

1 **Key variants via Alzheimer's Disease Sequencing Project whole genome sequence data**

2  
3 **Authors**

4 Yanbing Wang,<sup>a</sup> Chloé Sarnowski,<sup>\*a,b</sup> Honghuang Lin,<sup>c</sup> Achilleas N Pitsillides,<sup>a</sup> Nancy L  
5 Heard-Costa,<sup>a,d</sup> Seung Hoan Choi,<sup>a</sup> Dongyu Wang,<sup>a</sup> Joshua C Bis,<sup>e</sup> Elizabeth E Blue,<sup>f,g</sup>  
6 Alzheimer's Disease Neuroimaging Initiative (ADNI), Eric Boerwinkle,<sup>b</sup> Philip L De Jager,<sup>h,i</sup>  
7 Myriam Fornage,<sup>b,j</sup> Ellen M Wijsman,<sup>k</sup> Sudha Seshadri,<sup>d,l,m</sup> Josée Dupuis,<sup>a,n</sup> Gina M Peloso,<sup>a</sup>  
8 Anita L DeStefano<sup>a,d</sup>, for the Alzheimer's Disease Sequencing Project (ADSP)

9  
10 **Affiliations**

- 11 a. Department of Biostatistics, Boston University, School of Public Health, Boston, MA,  
12 USA
- 13 b. Human Genetics Center, Department of Epidemiology, Human Genetics, and  
14 Environmental Sciences, School of Public Health, The University of Texas Health  
15 Science Center at Houston, Houston, TX, USA
- 16 c. Department of Medicine, University of Massachusetts Chan Medical School,  
17 Worcester, MA, USA
- 18 d. The Framingham Heart Study, Framingham, MA, USA
- 19 e. Cardiovascular Health Research Unit, Department of Medicine, University of  
20 Washington, Seattle, WA, USA
- 21 f. Department of Medicine, Division of Medical Genetics, University of Washington,  
22 Seattle, WA, USA
- 23 g. Brotman Baty Institute, Seattle, WA, USA
- 24 h. Center for Translational & Computational Neuroimmunology, Department of  
25 Neurology, Columbia University Irving Medical Center, New York, NY, USA
- 26 i. Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Columbia  
27 University Irving Medical Center, New York, NY, USA

**NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.**

- 28 j. Brown Foundation Institute of Molecular Medicine, McGovern Medical School,  
29 University of Texas Health Science Center at Houston, Houston, TX, USA  
30 k. Div. of Medical Genetics and Dept. Biostatistics Statistical Genetics Lab, University of  
31 Washington, Seattle, WA, USA  
32 l. Glenn Biggs Institute for Alzheimer's & Neurodegenerative Diseases, The University  
33 of Texas Health Science Center at San Antonio, San Antonio, TX, USA  
34 m. Boston University School of Medicine, Department of Neurology, Boston, MA, USA  
35 n. Department of Epidemiology, Biostatistics and Occupational Health, School of  
36 Population and Global Health, McGill University, Montreal, Canada

37

38 Corresponding authors

39 Dr. Chloé Sarnowski, email: [Chloe.Sarnowski@uth.tmc.edu](mailto:Chloe.Sarnowski@uth.tmc.edu) (editorial office correspondence)

40 Dr. Anita L Destefano, email: [adestef@bu.edu](mailto:adestef@bu.edu)

41

42 Drs. Yanbing Wang and Chloé Sarnowski contributed equally to this work. Drs. Gina Peloso  
43 and Anita L DeStefano contributed equally to the supervision of this work.

44

45 **Abbreviations:** Alzheimer's disease (AD), the Alzheimer's Disease Sequencing Project  
46 (ADSP), Black/African-American (AA), Combined Multivariate and Collapsing (CMC),  
47 White/European-ancestry (EA), Genome-Wide Association Studies (GWAS), Genetic  
48 Relationship Matrix (GRM), Hispanic/Latino (HI), Linkage Disequilibrium (LD), Minor Allele  
49 Count (MAC), Minor Allele Frequency (MAF), Mild Cognitive Impairment (MCI), Principal  
50 Component Analysis (PCA), Quality Control (QC), Sequence Kernel Association Test (SKAT),  
51 variant-Set Test for Association using Annotation infoRmation (STAAR), Whole Genome  
52 Sequencing (WGS)

53

54

55

56 **Abstract**

57 INTRODUCTION: Genome-wide association studies (GWAS) have identified loci associated  
58 with Alzheimer's disease (AD) but did not identify specific causal genes or variants within those  
59 loci. Analysis of whole genome sequence (WGS) data, which interrogates the entire genome  
60 and captures rare variations, may identify causal variants within GWAS loci.

61 METHODS: We performed single common variant association analysis and rare variant  
62 aggregate analyses in the pooled population (N cases=2,184, N controls=2,383) and targeted  
63 analyses in sub-populations using WGS data from the Alzheimer's Disease Sequencing  
64 Project (ADSP). The analyses were restricted to variants within 100 kb of 83 previously  
65 identified GWAS lead variants.

66 RESULTS: Seventeen variants were significantly associated with AD within five genomic  
67 regions implicating the genes OARD1/NFYA/TREML1, JAZF1, FERMT2, and SLC24A4.  
68 KAT8 was implicated by both single variant and rare variant aggregate analyses.

69 DISCUSSION: This study demonstrates the utility of leveraging WGS to gain insights into AD  
70 loci identified via GWAS.

71

72 **Keywords:** Alzheimer's disease; Whole Genome Sequencing; Association Analyses; Single  
73 Nucleotide Variations; Diverse Populations; Genome Wide Association Study

74

75

76

77

78

79

80

81

82

83

## 84 1. Introduction

85 Alzheimer's disease (AD), the most common cause of dementia, has been ranked as the 6th  
86 leading cause of death in the United States and the 5th leading cause of death in older people  
87 ( $\geq 65$  years old). Although the role of genetic factors in the development of AD has been widely  
88 recognized, genome-wide association studies (GWAS) typically identify regions or loci rather  
89 than specific genes and/or variants. Additionally, the loci identified by GWAS only explain a  
90 small portion of the total heritability of AD ( $h^2_{AD} = [0.58-0.79]$ ).<sup>1</sup> Next-generation sequencing  
91 technology applied in diverse populations as part of the Alzheimer's Disease Sequencing  
92 Project (ADSP) may help to elucidate the genetic architecture of AD, and thus, aid in the  
93 development of effective strategies to diagnose, prevent and treat AD.<sup>2</sup>

94 A recent large GWAS totalling 111,326 clinically diagnosed/'proxy' AD cases and 677,663  
95 controls has identified over 70 loci associated with AD and related dementias.<sup>3</sup> However, the  
96 characterization of these loci remains incomplete. Leveraging whole genome sequence  
97 (WGS) data that encompasses the full spectrum of genetic variation including common and  
98 rare variants might identify important AD genes within these GWAS loci and provide a better  
99 understanding of the biological mechanisms involved in the pathophysiology of AD. Previous  
100 studies used WGS to identify genetic loci associated with AD.<sup>3-6</sup> A family-based study  
101 conducted in 2,247 subjects from NIMH/NIA with replication in 1,669 independent participants  
102 from the ADNI/ADSP identified 13 novel AD candidate loci with rare-variant signals (FNBP1L,  
103 SEL1L, LINC00298, PRKCH, C15ORF41, C2CD3, KIF2A, APC, LHX9, NALCN, CTNNA2,  
104 SYTL3, and CLSTN2).<sup>3</sup> More recently, the same team investigated association of groups of  
105 rare variants in the same datasets using a sliding-window approach and identified two novel  
106 genes (DTNB and DLG2) associated with AD.<sup>4</sup> Additional studies conducted in Asian  
107 populations highlighted the importance of increasing representation of understudied  
108 population groups and value of WGS to uncover population-specific genetic loci.<sup>5,6</sup>

109 In this work, we focused on deep interrogation of known AD GWAS loci<sup>7</sup> using the ADSP WGS  
110 data. The ADSP aims to identify protective or risk genetic contributors for AD in populations  
111 with diverse ancestry. The ADSP has generated single nucleotide variant and

112 insertion/deletion (indel) calls based on WGS data from 4,789 participants, which are publicly  
113 available (R1 data release <https://dss.niagads.org/datasets/ng00067-v1/>). The goal of the  
114 current study is not replication of prior GWAS findings as we are underpowered to do so. In  
115 addition, the ADSP sample in the current analyses is not independent of the sample used in  
116 Bellenguez et al.<sup>7</sup> Instead we aim to provide a more comprehensive look at GWAS loci.  
117 We conducted single variant association analyses and rare variant aggregation association  
118 tests using the R1 WGS data of ADSP to identify specific genetic variants, genes and non-  
119 coding regions associated with AD within previously identified AD loci. We also examined  
120 multi-ancestry evidence for AD associations through population-specific analyses in  
121 White/European-ancestry (EA), Black/African-American (AA) and Hispanic/Latino (HI)  
122 subgroups, and a multi-population meta-analysis. The insights gained from our analysis will  
123 contribute to a better understanding of the AD pathogenesis and to potentially identify new  
124 targets for AD drug and treatment.

125

## 126 **2. Methods**

### 127 2.1 Study Participants

128 Data from the ADSP is available to qualified investigators via the National Institute on Aging  
129 Genetics of Alzheimer's Disease Data Storage Site (NIAGADS) (<https://dss.niagads.org/>).  
130 This study was done under an approved NIAGADS research use statement and local  
131 Institutional Review Board approval. The current analyses focused on participants with WGS  
132 data in the NIAGADS file set named "R1 5K WGS Project Level VCF". WGS data have been  
133 generated in multiple cohorts as part of the ADSP. The ADSP data included in this study are  
134 comprised of distinct phases including the Discovery, Discovery Extension, and Augmentation  
135 phases. The Discovery phase WGS was generated from individuals of multiplex AD families  
136 as previously described.<sup>8-10</sup> The Discovery Extension phase consisted of a family component  
137 and a case control component. The Discovery Extension family component WGS was  
138 generated on additional members of selected families from the Discovery phase as well as

139 members of 77 additional families. A set of 114 Hispanic control individuals was also  
140 sequenced with the family component.

141 A focus of the Discovery Extension case control component was to increase the diversity of  
142 the ADSP samples. The ADSP Discovery Extension WGS was generated on 3,082  
143 individuals, with approximately one third from EA, AA, and HI populations. In the ADSP  
144 Discovery and Discovery Extension phases sequencing was performed at three sequencing  
145 centers via the National Human Genome Research Institute (NHGRI). Sequence data for  
146 ADSP Augmentation Studies were supported by NIA and private funding and are shared with  
147 the research community via NIAGADS. The ADSP data coordinating center, the Genomic  
148 Center for AD (GCAD), produced a jointly called and quality controlled (QC'ed) data set for  
149 WGS10 that included the ADSP WGS Discovery, Discovery Extension, and from the  
150 Augmentation phase, the Alzheimer's Disease Neuroimaging Initiative (ADNI) study. Details  
151 of studies included in the ADSP can be found at NIAGADS under dataset: NG00067 ADSP  
152 Umbrella Study (<https://dss.niagads.org/datasets/ng00067/>).

153

## 154 2.2 WGS Quality Control

155 Low-quality variants were filtered out based on the GCAD provided flags, which were  
156 generated separately for the Family, Case-Control, and ADNI sub-studies.<sup>10</sup> In addition, GCAD  
157 provided the ABHet ratio computed as (the total reference reads over all heterozygous  
158 genotypes)/(total alternative and reference reads over all heterozygous genotypes). A variant  
159 was excluded if it failed the GATK Variant Quality Score Recalibration (VQSR) filter, all  
160 genotypes were missing, was monomorphic, or if it had low call rate across all studies.  
161 Additional filtering was implemented within sub-study. If a variant had high read depth (>500  
162 reads) within a study or had ABHet < 0.25 or ABHet > 0.75 within a sub-study, all the  
163 genotypes within that sub-study were set to missing. After these filters were applied, a final  
164 call rate filter of 95% across all sub-studies was implemented.

165

## 166 2.3 AD Phenotype Definition

167 The ADSP provides different AD status variable definitions for participants included via case-  
168 control versus family-based studies. In the current analysis, for individuals in the ADSP case-  
169 control study, we defined AD cases as individuals with either prevalent or incident AD. Case-  
170 control individuals with no prevalent or incident AD were defined as controls and those with  
171 missing status were defined as unknown. In the ADSP family phenotype file, possible values  
172 for the AD status variable include no dementia, definite AD, probable AD, possible AD, family-  
173 reported AD, other dementia, family reported no dementia, and unknown. For family-based  
174 individuals, we defined an AD case as either possible, probable or definite AD. AD controls  
175 were defined as individuals coded as no dementia. We redefined individuals with family-  
176 reported AD, other dementia, or unknown status as missing AD status. The ADNI phenotype  
177 data, which is part of the ADSP Augmentation study, provides information on mild cognitive  
178 impairment (MCI) in addition to AD status. Individuals with a current diagnosis of MCI (N=320)  
179 were included as AD controls in the current study. After selecting genetically unique individuals  
180 with AD status available, a total of 4,567 participants (2,383 controls and 2,184 cases) with  
181 WGS were included in the analyses.

182

#### 183 2.4 Pooled sample single-variant association analysis

184 Single-variant association analysis of AD was performed on variants within GWAS loci for  
185 participants with both phenotype and genotype data available using GENESIS.<sup>11</sup> Principal  
186 component analysis (PCA) was performed as described in the supplemental methods to  
187 assess and adjust for genetic ancestry of the study participants (**Figure A1**). The WGS  
188 samples included in the ADSP R1 WGS data set were sequenced across four sequencing  
189 centers (Baylor College of Medicine Human Genome Sequencing Center, The Broad Institute,  
190 McDonnell Genome Institute at Washington University School of Medicine, and Illumina) and  
191 2 sequencing platforms (Illumina HiSeq 2000/2500, and Illumina HiSeq X Ten). In order to  
192 control for the effects from study design and technical differences, we generated indicator  
193 variables (study $\times$ sequencing center $\times$ sequencing platform) with 10 categories based on **Table**  
194 **1**. We considered these indicator variables as technical covariates and defined case-control  $\times$



195 Broad x HiSeq X Ten, which had the largest number of observations, as the reference group.  
196 We used a generalized logistic mixed-effects model to account for relatedness through a  
197 genetic relationship matrix (GRM). The GRM was estimated based on the same variants used  
198 in the PCA. We included sex, the technical covariates, and PC2 (based on a Bonferroni  
199 corrected significant  $p < 0.0016$  for testing 32 PCs) as covariates in the null model. We  
200 performed the analysis across autosomes, and kept variants satisfying the criteria: call rate  
201 higher than 95% and minor allele count (MAC) higher than 20.  
202 To determine if significant variants identified provided distinct signals from the lead GWAS  
203 variants,<sup>7</sup> conditional analyses were performed in loci displaying significant associations.  
204 Genotype data, coded as 0/1/2, for the lead GWAS variants in these loci was included in the  
205 null model in addition to the covariates. Association analyses conditioned on the lead GWAS  
206 variants were then rerun for the loci of interest.

207

## 208 2.5 Population specific association analysis

209 We conducted population specific analyses (null model and association analyses) for AD using  
210 GENESIS, accounting for genetic relatedness using a GRM. We defined three population  
211 groups (EA, AA, and HI). We selected a total of 2,144 EA participants based on PCA analysis  
212 performed using both the ADSP and the Human Genome Diversity Project (HGDP). Only  
213 participants who were not outliers based on 6 standard deviations (SD) from the mean for PCs  
214 1-4 calculated in the European HGDP groups (Adygei, Basque, French, BergamoItalian,  
215 Orcadian, Russian, Sardinian, and Tuscan) were retained. We selected a total of 1,028 AA  
216 and 1,548 HI participants based on reported race and ethnicity. A total of 38 participants who  
217 identified as both African-American and Hispanic were placed in the Hispanic population. We  
218 included in the null model, in each population group, covariates associated with AD status at  
219  $P \leq 0.05$ . The EA null model included sex, ADSP family study status, Illumina sequencing  
220 center, HiSeq X Ten platform, PC 2, PC 9, and PC 15. The HI null model included sex, all  
221 sequencing centers, HiSeq X Ten platform, PC 13, PC 16, and PC 17. The AA null model  
222 included sex, Illumina sequencing center, and PC 1. We performed association analyses, in



223 each population group, and retained the results with call rate higher than 95% and minor allele  
224 count (MAC) higher than 20. In addition, we performed a multi-population meta-analysis using  
225 three different models (fixed-effect, random-effect, and Han & Eskin's modified random-effect)  
226 implemented in Metasoft<sup>12</sup> by combining the population specific results satisfying the criteria  
227 of a within population MAC higher than 10. We then kept the meta-analysis results passing a  
228 total MAC across population groups higher than 20.

229

## 230 2.6 Gene-based tests

231 We tested the association of aggregate groups of low frequency (minor allele frequency (MAF)  
232 < 5%) or rare (MAF < 1%) genetic variants with AD status. Annotation for all called variants  
233 was generated using Ensembl VEP91 by the ADSP annotation working group. We selected  
234 missense or loss of function (lof) genetic variants based on the most severe variant  
235 consequence according to the ADSP Annotation WG Ranking Process and listed in the  
236 annotation file (frameshift variant, inframe deletion, inframe insertion, missense variant,  
237 protein-altering variant, splice acceptor variant, splice donor variant, start lost, stop gained,  
238 and stop lost). We conducted Sequence Kernel Association Test (SKAT, mmskat) and burden  
239 tests (combined multivariate and collapsing (CMC), emmaxCMC) with EFACTS (Efficient and  
240 Parallelizable Association Container Toolbox) using mixed-effect models adjusted for sex,  
241 technical covariates, and PCs significantly associated with AD status (PC2) accounting for  
242 genetic relatedness (GRM).

243

## 244 2.7 Non-coding rare variant analysis

245 For non-coding rare variant analysis, we used annotations from WGS v0.8<sup>13</sup> including  
246 annotations from ANNOVAR, VEP, SnpEff, COSMIC and SPIDEX. We conducted rare variant  
247 analysis using the variant-Set Test for Association using Annotation infoRmation (STAAR)  
248 method,<sup>14</sup> which was developed to boost power of rare variant analyses by effectively  
249 incorporating both variant functional categories and multiple complementary functional  
250 annotations while accounting for relatedness and population structures. We used the same

251 covariates (sex, technical covariates, and significant PCs) in the model as in single-variant  
252 analysis. The GRM was incorporated to account for relatedness among samples.  
253 We aggregated sites that overlap enhancers and promoters around gene transcription start  
254 sites (TSS). The promoters within 5KB of a TSS that overlap DNase hypersensitivity sites  
255 (DHS) are defined as at least one WGS H3K4me3 annotation for brain tissues (E067, E068,  
256 E069, E070, E071, E072, E073, E074, E081, E082), and the enhancers within 20KB of a TSS  
257 are defined by EnhancerFinder in Brain. We incorporated annotations from WGS in the  
258 analysis, which include MAF, functional scores (GERP, GenoCanyon, RegulomeDB,  
259 FUNSEQ, CADD, Fathmm, EIGEN-PC), and the ENCODE score (DNASE). We then  
260 transformed the annotation scores to phred-scaled scores using  $-10 \times \log_{10}(\text{rank}(-\text{score})/M)$ ,  
261 where M is the total number of variants tested in the analysis.

262

## 263 2.8 Focus on GWAS loci

264 Given the limited power to detect novel loci with the current sample size, we focused on  
265 exploiting WGS to provide insights on previously reported AD GWAS loci. We used the  
266 variants listed in Supplemental Table 5 in Bellenguez et. al as the previously reported AD  
267 GWAS top variants. For single variant association analyses, we looked up these lead variants  
268 in the ADSP WGS data. We then assessed ADSP WGS associations within 100KB of each  
269 lead GWAS variant. For gene-based and non-coding rare variant analysis, we obtained the  
270 results for genes or regions in the 100KB window around each lead variant. We included  
271 genes or non-coding regions for which any portion overlapped with the specified window.  
272 Using this paradigm, we identified 303 genes within 100kb of the index SNPs.

273 In general, we defined a threshold for statistical significance equal to  $0.05/\text{number of statistical}$   
274  $\text{tests}$  and a suggestive threshold as  $1/\text{number of statistical tests}$ . Within a 100kb window, many  
275 single variant tests were highly correlated. Therefore, we computed the effective number of  
276 independent tests using the simpleM approach<sup>15</sup> and used the effective number of tests in the  
277 denominator when computing a window-specific threshold for single variant association

278 testing. Effective number of tests were computed across the pooled sample and within each  
279 population sub-group (**Table A1**). **Figure 1** provides an overview of our analysis workflow.  
280 We leveraged publicly-available multi-omic resource generated by applying quantitative trait  
281 locus (xQTL) analyses to RNA sequence and DNA methylation from the dorsolateral prefrontal  
282 cortex of 411 older adults from the Religious Orders Study (ROS) and Memory and Aging  
283 Project (MAP) studies<sup>16</sup> to look-up the main genetic variants from the pooled association  
284 analysis.

285

### 286 **3. Results**

#### 287 3.1 Description of ADSP data

288 After the QC of the ADSP data release NG00067.v2, there were over 95 million variants across  
289 4,733 participants. A total of 4,567 individuals (2,383 controls, 2,184 cases) have available  
290 AD status and contributed to the analyses, among which 807 are from the ADSP family study,  
291 2,963 are from the ADSP case-control study, and 797 are from ADNI. The participants  
292 included in the analyses were more likely to be women (61.8%) than men. The distribution of  
293 study design membership, sequencing centers, and sequencing platforms is summarized in  
294 **Table 1**.

295

#### 296 3.2 Pooled sample single-variant association analysis

297 Genome-wide, there were about 20 million variants with call rate higher than 95% and MAC  
298 higher than 20 in the pooled sample analysis. Our model that included GRM and PC  
299 adjustments showed acceptable type-I error control ( $\lambda = 1.05$ , **Figure A2**). As expected, the  
300 strongest association was observed at the *APOE* locus, where the major *APOE* variant  
301 rs429358 ( $p = 7.2 \times 10^{-77}$ ) was the top hit.

302 Among the specific lead GWAS variants from Bellenguez et al,<sup>7</sup> none reached the strict  
303 significance threshold ( $p < 6 \times 10^{-4}$ , Bonferroni correction for the total number of variants  
304 tested) in the pooled sample association analysis. Using the suggestive significance threshold  
305 ( $p < 0.012$ ; 1/83), we found associations for rs7401792 ( $p = 7.3 \times 10^{-4}$ , MAF= 49.2%) in the

306 *SLC24A4* locus, rs75932628 ( $p = 3 \times 10^{-3}$ , MAF= 0.35%) in the *TREM2* locus, rs616338 ( $p =$   
307  $4.4 \times 10^{-3}$ , MAF= 0.71%) in the *ABI3* locus, rs1358782 ( $p = 4.9 \times 10^{-3}$ , MAF= 22.9%) in the  
308 *RBCK1* locus, rs1160871 ( $p = 6.9 \times 10^{-3}$ , MAF= 40.2%) in the *JAZF1* locus, and rs602602 ( $p$   
309  $= 0.12$ , MAF= 27.7%) in the *MINDY2* locus. Full results for the 83 lead GWAS variants are  
310 provided in the supplement (**Table A2**).

311 Applying the significance thresholds based on the effective number of tests within 100 kb  
312 windows around the lead GWAS variants (**Table A1**), we identified 17 significant variants in  
313 the single variant association analysis in the pooled sample (**Table 2**). These 17 variants occur  
314 in five genomic regions on chromosomes 6, 7, 14, and 16. Forest plots for the top variant in  
315 each of these five regions are presented in **Figure 2**. Examination of linkage disequilibrium  
316 (LD) patterns show near perfect LD among the variants identified on chromosome 6, on  
317 chromosome 7 or for the one region on chromosome 14 with multiple variants. Only a single  
318 variant was identified on chromosome 16 and in one region on chromosome 14. Detailed LD  
319 information is provided in the supplement (**Figure A3**).

320 Conditional analyses were performed to determine if these associations represented the same  
321 signal as the lead GWAS variant from Bellenguez et al or a distinct signal. As shown in **Table**  
322 **2**, the inclusion of the lead GWAS variant in the association model did not mitigate the  
323 association indicating the variants identified represent a distinct signal from the lead GWAS  
324 variants.

325

### 326 3.3 Population specific single-variant association analysis and multi-population meta-analysis

327 We conducted population specific association analyses in the three main subgroups (N=2,043  
328 EA; N=995 AA, and N=1,516 HI participants). There was acceptable type-I error in the  
329 population specific analyses and the multi-population meta-analysis (**Figures A4-A7**). We  
330 confirmed the significant association of the *APOE* locus (rs429358) in both the population  
331 specific analyses and the multi-population meta-analysis. However, as found in previous  
332 studies<sup>17,18</sup> the association was weaker in the Hispanic population (beta = 1.17 in EA, 1.02 in  
333 AA, and 0.59 in HI).

334 Population specific single variant analyses identified 23 significant variants in 11 loci within  
335 100 kb of the lead GWAS variants (**Table 3**). Of these, 15 variants (8 loci) were identified in  
336 the EA population and 8 variants (3 loci) in the AA population. No significant variants were  
337 identified in the HI population. The only overlap in significant variants between the pooled  
338 sample and the population specific single-variant analyses was a missense variant in *KAT8*,  
339 which was very rare in the AA and HI subsamples (**Table A3**).

340

### 341 3.4 Gene-based tests

342 QQ plots for the SKAT and burden tests are provided in **Figure A8**. Using the multiple-testing  
343 correction threshold for 303 genes ( $p < 1.7 \times 10^{-4}$ ), *KAT8* ( $p = 2.2 \times 10^{-5}$ , MAF < 5%) lying  
344 within 100 KB of a GWAS variant was detected to be significantly associated with AD status  
345 by SKAT, and it was also shown as a suggestive association ( $p < 3.4 \times 10^{-3}$ ) using CMC ( $p =$   
346  $9.2 \times 10^{-4}$ ). SKAT detected suggestive associations in *LAIR1* ( $p = 0.0023$ , MAF < 5%) and  
347 *ATF5* ( $p = 5.7 \times 10^{-4}$ , MAF < 5%) within 100KB. CMC identified *TREM2* ( $p = 3.3 \times 10^{-3}$  for MAF  
348 < 1% and  $8.9 \times 10^{-4}$  for MAF < 5%) within 100KB of GWAS variants.

349

### 350 3.5 Non-coding rare variant analysis

351 QQ plots showed deflated type-I error ( $\lambda = 0.75$ ), most likely due to small sample size, in the  
352 STAAR rare-variant analysis. (**Figure A9**). No regions were identified as significant using the  
353 STAAR approach. The top STAAR results overlapping GWAS loci are shown in **Table A3**.

354

### 355 3.6 xQTL analysis lookup

356 We did not identify significant mQTL or eQTL associations for the main genetic variants  
357 identified in the pooled analysis. We could not look up some of the less frequent variants on  
358 chr7, 14 and 16 in the QTL results as these analyses were restricted to common variants.  
359 Suggestive associations have been reported between rs10947950 (chr6) and cg25473438  
360 (beta = 0.24,  $P = 1.4 \times 10^{-8}$ ) and between rs7155002 (chr14) and cg12072028 (beta = 0.19,  $P$   
361 =  $5.8 \times 10^{-6}$ ), **Table A4**. The CpG cg12072028 is located in the intron 1 of *RIN3* and modest

362 associations have been described between rs7155002 and *RIN3* expression in the brain (beta  
363 = 0.04, P = 0.01), **Table A4**.

364

## 365 **Discussion**

366 GWAS have been essential in identifying genetic loci associated with AD. However, GWAS  
367 loci typically contain scores of genes and thousands of variants. Additional studies are needed  
368 to pinpoint specific genes or variants as the ones influencing risk for AD. WGS provides  
369 complete genomic sequence and hence enumerates both common and rare variants. WGS  
370 therefore has the potential to provide information beyond common variants that are the  
371 cornerstone of GWAS. In the current study, we have examined WGS from the ADSP R1 data  
372 set focusing on previously implicated regions to better understand important variants within  
373 AD GWAS loci in a diverse study sample. We identified 17 significant variants in five genomic  
374 regions using single variant association analysis in the pooled sample. The majority of these  
375 variants were intronic, although two intergenic and one missense variant were also identified.  
376 Bellenguez et al<sup>7</sup> identified multiple lead GWAS variants on chromosome 6, which yielded  
377 overlapping 100kb windows defined by our approach. We identified seven significant variants  
378 on chromosome 6 that are in nearly complete LD. Six of these variants are located within  
379 intronic regions of *OARD1* and *NFYA* genes, and one variant is very close to *APOBEC2*. One  
380 variant was intergenic with the closest gene being *TREML1*. This region contains *TREM2*, for  
381 which several rare coding variants have been implicated as conferring risk for AD.<sup>19-22</sup> *TREM2*  
382 showed suggestive evidence of association in gene-based analyses indicating multiple  
383 variants in this region are likely to play an important role in AD. The missense variant  
384 rs75932628 was one of the lead GWAS variants from Bellenguez et al. and has been identified  
385 as a functional variant for AD.<sup>23-25</sup> The variant rs75932628 has an MAF=0.0035 in the ADSP  
386 pooled sample and p=0.003 for single variant association with AD. This suggestive association  
387 is driven by the EA population (MAF=0.007) as this variant is less frequently observed in the  
388 AA (MAF=0.001) or HI (MAF=0.0003) populations. Our conditional analysis indicates that the  
389 variants we identified implicating *OARD1/NFYA/TREML1* have a distinct effect from

390 rs75932628. *OARD1* encodes a deacylase with a function to catalyse O-acetyl-ADP-ribose  
391 during multiple cellular processes. A homozygous mutation could lead to cell death and cause  
392 a form of childhood neurodegenerative disorder.<sup>26</sup> *NFYA* encodes a subunit of nuclear  
393 transcription factor Y, which is a ubiquitous transcription factor. The gene is involved in post-  
394 transcriptional regulation with tissue-specific preference, and it is suppressed in the brain of  
395 model mice with Huntington's disease<sup>27</sup> and spinal and bulbar muscular atrophy.<sup>28</sup> *TREML1*  
396 encodes a protein belonging to the family of triggering receptors expressed on myeloid cells-  
397 like (TREM). A deficiency of *TREML1* might result in haemorrhage due to localized  
398 inflammatory lesions.<sup>29</sup>

399 The five significant variants on chromosome 7 are in a strong LD block and all variants are  
400 intronic for *JAZF1*, which encodes a transcriptional repressor. The gene has been linked with  
401 diabetes mellitus and cancer, but also has a role in lipid metabolism supporting the connection  
402 between lipid levels and AD.<sup>30</sup> The *JAZF1* GWAS variant (rs1160871) is a strong eQTL in  
403 microglia and considered a Tier 1 (highly plausible) AD gene.<sup>7</sup> The significance of the variants  
404 at this locus was slightly attenuated by conditioning on the nearby GWAS variant (rs1160871),  
405 indicating that this may be a shared effect with the lead GWAS variant.

406 We identified two regions on chromosome 14, with significant variants intronic to *FERMT2* and  
407 *SLC24A4*. The association of the intronic variant rs7155002 for *SLC24A4* was slightly  
408 attenuated by the conditional analysis indicating that this may be a shared effect with the lead  
409 GWAS variant. Lookup in brain xQTL data shed light on potential biological regulatory  
410 mechanisms in *RIN3* that has also been implicated in AD.<sup>31,32</sup> *FERMT2* encodes plekstrin  
411 homology domain-containing family C member 1 and is known to be involved in APP  
412 metabolism.<sup>33</sup> The under-expression of *FERMT2* was associated with mature APP level  
413 increment in the cell surface.<sup>33</sup> Previous studies reported that *FERMT2* is also involved in  
414 cardiac and skeletal muscle development<sup>34</sup> and cancer progression.<sup>35,36</sup> *SLC24A4* encodes a  
415 member of the potassium-dependent sodium/calcium exchanger protein family and is  
416 associated with neural development.<sup>37</sup> A homozygous mutation in *SLC24A4* may cause  
417 amelogenesis imperfecta<sup>38,39</sup> but the function of *SLC24A4* in AD is not clear yet.



418 A rare missense variant (rs201871085, MAF = 0.0108 in the pooled sample) within *KAT8*  
419 (lysine acetyltransferase 8) on chromosome 16 was significantly associated with AD. *KAT8*  
420 was also significant in the low-frequency variant gene-based analyses. *KAT8* encodes a  
421 member of the MYST histone acetylase protein family that has a characteristic MYST domain  
422 containing an acetyl-CoA-binding site, a chromodomain typical of proteins which bind  
423 histones, and a C2HC-type zinc finger. This gene has been recently identified by two large-  
424 scale GWAS of clinically diagnosed AD and family history of AD<sup>40,41</sup> and by a novel knockoff  
425 method when applied to the ADSP Data.<sup>42</sup> Aberrant expression patterns of *KAT8* might be  
426 associated with AD progression.<sup>43</sup> *KAT8* appears like a promising candidate gene that is  
427 involved in cerebral development<sup>44</sup> and may play a role in neurodegeneration in both AD and  
428 Parkinson's disease.<sup>45,46</sup> We were not able to look-up the variant rs201871085 in the brain  
429 xQTL data. This variant might not have been analyzed due to a low frequency or a low quality  
430 of imputation, thus highlighting the importance of leveraging whole genome sequence data.  
431 The ADSP represents a diverse population sample, although in this early release of ADSP  
432 WGS data the sample size within a specific population is limited ( $N_{EA} = 2,043$ ,  $N_{AA} = 995$ ,  $N_{HI}$   
433  $= 1,516$ ). Population specific analyses provide information about patterns of allele frequency  
434 for AD associated variants among populations. Among the five loci identified as significant in  
435 the pooled single variant association analysis, two regions (chr6 & 16) displayed EA-specific  
436 associations and corresponded to low frequency variants in EA that were rare in other  
437 population groups. One signal (chr7) was driven by a variant common in AA with a low  
438 frequency in HI, and rare in EA. Finally, two regions (chr14) were driven by HI signals with one  
439 variant common in all population groups, and one variant common in HI but rare in EA and  
440 AA. All these results are summarized in **Table A3**.  
441 The signals identified only in AA in the population specific analyses (chr4 and 14)  
442 corresponded to SNPs common in AA that were less common in HI and rare in EA. A few  
443 signals identified only in EA corresponded to variants that were rare in all population groups  
444 (chr8, 17 and 20). Two low frequency signals identified only in EA (chr16 and 17)  
445 corresponded to SNPs that were rare in AA and HI. Finally, three signals identified only in EA

446 (chr5, 14, and 16) corresponded to SNPs that were common in different population groups.

447 All these results are summarized in **Table 3**.

448 A strength of this study is the analysis of WGS data that was jointly called and QC'ed by a  
449 single data coordinating center. The diversity in genetic ancestry of the participants included  
450 is another strength. Despite this diversity, a limitation of the study is the moderate sample size  
451 within each population analysed. To overcome this limitation, main analyses were focused on  
452 the pooled sample. The ADSP is ongoing with larger WGS data sets being publicly released  
453 and planned. Future analyses with larger sample size may yield additional insights, specifically  
454 for population specific effects. The current study demonstrates the importance of leveraging  
455 whole genome sequence data to gain insights into loci identified via GWAS and highlights the  
456 contribution of low frequency variants to AD risk.

457

458

459

460

461

462

463

464

465

466

467

468

469

470

471

472

473

474 **References**

- 475 1. Karlsson IK, Escott-Price V, Gatz M, et al. Measuring heritable contributions to  
476 alzheimer's disease: Polygenic risk score analysis with twins. *Brain Commun.*  
477 2022;4(1):fcab308. doi: 10.1093/braincomms/fcab308.
- 478 2. Beecham GW, Bis JC, Martin ER, et al. The alzheimer's disease sequencing project:  
479 Study design and sample selection. *Neurol Genet.* 2017;3(5):e194. doi:  
480 10.1212/NXG.0000000000000194.
- 481 3. Prokopenko D, Morgan SL, Mullin K, et al. Whole-genome sequencing reveals new  
482 alzheimer's disease-associated rare variants in loci related to synaptic function and neuronal  
483 development. *Alzheimers Dement.* 2021;17(9):1509-1527. doi: 10.1002/alz.12319.
- 484 4. Prokopenko D, Lee S, Hecker J, et al. Region-based analysis of rare genomic variants in  
485 whole-genome sequencing datasets reveal two novel alzheimer's disease-associated genes:  
486 DTNB and DLG2. *Mol Psychiatry.* 2022;27(4):1963-1969. doi: 10.1038/s41380-022-01475-0.
- 487 5. Shigemizu D, Asanomi Y, Akiyama S, Mitsumori R, Niida S, Ozaki K. Whole-genome  
488 sequencing reveals novel ethnicity-specific rare variants associated with alzheimer's  
489 disease. *Mol Psychiatry.* 2022;27(5):2554-2562. doi: 10.1038/s41380-022-01483-0.
- 490 6. Park J, Park I, Youm EM, et al. Novel alzheimer's disease risk variants identified based on  
491 whole-genome sequencing of APOE epsilon4 carriers. *Transl Psychiatry.* 2021;11(1):296-9.  
492 doi: 10.1038/s41398-021-01412-9.
- 493 7. Bellenguez C, Kucukali F, Jansen IE, et al. New insights into the genetic etiology of  
494 alzheimer's disease and related dementias. *Nat Genet.* 2022;54(4):412-436. doi:  
495 10.1038/s41588-022-01024-z.
- 496 8. Tsapanou A, Scarmeas N, Gu Y, et al. Data from a cross-sectional study on  
497 apolipoprotein E (APOE-epsilon4) and snoring/sleep apnea in non-demented older adults.  
498 *Data Brief.* 2015;5:351-353. doi: 10.1016/j.dib.2015.09.014.
- 499 9. Chung J, Zhang X, Allen M, et al. Genome-wide pleiotropy analysis of neuropathological  
500 traits related to alzheimer's disease. *Alzheimers Res Ther.* 2018;10(1):22-z. doi:  
501 10.1186/s13195-018-0349-z.

- 502 10. Naj AC, Lin H, Vardarajan BN, et al. Quality control and integration of genotypes from  
503 two calling pipelines for whole genome sequence data in the alzheimer's disease  
504 sequencing project. *Genomics*. 2019;111(4):808-818. doi: 10.1016/j.ygeno.2018.05.004.
- 505 11. Gogarten SM, Sofer T, Chen H, et al. Genetic association testing using the GENESIS  
506 R/bioconductor package. *Bioinformatics*. 2019;35(24):5346-5348. doi:  
507 10.1093/bioinformatics/btz567.
- 508 12. Han B, Eskin E. Random-effects model aimed at discovering associations in meta-  
509 analysis of genome-wide association studies. *Am J Hum Genet*. 2011;88(5):586-598. doi:  
510 10.1016/j.ajhg.2011.04.014.
- 511 13. Liu X, White S, Peng B, et al. WGS: An annotation pipeline for human genome  
512 sequencing studies. *J Med Genet*. 2016;53(2):111-112. doi: 10.1136/jmedgenet-2015-  
513 103423.
- 514 14. Li X, Li Z, Zhou H, et al. Dynamic incorporation of multiple in silico functional annotations  
515 empowers rare variant association analysis of large whole-genome sequencing studies at  
516 scale. *Nat Genet*. 2020;52(9):969-983. doi: 10.1038/s41588-020-0676-4.
- 517 15. Gao X, Starmer J, Martin ER. A multiple testing correction method for genetic  
518 association studies using correlated single nucleotide polymorphisms. *Genet Epidemiol*.  
519 2008;32(4):361-369. doi: 10.1002/gepi.20310.
- 520 16. Ng B, White CC, Klein H, et al. An xQTL map integrates the genetic architecture of the  
521 human brain's transcriptome and epigenome. *Nat Neurosci*. 2017;20(10):1418-1426. doi:  
522 10.1038/nn.4632.
- 523 17. Blue EE, Horimoto ARVR, Mukherjee S, Wijsman EM, Thornton TA. Local ancestry at  
524 APOE modifies alzheimer's disease risk in caribbean hispanics. *Alzheimers Dement*.  
525 2019;15(12):1524-1532. doi: 10.1016/j.jalz.2019.07.016.
- 526 18. Campos M, Edland SD, Peavy GM. Exploratory study of apolipoprotein E epsilon4  
527 genotype and risk of alzheimer's disease in mexican hispanics. *J Am Geriatr Soc*.  
528 2013;61(6):1038-1040. doi: 10.1111/jgs.12292.

- 529 19. Jonsson T, Stefansson H, Steinberg S, et al. Variant of TREM2 associated with the risk  
530 of alzheimer's disease. *N Engl J Med*. 2013;368(2):107-116. doi: 10.1056/NEJMoa1211103.
- 531 20. Li R, Wang X, He P. The most prevalent rare coding variants of TREM2 conferring risk of  
532 alzheimer's disease: A systematic review and meta-analysis. *Exp Ther Med*. 2021;21(4):347.  
533 doi: 10.3892/etm.2021.9778.
- 534 21. Jin SC, Benitez BA, Karch CM, et al. Coding variants in TREM2 increase risk for  
535 alzheimer's disease. *Hum Mol Genet*. 2014;23(21):5838-5846. doi: 10.1093/hmg/ddu277.
- 536 22. Guerreiro R, Wojtas A, Bras J, et al. TREM2 variants in alzheimer's disease. *N Engl J*  
537 *Med*. 2013;368(2):117-127. doi: 10.1056/NEJMoa1211851.
- 538 23. Song WM, Joshita S, Zhou Y, Ulland TK, Gilfillan S, Colonna M. Humanized TREM2  
539 mice reveal microglia-intrinsic and -extrinsic effects of R47H polymorphism. *J Exp Med*.  
540 2018;215(3):745-760. doi: 10.1084/jem.20171529.
- 541 24. Guerreiro R, Wojtas A, Bras J, et al. TREM2 variants in alzheimer's disease. *N Engl J*  
542 *Med*. 2013;368(2):117-127. doi: 10.1056/NEJMoa1211851.
- 543 25. Xiang X, Piers TM, Wefers B, et al. The Trem2 R47H alzheimer's risk variant impairs  
544 splicing and reduces Trem2 mRNA and protein in mice but not in humans. *Mol*  
545 *Neurodegener*. 2018;13(1):49-6. doi: 10.1186/s13024-018-0280-6.
- 546 26. Sharifi R, Morra R, Appel CD, et al. Deficiency of terminal ADP-ribose protein  
547 glycohydrolase TARG1/C6orf130 in neurodegenerative disease. *EMBO J*. 2013;32(9):1225-  
548 1237. doi: 10.1038/emboj.2013.51.
- 549 27. Yamanaka T, Miyazaki H, Oyama F, et al. Mutant huntingtin reduces HSP70 expression  
550 through the sequestration of NF-Y transcription factor. *EMBO J*. 2008;27(6):827-839. doi:  
551 10.1038/emboj.2008.23.
- 552 28. Katsuno M, Adachi H, Minamiyama M, et al. Disrupted transforming growth factor-beta  
553 signaling in spinal and bulbar muscular atrophy. *J Neurosci*. 2010;30(16):5702-5712. doi:  
554 10.1523/JNEUROSCI.0388-10.2010.

- 555 29. Washington AV, Gibot S, Acevedo I, et al. TREM-like transcript-1 protects against  
556 inflammation-associated hemorrhage by facilitating platelet aggregation in mice and  
557 humans. *J Clin Invest*. 2009;119(6):1489-1501. doi: 10.1172/JCI36175.
- 558 30. Li L, Yang Y, Yang G, et al. The role of JAZF1 on lipid metabolism and related genes in  
559 vitro. *Metabolism*. 2011;60(4):523-530. doi: 10.1016/j.metabol.2010.04.021.
- 560 31. Shen R, Zhao X, He L, et al. Upregulation of RIN3 induces endosomal dysfunction in  
561 alzheimer's disease. *Transl Neurodegener*. 2020;9(1):26-1. doi: 10.1186/s40035-020-00206-  
562 1.
- 563 32. Bhattacharyya R, Teves CAF, Long A, Hofert M, Tanzi RE. The neuronal-specific  
564 isoform of BIN1 regulates beta-secretase cleavage of APP and abeta generation in a RIN3-  
565 dependent manner. *Sci Rep*. 2022;12(1):3486-4. doi: 10.1038/s41598-022-07372-4.
- 566 33. Chapuis J, Flaig A, Grenier-Boley B, et al. Genome-wide, high-content siRNA screening  
567 identifies the alzheimer's genetic risk factor FERMT2 as a major modulator of APP  
568 metabolism. *Acta Neuropathol*. 2017;133(6):955-966. doi: 10.1007/s00401-016-1652-z.
- 569 34. Dowling JJ, Gibbs E, Russell M, et al. Kindlin-2 is an essential component of intercalated  
570 discs and is required for vertebrate cardiac structure and function. *Circ Res*.  
571 2008;102(4):423-431. doi: 10.1161/CIRCRESAHA.107.161489.
- 572 35. Shen Z, Ye Y, Dong L, et al. Kindlin-2: A novel adhesion protein related to tumor  
573 invasion, lymph node metastasis, and patient outcome in gastric cancer. *Am J Surg*.  
574 2012;203(2):222-229. doi: 10.1016/j.amjsurg.2011.06.050.
- 575 36. Sossey-Alaoui K, Pluskota E, Szpak D, Plow EF. The Kindlin2-p53-SerpinB2 signaling  
576 axis is required for cellular senescence in breast cancer. *Cell Death Dis*. 2019;10(8):539-z.  
577 doi: 10.1038/s41419-019-1774-z.
- 578 37. Larsson M, Duffy DL, Zhu G, et al. GWAS findings for human iris patterns: Associations  
579 with variants in genes that influence normal neuronal pattern development. *Am J Hum*  
580 *Genet*. 2011;89(2):334-343. doi: 10.1016/j.ajhg.2011.07.011.

- 581 38. Parry DA, Poulter JA, Logan CV, et al. Identification of mutations in SLC24A4, encoding  
582 a potassium-dependent sodium/calcium exchanger, as a cause of amelogenesis imperfecta.  
583 *Am J Hum Genet.* 2013;92(2):307-312. doi: 10.1016/j.ajhg.2013.01.003.
- 584 39. Khan SA, Khan MA, Muhammad N, et al. A novel nonsense variant in SLC24A4 causing  
585 a rare form of amelogenesis imperfecta in a pakistani family. *BMC Med Genet.*  
586 2020;21(1):97-6. doi: 10.1186/s12881-020-01038-6.
- 587 40. Jansen IE, Savage JE, Watanabe K, et al. Genome-wide meta-analysis identifies new  
588 loci and functional pathways influencing alzheimer's disease risk. *Nat Genet.*  
589 2019;51(3):404-413. doi: 10.1038/s41588-018-0311-9.
- 590 41. Marioni RE, Harris SE, Zhang Q, et al. GWAS on family history of alzheimer's disease.  
591 *Transl Psychiatry.* 2018;8(1):99-6. doi: 10.1038/s41398-018-0150-6.
- 592 42. He Z, Liu L, Wang C, et al. Identification of putative causal loci in whole-genome  
593 sequencing data via knockoff statistics. *Nat Commun.* 2021;12(1):3152-4. doi:  
594 10.1038/s41467-021-22889-4.
- 595 43. Chen F, Chen H, Jia Y, Lu H, Tan Q, Zhou X. miR-149-5p inhibition reduces alzheimer's  
596 disease beta-amyloid generation in 293/APPsw cells by upregulating H4K16ac via KAT8.  
597 *Exp Ther Med.* 2020;20(5):88. doi: 10.3892/etm.2020.9216.
- 598 44. Li L, Ghorbani M, Weisz-Hubshman M, et al. Lysine acetyltransferase 8 is involved in  
599 cerebral development and syndromic intellectual disability. *J Clin Invest.* 2020;130(3):1431-  
600 1445. doi: 10.1172/JCI131145.
- 601 45. Dumitriu A, Golji J, Labadorf AT, et al. Integrative analyses of proteomics and RNA  
602 transcriptomics implicate mitochondrial processes, protein folding pathways and GWAS loci  
603 in parkinson disease. *BMC Med Genomics.* 2016;9:5-y. doi: 10.1186/s12920-016-0164-y.
- 604 46. Wainberg M, Andrews SJ, Tripathy SJ. Shared genetic risk loci between alzheimer's  
605 disease and related dementias, parkinson's disease, and amyotrophic lateral sclerosis.  
606 *Alzheimers Res Ther.* 2023;15(1):113-3. doi: 10.1186/s13195-023-01244-3.
- 607  
608



## 609 **Acknowledgments**

610 We thank the contributors who collected samples used in this study, as well as patients and  
611 their families, whose help and participation made this work possible.

612 ADSP: Data for this study were prepared, archived, and distributed by the National Institute  
613 on Aging Alzheimer's Disease Data Storage Site (NIAGADS) at the University of Pennsylvania  
614 (U24-AG041689), funded by the National Institute on Aging (accession NG00067). The full  
615 acknowledgement statement for the ADSP can be found at:  
616 <https://dss.niagads.org/datasets/ng00067/>

617 ADNI: Data used in preparation of this article were obtained through NIAGADS. The  
618 investigators within the ADNI contributed to the design and implementation of ADNI and/or  
619 provided data but did not participate in analysis or writing of this report. A complete listing of  
620 ADNI investigators can be found at:

621 [http://adni.loni.usc.edu/wp-content/uploads/how\\_to\\_apply/ADNI\\_Acknowledgement\\_List.pdf](http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf)

622

## 623 **Sources of Funding**

624 This work was funded through U01 AG058589 and U01 AG068221 from the National Institute  
625 on Aging.

626 ADSP: All relevant funding is listed in the full acknowledgement statement for the ADSP that  
627 can be found at: <https://dss.niagads.org/datasets/ng00067/>

628 ADNI: Data collection and sharing for ADNI was funded by the Alzheimer's Disease  
629 Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD  
630 ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the  
631 National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering,  
632 and through generous contributions from the following: AbbVie, Alzheimer's Association;  
633 Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-  
634 Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli  
635 Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company  
636 Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy

637 Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research &  
638 Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.;  
639 NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer  
640 Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition  
641 Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI  
642 clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the  
643 National Institutes of Health ([www.fnih.org](http://www.fnih.org)). The grantee organization is the Northern  
644 California Institute for Research and Education, and the study is coordinated by the  
645 Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data  
646 are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

647

#### 648 **Disclosures**

649 The authors do not have declarations of interest to report. The funding sources of this study  
650 had no role in the study design, the collection, the analysis or the interpretation of data, in the  
651 writing of the report, or in the decision to submit the article for publication.

652

653

654

655

656

657

658

659

660

661

662

663

664

665 **Figure captions**

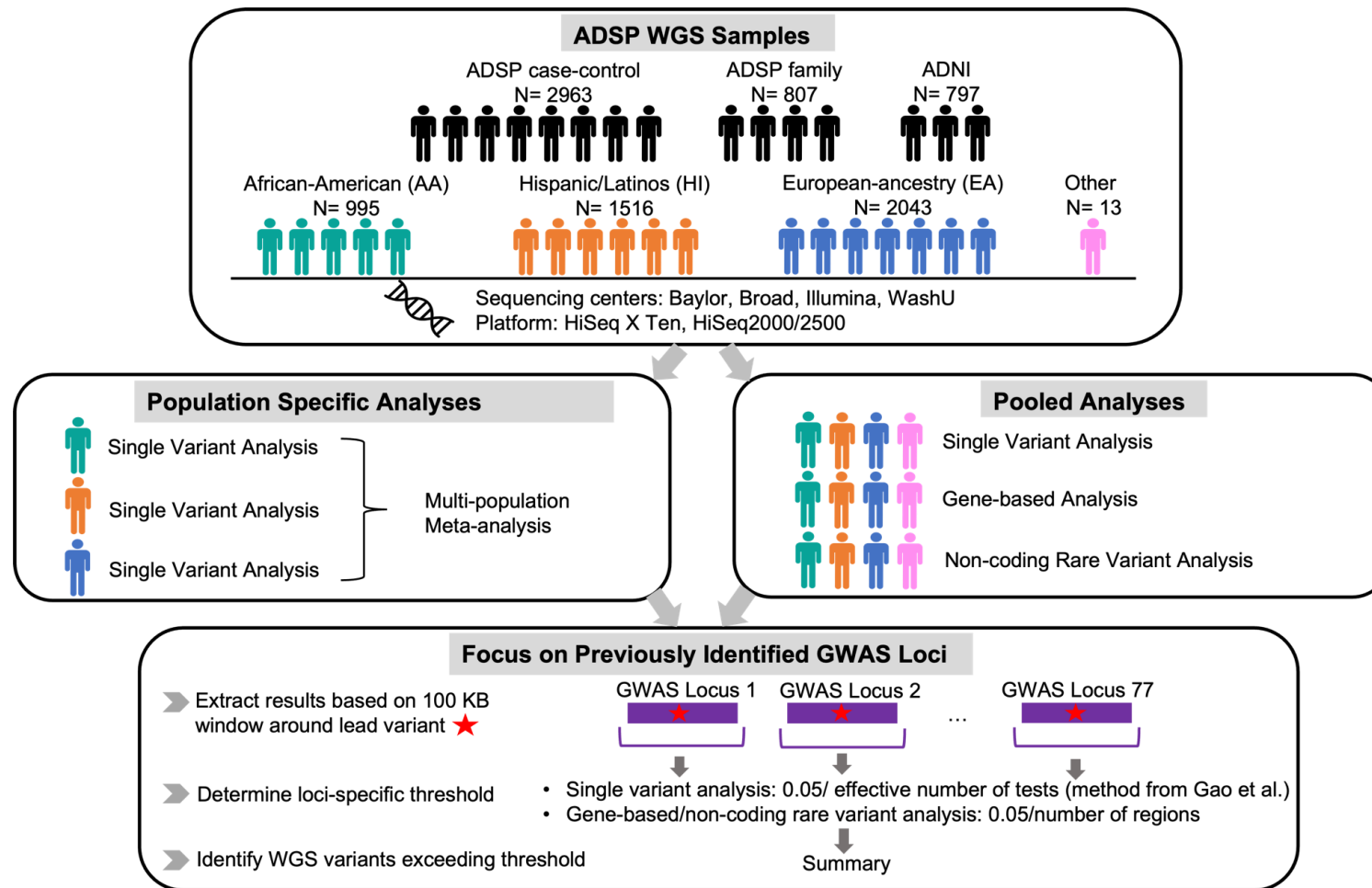
666 **Figure 1.** Schematic of the ADSP 5K analysis

667 **Figure 2.** Top variants identified from single variant association analysis in the pooled sample  
668 within 100kb of the 83 lead GWAS variants

669 Variant ID is in the form of chromosome:position (effect allele). Positions provided are on build  
670 38. EAF is the effect allele frequency, Meta-RE is the multi-population meta-analysis using a  
671 random effect model, Meta-FE is the multi-population meta-analysis using a fixed effect model.  
672 The p-value for Meta-RE is calculated using Han and Eskin's random effects model. The effect  
673 size and its 95% CI is not shown for variants with a minor allele count (MAC) < 10 in population  
674 specific analysis.

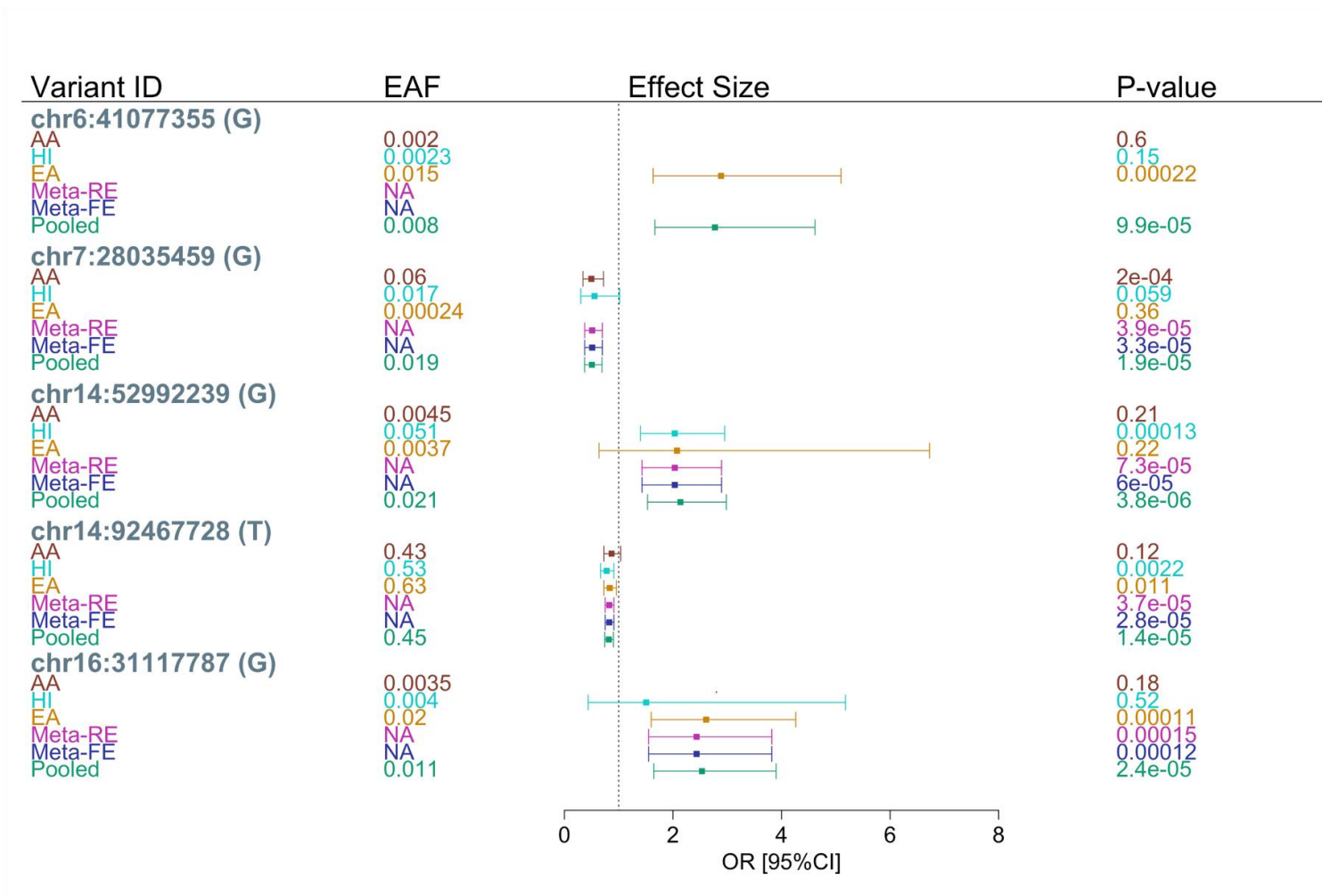
675

676



**Figure 1.** Schematic of ADSP 5K analysis

Color should be used.



**Figure 2.** Top variants identified from single variant association analysis in the pooled sample within 100kb of the 83 lead GWAS variants

Color should be used.

**Table 1.** Characteristics of the participants included in the ADSP R1 data set

		<b>All (N=4,567)</b>	<b>AA* (N=995)</b>	<b>HI* (N=1,516)</b>	<b>EA* (N=2,043)</b>	<b>Other (N=13)</b>
<b>Age (sd)</b>		76.9 (8.3)	79.2 (7.6)	75.1 (8.5)	77.1 (8.4)	77.3 (6.6)
<b>Alzheimer's Disease (%)</b>						
	<b>Case</b>	2184 (47.8%)	463 (46.5%)	795 (52.4%)	921 (45.1%)	5 (46.2%)
	<b>Control</b>	2383 (52.2%)	532 (53.5%)	721(47.6%)	1122 (54.9%)	8 (53.8%)
<b>Sex (%)</b>						
	<b>Female</b>	2822 (61.8%)	710 (71.4%)	1020 (67.3%)	1086 (53.2%)	6 (38.5%)
	<b>Male</b>	1745 (38.2%)	285 (28.6%)	496 (32.7%)	957 (46.8%)	7 (61.5%)
<b>Study (%)</b>						
	<b>ADNI</b>	797 (17.5%)	26 (2.6%)	10 (0.7%)	750 (36.7%)	11 (84.6%)
	<b>ADSP- case control</b>	2963 (64.9%)	944 (94.9%)	1107 (73.0%)	911 (44.6%)	1 (7.7%)
	<b>ADSP- family</b>	807 (17.7%)	25 (2.5%)	399 (26.3%)	382 (18.7%)	1 (7.7%)
<b>Sequencing center (%)</b>						
	<b>Baylor</b>	1241 (27.2%)	0 (0)	1103 (72.8%)	138 (6.8%)	0 (0)
	<b>Broad</b>	1263 (27.7%)	2 (0.2%)	286 (18.9%)	974 (47.7%)	1 (7.7%)
	<b>Illumina</b>	797 (17.5%)	26 (2.6%)	10 (0.7%)	750 (36.7%)	11 (84.6%)

		<b>All (N=4,567)</b>	<b>AA* (N=995)</b>	<b>HI* (N=1,516)</b>	<b>EA* (N=2,043)</b>	<b>Other (N=13)</b>
	<b>WashU</b>	1266 (27.7%)	967 (97.2%)	117 (7.7%)	181 (8.9%)	1 (7.7%)
<b>Platform (%)</b>						
	<b>HiSeq X Ten</b>	3227 (70.7%)	965 (97.0%)	1186 (78.2%)	1074 (52.6%)	2 (15.4%)
	<b>HiSeq2000/2500</b>	1340 (29.3%)	30 (3.0%)	330 (21.8%)	969 (47.4%)	11 (84.6%)

\*AA: Black/African-American, HI: Hispanic/Latino; EA: White/European ancestry. Populations defined are described in the Methods.



**Table 2.** Significant variants from single variant association analysis in the pooled ADSP sample within 100kb of the 83 lead GWAS variants

Chr:Pos:A2:A1*	rsid	Gene	Location	GWAS variants†	GWAS loci†	Pooled Single Variant Association Analysis			Conditional Analysis	
						MAF	Pvalue	Beta	Pvalue	Beta
6:41067923:C:T	rs115774857	OARD1 (close gene APOBEC2)	intronic	rs143332484, rs75932628, rs10947943	TREM2, UNC5CL	0.0080	1.0E-04	1.016	6.6E-05	1.047
6:41077355:A:G	rs145520578	OARD1, NFYA	intronic			0.0080	9.9E-05	1.018	6.4E-05	1.049
6:41077511:A:C	rs115202236		intronic			0.0081	9.9E-05	1.018	6.4E-05	1.049
6:41082030:G:A	rs12200736		intronic	rs143332484, rs75932628, rs60755019, rs10947943	TREM2, TREML2, UNC5CL	0.0080	1.0E-04	1.018	6.5E-05	1.048
6:41083056:C:T	rs10947945		intronic			0.0080	1.0E-04	1.018	6.4E-05	1.049
6:41088533:T:C	rs12210716		intronic			0.0080	1.0E-04	1.018	6.4E-05	1.049
6:41147490:T:G	rs12199328		intergenic (close gene TREML1)	intergenic	rs143332484, rs75932628, rs60755019	TREM2, TREML2	0.0080	1.0E-04	1.018	6.5E-05
6:41173956:G:A‡	rs10947950‡	intergenic	0.0128				1.4E-04	0.783	9.1E-05	0.806
7:28034934:T:A	rs73683942	JAZF1	intronic	rs1160871	JAZF1	0.0188	7.5E-05	-0.627	3.3E-03	-0.624

Chr:Pos:A2:A1*	rsid	Gene	Location	GWAS variants†	GWAS loci†	Pooled Single Variant Association Analysis			Conditional Analysis	
						MAF	Pvalue	Beta	Pvalue	Beta
7:28034935:C:G	rs78789160		intronic			0.0188	7.5E-05	-0.627	3.3E-03	-0.624
7:28035459:GAGAT:G	no rsids		intronic			0.0188	1.9E-05	-0.677	3.9E-03	-0.674
7:28037452:A:C	rs73683943		intronic			0.0191	3.8E-05	-0.644	3.6E-03	-0.641
7:28042506:C:G	rs60825597		intronic			0.0196	7.8E-05	-0.607	3.3E-03	-0.6
14:52885670:T:TA	rs1310103853	FERMT2	intronic	rs17125924	FERMT2	0.0174	4.5E-06	0.806	3.8E-06	0.817
14:52932032:A:G	rs60609189		intronic			0.0189	7.4E-06	0.768	8.4E-06	0.767
14:52992239:A:G	rs12431954	intergenic				0.0205	3.8E-06	0.763	5.6E-06	0.752
14:92467728:C:T	rs7155002	SLC24A4	intronic	rs12590654, rs7401792	SLC24A4	0.4460	1.4E-05	-0.198	4.6E-03	-0.26
16:31117787:C:G	rs201871085	KAT8 (close gene BCKDK)	missense variant, non coding transcript exon variant	rs889555	BCKDK	0.0108	2.4E-05	0.926	2.9E-05	0.916

\*A1 corresponds to the alternate (effect) allele; positions provided are on build 38

† The GWAS variants and GWAS loci are based on the GWAS list from Bellenguez et al (PMID: 35379992)

‡ Variant identified as significant in the multi-population meta-analysis (not in the pooled analysis)

MAF: Minor Allele Frequency

**Table 3.** Significant variants from single variant association analysis in the population subsamples within 100kb of the 83 lead GWAS variants

Chr:Pos:A2:A1*	White / European Ancestry					Black / African American					Hispanic / Latino				
	N	Freq	MAC	Pvalue	Beta	N	Freq	MAC	Pvalue	Beta	N	Freq	MAC	Pvalue	Beta
4:11052797:T:C	2040	0.0002	1	3.04E-01	2.069	984	0.137	270	8.37E-05	-0.534	1513	0.059	180	4.94E-01	0.113
4:11053332:C:T	2043	0.0002	1	3.04E-01	2.068	995	0.137	273	8.66E-05	-0.530	1516	0.059	179	4.83E-01	0.116
5:180211637:C:T	2042	0.143	584	7.14E-05	0.385	995	0.052	104	8.00E-02	-0.351	1514	0.110	333	9.48E-01	0.009
5:180214978:G:A	2042	0.143	586	5.37E-05	0.392	994	0.051	101	1.45E-01	-0.296	1514	0.109	330	9.39E-01	0.010
5:180216117:C:T	2041	0.143	582	3.69E-05	0.400	995	0.050	100	1.72E-01	-0.278	1512	0.096	289	6.05E-01	0.073
5:180222862:G:A	2042	0.144	590	3.70E-05	0.398	995	0.079	157	3.94E-01	-0.144	1516	0.124	377	9.75E-01	0.004
5:180224704:C:T	2041	0.145	591	2.85E-05	0.403	989	0.081	160	3.81E-01	-0.147	1511	0.124	375	9.48E-01	-0.008
8:11868930:C:G	2043	0.007	27	6.52E-05	1.627	995	0.003	6	5.52E-01	-0.494	1516	0.009	26	3.26E-01	-0.414
11:86068255:T:G	2042	0.810	774	6.19E-01	0.044	992	0.886	226	5.53E-05	0.576	1516	0.811	572	8.58E-01	-0.019
11:86068268:A:G	2042	0.810	774	6.19E-01	0.044	990	0.887	224	2.49E-05	0.607	1515	0.811	572	8.45E-01	-0.021
11:86072833:A:G	2042	0.815	755	6.67E-01	0.038	992	0.888	222	5.37E-05	0.577	1516	0.816	559	7.35E-01	-0.037
11:86186529:G:A	2038	0.142	577	8.10E-02	0.169	993	0.264	525	1.21E-04	-0.403	1516	0.245	742	1.52E-01	0.137
14:105740487:C:T	1863	0.524	1772	2.12E-05	-0.479	981	0.308	605	5.33E-01	0.074	1446	0.430	1244	7.48E-02	0.204
15:63375962:G:A	2042	0.0002	1	2.53E-01	2.318	994	0.101	200	1.18E-04	0.591	1514	0.035	105	9.71E-01	0.008

Chr:Pos:A2:A1*	White / European Ancestry					Black / African American					Hispanic / Latino				
	N	Freq	MAC	Pvalue	Beta	N	Freq	MAC	Pvalue	Beta	N	Freq	MAC	Pvalue	Beta
15:63376200:A:G	2042	0.0002	1	2.53E-01	2.317	994	0.106	210	1.24E-04	0.576	1515	0.040	122	7.96E-01	-0.052
16:31117787:C:G	1917	0.020	76	1.11E-04	0.961	991	0.004	7	1.77E-01	1.033	1489	0.004	12	5.16E-01	0.412
16:86357432:T:C	2043	0.088	359	3.53E-05	-0.497	995	0.132	262	7.55E-01	-0.043	1516	0.097	293	4.62E-01	0.100
17:46724128:T:C	2043	0.007	30	3.58E-05	1.607	995	0.002	3	5.80E-01	-0.643	1515	0.006	17	3.75E-01	0.514
17:46747538:C:T	2043	0.007	28	3.69E-05	1.649	994	0.002	3	5.79E-01	-0.645	1515	0.006	17	3.76E-01	0.513
17:58269710:G:A	2042	0.016	67	7.96E-05	1.069	995	0.004	8	7.23E-01	0.256	1515	0.005	14	7.84E-01	-0.169
20:56407698:G:A	2037	0.004	16	3.43E-05	2.328	991	0.001	2	8.15E-01	0.339	1516	0.002	6	8.86E-01	-0.172
20:56490678:G:T	2041	0.005	19	4.74E-05	2.021	995	0.002	4	4.93E-01	-0.703	1516	0.002	6	8.86E-01	-0.172
20:56505267:G:A	2041	0.005	20	1.57E-05	2.095	995	0.002	4	4.93E-01	-0.703	1513	0.002	6	8.83E-01	-0.177

\*A1 corresponds to the alternate (effect) allele; positions provided are on build 38

Variants with a minor allele count (MAC) < 10 in a population subsample were not included in the meta-analysis