2013;8:e79771.
16. Wang H, Bennett DA, De Jager PL, Zhang QY, Zhang HY. Genome-wide epistasis analysis for Alzheimer's disease and implications for genetic risk prediction. *Alzheimers Res Ther*. 2021;13:55.
17. Farrer LA, Cupples LA, Haines JL, et al. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. *JAMA*. 1997;278:1349-1356.
18. Tang MX, Stern Y, Marder K, et al. The APOE-epsilon4 allele and the risk of Alzheimer disease among African Americans, whites, and Hispanics. *JAMA*. 1998;279:751-755.
19. Choi KY, Lee JJ, Gunasekaran TI, et al. APOE Promoter Polymorphism-219T/G is an effect modifier of the influence of APOE epsilon4 on Alzheimer's disease risk in a multiracial sample. *J Clin Med*. 2019;8:1236.
20. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*. 2013;1303.
21. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011;43:491-498.
22. Leung YY, Valladares O, Chou YF, et al. VCPA: genomic variant calling pipeline and data management tool for Alzheimer's Disease Sequencing Project. *Bioinformatics*. 2019;35:1768-1770.
23. Naj AC, Lin H, Vardarajan BN, et al. Quality control and integration of genotypes from two calling pipelines for whole genome sequence data in the Alzheimer's disease sequencing project. *Genomics*. 2019;111:808-818.
24. Genomes Project C, Auton A, Brooks LD, et al. A global reference for human genetic variation. *Nature*. 2015;526:68-74.
25. Byrska-Bishop M, Evani US, Zhao X, et al. High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. *Cell*. 2022;185:3426-3440. e19.
26. Jin Y, Schaffer AA, Feolo M, Holmes JB, Kattman BL. GRAF-pop: a fast distance-based method to infer subject ancestry from multiple genotype datasets without principal components analysis. *G3 (Bethesda)*. 2019;9:2447-2461.
27. Zhang Y, Qi G, Park JH, Chatterjee N. Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits. *Nat Genet*. 2018;50:1318-1326.
28. Kwong AM, Blackwell TW, LeFaive J, et al. Robust, flexible, and scalable tests for Hardy-Weinberg equilibrium across diverse ancestries. *Genetics*. 2021;218:iyab044.
29. Hollingworth P, Harold D, Sims R, et al. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat Genet*. 2011;43:429-435.
30. STAARpipeline: an all-in-one rare-variant tool for biobank-scale whole-genome sequencing data. *Nat Methods*. 2022;19:1532-1533.
31. Li X, Li Z, Zhou H, et al. Dynamic incorporation of multiple in silico functional annotations empowers rare variant association analysis of large whole-genome sequencing studies at scale. *Nat Genet*. 2020;52:969-983.
32. Maples BK, Gravel S, Kenny EE, Bustamante CD. RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. *Am J Hum Genet*. 2013;93:278-288.
33. Uren C, Hoal EG, Moller M. Putting RFMix and ADMIXTURE to the test in a complex admixed population. *BMC Genet*. 2020;21:40.
34. Delaneau O, Zagury JF, Robinson MR, Marchini JL, Dermitzakis ET. Accurate, scalable and integrative haplotype estimation. *Nat Commun*. 2019;10:5436.
35. Kent WJ, Sugnet CW, Furey TS, et al. The human genome browser at UCSC. *Genome Res*. 2002;12:996-1006.
36. Chen S, Francioli LC, Goodrich JK, et al. A genomic mutational constraint map using variation in 76,156 human genomes. *Nature*. 2024;625:92-100. doi:10.1038/s41586-023-06045-0
37. Liu CC, Liu CC, Kanekiyo T, Xu H, Bu G. Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. *Nat Rev Neurol*. 2013;9:106-118.
38. Gao P, Ye L, Cheng H, Li H. The mechanistic role of bridging integrator 1 (BIN1) in Alzheimer's disease. *Cell Mol Neurobiol*. 2021;41:1431-1440.
39. Naj AC, Leonenko G, Jian X, et al. Genome-wide meta-analysis of late-onset Alzheimer's disease using rare variant imputation in 65,602 subjects identifies novel rare variant locus NCK2: the International Genomics of Alzheimer's Project (IGAP). *medRxiv*. 2021:2021031421253553. doi:10.1101/2021.03.14.21253553v1
40. Athan ES, Williamson J, Ciappa A, et al. A founder mutation in presenilin 1 causing early-onset Alzheimer disease in unrelated Caribbean Hispanic families. *JAMA*. 2001;286:2257-2263.
41. Rogaeva EA, Fafel KC, Song YQ, et al. Screening for PS1 mutations in a referral-based series of AD cases: 21 novel mutations. *Neurology*. 2001;57:621-625.
42. Lee JH, Kahn A, Cheng R, et al. Disease-related mutations among Caribbean Hispanics with familial dementia. *Mol Genet Genomic Med*. 2014;2:430-437.
43. Ravenscroft TA, Pottier C, Murray ME, et al. The presenilin 1 p.Gly206Ala mutation is a frequent cause of early-onset Alzheimer's disease in Hispanics in Florida. *Am J Neurodegener Dis*. 2016;5:94-101.
44. Taliun D, Harris DN, Kessler MD, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed program. *Nature*. 2021;590:290-299.
45. Novembre J, Johnson T, Bryc K, et al. Genes mirror geography within Europe. *Nature*. 2008;456:98-101.
46. Poblete J, Dokuru D, Coppola G, et al. Historical migration revealed through a case of autosomal dominant Alzheimer's disease. *P R Health Sci J*. 2019;38:144-147.

15525279, 0, Downloaded from https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.14283 by Massey University Library, Wiley Online Library on [18/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

47. Aikawa T, Holm ML, Kanekiyo T. ABCA7 and pathogenic pathways of Alzheimer's disease. *Brain Sci*. 2018;8:27.
48. May P, Pichler S, Hartl D, et al. Rare ABCA7 variants in 2 German families with Alzheimer disease. *Neurol Genet*. 2018;4:e224.
49. Le Guen Y, Belloy ME, Napolioni V, et al. A novel age-informed approach for genetic association analysis in Alzheimer's disease. *Alzheimers Res Ther*. 2021;13:72.
50. De Roeck A, Van den Bossche T, van der Zee J, et al. Deleterious ABCA7 mutations and transcript rescue mechanisms in early onset Alzheimer's disease. *Acta Neuropathol*. 2017;134:475-487.
51. Vardarajan BN, Ghani M, Kahn A, et al. Rare coding mutations identified by sequencing of Alzheimer disease genome-wide association studies loci. *Ann Neurol*. 2015;78:487-498.
52. Acosta-Uribe J, Aguillon D, Cochran JN, et al. A neurodegenerative disease landscape of rare mutations in Colombia due to founder effects. *Genome Med*. 2022;14:27.
53. Giau VV, Bagyinszky E, Yang YS, Youn YC, An SSA, Kim SY. Genetic analyses of early-onset Alzheimer's disease using next generation sequencing. *Sci Rep*. 2019;9:8368.
54. Cuyvers E, De Roeck A, Van den Bossche T, et al. Mutations in ABCA7 in a Belgian cohort of Alzheimer's disease patients: a targeted resequencing study. *Lancet Neurol*. 2015;14:814-822.
55. Teerlink CC, Miller JB, Vance EL, et al. Analysis of high-risk pedigrees identifies 11 candidate variants for Alzheimer's disease. *Alzheimers Dement*. 2022;18:307-317.
56. He Z, O'Roak BJ, Smith JD, et al. Rare-variant extensions of the transmission disequilibrium test: application to autism exome sequence data. *Am J Hum Genet*. 2014;94:33-46.
57. Czubowicz K, Jesko H, Wencel P, Lukiw WJ, Strosznajder RP. The role of ceramide and sphingosine-1-phosphate in Alzheimer's disease and other neurodegenerative disorders. *Mol Neurobiol*. 2019;56:5436-5455.
58. Le Stunff H, Veret J, Kassis N, et al. Deciphering the link between hyperhomocysteinemia and ceramide metabolism in Alzheimer-type neurodegeneration. *Front Neurol*. 2019;10:807.
59. Grimm MO, Grosgen S, Rothhaar TL, et al. Intracellular APP domain regulates serine-palmitoyl-CoA transferase expression and is affected in Alzheimer's disease. *Int J Alzheimers Dis*. 2011;2011:695413.
60. Stepler KE, Gillyard TR, Reed CB, Avery TM, Davis JS, Robinson RAS. ABCA7, a genetic risk factor associated with Alzheimer's disease risk in African Americans. *J Alzheimers Dis*. 2022;86:5-19.
61. Cukier HN, Kunkle BW, Vardarajan BN, et al. ABCA7 frameshift deletion associated with Alzheimer disease in African Americans. *Neurol Genet*. 2016;2:e79.
62. Holstege H, Hulsman M, Charbonnier C, et al. Exome sequencing identifies rare damaging variants in ATP8B4 and ABCA1 as risk factors for Alzheimer's disease. *Nat Genet*. 2022;54:1786-1794.
63. Khan AT, Gogarten SM, McHugh CP, et al. Recommendations on the use and reporting of race, ethnicity, and ancestry in genetic research: experiences from the NHLBI TOPMed program. *Cell Genom*. 2022;2:100155.
64. Wang Y, Sarnowski C, Lin H, et al. Key variants via the Alzheimer's Disease Sequencing Project whole genome sequence data. *Alzheimers Dement*. 2024;20:3290-3304.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Lee W-P, Choi SH, Shea MG, et al. Association of common and rare variants with Alzheimer's disease in more than 13,000 diverse individuals with whole-genome sequencing from the Alzheimer's Disease Sequencing Project. *Alzheimer's Dement*. 2024;1-14. https://doi.org/10.1002/alz.14283

15525279, 0, Downloaded from https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.14283 by Massey University Library, Wiley Online Library on [18/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License