



## Original Investigation | Neurology

# Association of Uncommon, Noncoding Variants in the *APOE* Region With Risk of Alzheimer Disease in Adults of European Ancestry

Elizabeth E. Blue, PhD; Anqi Cheng, PhD; Sunny Chen, MS; Chang-En Yu, PhD; for the Alzheimer's Disease Genetics Consortium

## Abstract

**IMPORTANCE** The  $\epsilon 2$  and  $\epsilon 4$  alleles of the apolipoprotein E (*APOE*) gene are associated with Alzheimer disease (AD) risk. Although nearby genetic variants have also been shown to be associated with AD, including rs2075650 in the *TOMM40* gene and rs4420638 near the *APOC1* gene, it is unknown whether these associations are independent of the  $\epsilon 2$  and  $\epsilon 4$  alleles.

**OBJECTIVE** To assess whether variants near *APOE* are associated with AD independently of the  $\epsilon 2/\epsilon 3/\epsilon 4$  genotype.

**DESIGN, SETTING, AND PARTICIPANTS** In this genetic association study of the Alzheimer's Disease Genetics Consortium imputed genotype at data, 14 415 variants near *APOE* ( $\pm 500$  kilobase) for 18 795 individuals with European ancestry were tested for association with AD using 4 logistic mixed models adjusting for sex, cohort, population structure, and relatedness. Model 1 had no *APOE* adjustment, and model 2 adjusted for the count of  $\epsilon 2$  and  $\epsilon 4$  alleles. Model 3 was restricted to  $\epsilon 3$  homozygotes, and model 4 was restricted to  $\epsilon 4$  homozygotes. Data were downloaded from May 31, 2018, to June 3, 2018, and analyzed from November 1, 2018, to June 24, 2020.

**MAIN OUTCOMES AND MEASURES** Alzheimer disease affectation status was defined by clinicians using standard National Institute of Neurological and Communicative Disorders and Stroke and Alzheimer Disease and Related Disorders Association criteria. Association was evaluated using Score tests; results with  $P < .05$  divided by the number of independent tests per model were considered statistically significant.

**RESULTS** Among the 18 795 individuals in the study, 9704 were affected by AD and 9066 were control individuals; the median age at onset/evaluation was 76 (interquartile range, 70-82) years; and 11 167 were female (59.4%). Associations with AD were found for rs2075650 (odds ratio [OR], 2.59; 95% CI, 2.45-2.75;  $P = 3.19 \times 10^{-228}$ ) and rs4420638 (OR, 2.77; 95% CI, 2.62-2.94;  $P = 2.99 \times 10^{-254}$ ) without *APOE* adjustment. Although rs2075650 was nominally associated with AD among the  $\epsilon 4$  homozygotes (OR, 1.33; 95% CI, 1.00-1.77;  $P = .047$ ), the association between rs4420638 and AD was eliminated by *APOE* adjustment (model 2 OR, 1.06 [95% CI, 0.96-1.18;  $P = .24$ ]; model 3 OR, 1.13 [95% CI, 0.95-1.34;  $P = .18$ ]; model 4 OR, 0.90 [95% CI, 0.56-1.45;  $P = .66$ ]). There was a significant association between rs192879175 and AD among  $\epsilon 3$  homozygotes (OR, 0.50; 95% CI, 0.37-0.68;  $P = 8.30 \times 10^{-6}$ ).

**CONCLUSIONS AND RELEVANCE** The results of this genetic association study suggest that  $\epsilon 2/\epsilon 3/\epsilon 4$  alleles are not the only variants in the *APOE* region that are associated with AD risk. Additional work with independent data is needed to replicate these results.

JAMA Network Open. 2020;3(10):e2017666. doi:10.1001/jamanetworkopen.2020.17666

## Key Points

**Question** Is genetic variation near the apolipoprotein E gene (*APOE*) associated with risk of Alzheimer disease (AD) independently of the  $\epsilon 2/\epsilon 3/\epsilon 4$  genotype?

**Findings** In this genetic association study of 18 795 participants of European ancestry from the Alzheimer's Disease Genetics Consortium, an association was found between rs2075650 and AD risk among  $\epsilon 4$  homozygotes and a significant association was found between rs192879175 and AD risk among  $\epsilon 3$  homozygotes.

**Meaning** These findings suggest that even among individuals with the same  $\epsilon 2/\epsilon 3/\epsilon 4$  genotype, genetic variation within the *APOE* neighboring region may be associated with risk of AD.

## + Supplemental content

Author affiliations and article information are listed at the end of this article.

**Open Access.** This is an open access article distributed under the terms of the CC-BY License.

## Introduction

The association between the apolipoprotein E (*APOE* [OMIM 107741]) gene and Alzheimer disease (AD) has been known for longer than 25 years<sup>1,2</sup> and has remained the strongest and most consistent association between AD risk and a common DNA variant.<sup>3,4</sup> Dozens of genetic loci are associated with risk of AD, and hundreds of variants across 3 genes (*APP* [OMIM 605714], *PSEN1* [OMIM 104311], and *PSEN2* [OMIM 600759]) are known to cause early-onset, autosomal dominant forms of AD.<sup>4-6</sup> This genetic heterogeneity has also been observed at the *APOE* locus. Two independent missense variants in *APOE*, rs429358 and rs7412, are consistently associated with large effects on AD risk, and together define the  $\epsilon 2/\epsilon 3/\epsilon 4$  alleles. Associations between many other single-nucleotide variants (SNVs) at the *APOE* locus with AD risk, age at onset, and/or biomarkers have been reported.<sup>7,8</sup>

Whether the association between SNVs in the *APOE* region and AD is independent of the effects of rs429358 and rs7412 is not settled. Many of these SNVs are in linkage disequilibrium (LD) with rs429358 in European ancestry samples, and most are noncoding changes that could affect gene expression.<sup>7,8</sup> The *APOE* locus includes a long cluster of genes transcribed in the same direction, suggesting that they may be coregulated by *cis* regulatory elements. These genes have also been implicated in shared biological pathways, including lipid metabolism, the immune system, and mitochondrial function,<sup>5,7</sup> which suggests that changes in either quality or quantity of the products of these genes may also be associated with AD.

Two noncoding SNVs at the *APOE* locus have consistently shown an association with AD risk and related traits: rs2075650 (the *TOMM40* SNV [OMIM 608061]) and rs4420638 (the *APOC1* SNV [OMIM 107710]). The association between these SNVs and AD is not always robust to *APOE* adjustment.<sup>9,10</sup> Both SNVs are also associated with memory and cognitive function, cerebral spinal fluid biomarkers for immune response,<sup>11</sup> oxidative stress markers,<sup>9</sup> and longevity.<sup>12-14</sup> However, because both SNVs are in moderate LD ( $0.2 < r^2 < 0.8$ ) with rs429358, these associations may not be independent of the  $\epsilon 4$  allele.

We investigated whether rs2075650, rs4420638, or other SNVs in the extended *APOE* locus are associated with risk of AD independently of  $\epsilon 2/\epsilon 3/\epsilon 4$  genotype in a large cohort with European ancestry. We hypothesized that the analytical strategy to adjust for *APOE* effects may influence these association signals.

## Methods

### Samples and Genotype Data

This genetic association study used Alzheimer's Disease Genetics Consortium (ADGC) data, which were accessed through an application on the ADGC website.<sup>15</sup> All participants reported European ancestry. This study was approved by the University of Washington institutional review board and followed the Strengthening the Reporting of Genetic Association Studies (STREGA) reporting guideline. This study evaluated publicly available deidentified data provided by the ADGC. Informed consent was obtained for all research participants as previously described.<sup>16</sup>

The ADGC imputed genotype data were previously generated using the segmented haplotype estimation and imputation tool (SHAPEIT)<sup>17</sup> and IMPUTE, version 2,<sup>18</sup> or MaCH<sup>19</sup> and Minimac<sup>20</sup> software and the 1000 Genomes Project (1KGP) sequence data as reference (phase 3; hg19/GRCh37),<sup>5,21</sup> in which imputed variants with minor allele frequencies (MAFs) of at least 0.01 and either an  $r^2$  or an information measure of less than 0.40 were removed. After excluding 2 data sets owing to incomplete data files, we extracted the SNVs on a bead chip array (Infinium OmniExpress; Illumina) to create a genome-wide association study (GWAS) panel used to estimate principal components, relatedness, and genomic inflation ( $\lambda$  statistic).<sup>22</sup> Single-nucleotide variants with an MAF of less than 0.05, variant-level missing rate of greater than 0.05, or ambiguous alleles were excluded from analysis, as were samples with individual-level missing rate of greater than 0.05; 510 665 variants in 18 795 participants remained. We extracted the 14 415 imputed SNVs within the *APOE* gene ( $\pm 500$  kilobase [kb]) (chromosome 19: 44 909 039-45 912 650) for association testing.

Case individuals were defined as those affected by AD as determined by clinicians using the National Institute of Neurological and Communicative Disorders and Stroke and Alzheimer Disease and Related Disorders Association criteria,<sup>23,24</sup> and control individuals were those not affected by AD. *APOE* genotypes ( $\epsilon 2/\epsilon 3/\epsilon 4$ ) were extracted from the cohort-specific covariate files. *APOE* was genotyped differently across ADGC cohorts.<sup>16</sup>

## Statistical Analysis

Data were downloaded from May 31, 2018, to June 3, 2018, and analyzed from November 1, 2018, to June 24, 2020. The GENESIS package was used to test for the association between SNVs and AD risk,<sup>25,26</sup> an approach that accounts for both population and pedigree structure (eMethods in the Supplement). PC-AiR<sup>27</sup> performed a principal components analysis on the GWAS panel to detect population structure, accounting for kinship estimates provided using the KING approach for robust inference.<sup>28</sup> PC-Relate<sup>29</sup> then used these principal components to estimate a genetic relatedness matrix that is adjusted for population structure. Plots of the first 2 principal components were used to identify outliers among those with self-reported European ancestry. We fit 4 logistic mixed models adjusted for sex, cohort, the first 10 principal components, and a polygenic random effect with covariance structure given by the genetic relatedness matrix. Model 1 included all samples with no *APOE* adjustment, model 2 included all samples and adjusted for  $\epsilon 2$  and  $\epsilon 4$  allele counts, model 3 was restricted to  $\epsilon 3$  homozygotes, and model 4 was restricted to  $\epsilon 4$  homozygotes. Score tests were performed for each logistic model for all SNVs with an MAF of greater than 0.01, with missing genotype data imputed using observed allele frequencies within the data. We estimated the odds ratio (OR) and its 95% CI as follows:  $OR = \exp \times (\text{score statistic}/\text{standard error}^2)$  and 95% CI =  $\pm 1.96 \times (1/\text{standard error})$ . For each model  $m$ , the number of independent tests  $t_m$  was estimated using the genetic type 1 error calculator.<sup>30</sup> Statistical significance was defined as  $P < .05/t_m$ . Linkage disequilibrium between pairs of SNVs was measured using PLINK, version 1.07.<sup>31</sup> Correlations between  $\epsilon 2$  and  $\epsilon 4$  genotypes and imputed genotypes at rs7412 and rs429358 were estimated using R, version 3.5.2.<sup>32</sup> The mismatch between observed and expected  $\epsilon 2$  and  $\epsilon 4$  genotypes was calculated as the number of alleles differing between the observed and imputed genotypes divided by the number of alleles observed. Basic variant annotations, including LD in the 1KGP subset with European ancestry, were performed using HaploReg, version 4.1.<sup>33</sup> Ancestry-matched reference allele frequencies (European MAF) were extracted for non-Finnish Europeans in the gnomAD database, version 2.1.<sup>34</sup>

## Results

### Summary Statistics

The data within the *APOE* region includes 14 415 SNVs and 18 795 individuals, of whom 11 167 were women (59.4%) and 7628 were men (40.6%) (median age at onset/evaluation, 76 [interquartile range, 70–82] years); 9704 were affected by AD (51.6%), and 9066 were controls (51.6%) (eTable 1 in the Supplement). Among cases, the  $\epsilon 2$  allele frequency was 680 of 19 408 (3.5%), and the  $\epsilon 4$  allele frequency was 7360 of 19 408 (37.9%); among controls, the  $\epsilon 2$  frequency was 1444 of 18 132 (8.0%) and the  $\epsilon 4$  frequency was 2490 of 18 132 (13.7%). We observed 71  $\epsilon 2$  homozygotes, 8848  $\epsilon 3$  homozygotes, and 1503  $\epsilon 4$  homozygotes. No outliers were identified by principal components analysis (eFigure 1 in the Supplement), and relatedness estimates were robust to the inclusion of genotypes from chromosome 19 (eFigure 2 in the Supplement). The number of independent tests within the *APOE* region was similar across analysis models ( $t_1$  and  $t_2$ , 1128;  $t_3$ , 1055; and  $t_4$ , 1013), with similar significance thresholds ( $t_1$  and  $t_2$ ,  $P = 4.43 \times 10^{-5}$ ;  $t_3$ ,  $P = 4.74 \times 10^{-5}$ ; and  $t_4$ ,  $P = 4.94 \times 10^{-5}$ ).

### Associations of rs2075650 and rs4420638 With $\epsilon 2$ , $\epsilon 4$ , and AD Risk

There was a stronger LD among rs2075650 (*TOMM40*), rs4420638 (*APOC1*), and rs429358 ( $\epsilon 4$ ) in the ADGC data than in 1KGP Europeans, and none of these SNVs were in LD with rs7412 ( $\epsilon 2$ ) (eTable 2

in the Supplement). Among the 1KGP Europeans, both rs2075650 ( $r^2 = 0.48$ ) and rs4420638 ( $r^2 = 0.65$ ) had moderate LD with rs429358 and modest LD with each other ( $r^2 = 0.30$ ). These correlations were strengthened in the ADGC data, in which  $r^2$  ranged from 0.50 to 0.83 among these 3 SNVs.

The association between AD status and the *TOMM40* and *APOC1* SNVs varied across models (Table 1), each showing no evidence for genomic inflation ( $\lambda_1 = 1.03$ ;  $\lambda_2 = 1.03$ ;  $\lambda_3 = 1.01$ ; and  $\lambda_4 = 0.99$ ) (eFigure 3 in the Supplement).

Each SNV was significantly associated with AD without *APOE* adjustment (model 1) (OR for rs2075650, 2.59 [95% CI, 2.45-2.75;  $P = 3.19 \times 10^{-228}$ ]; OR for rs4420638, 2.77 [95% CI, 2.62-2.94;  $P = 2.99 \times 10^{-254}$ ]), although these associations weakened with *APOE* adjustment or stratification. rs4420638 was not associated with AD with *APOE* adjustment (model 2: OR, 1.06; 95% CI, 0.96-1.18;  $P = .24$ ), among  $\epsilon 3$  homozygotes (model 3: OR, 1.13; 95% CI, 0.95-1.34;  $P = .18$ ), or among  $\epsilon 4$  homozygotes (model 4: OR, 0.90; 95% CI, 0.56-1.45;  $P = .66$ ). The association between rs2075650 and AD was nominally significant among  $\epsilon 4$  homozygotes (model 4) (OR, 1.33; 95% CI, 1.00-1.77;  $P = .047$ ) but failed to reach significance after *APOE* adjustment (model 2; OR, 1.09; 95% CI, 0.99-1.19;  $P = .07$ ) or among  $\epsilon 3$  homozygotes (model 3; OR, 1.16; 95% CI, 0.98-1.38;  $P = .09$ ).

Another *TOMM40* variant (rs10524523, also known as poly-T 523) has been reported to be associated with AD risk<sup>35</sup> but was not available in ADGC data. Using a proxy SNV, rs8106922, which best defines the phylogenetic clade separating long vs short poly-T alleles,<sup>36</sup> we found that rs2075650, rs4420638, rs429358, and rs7412 were not in LD with rs8106922 in ADGC data or 1KGP Europeans ( $r^2 < 0.20$ ). Although the minor allele at rs8106922 was significantly associated with reduced risk of AD under model 1 (OR, 0.69; 95% CI, 0.65-0.72;  $P < .001$ ), the association was not significant under any model adjusting for or stratifying by *APOE* genotype (eTable 3 in the Supplement).

### Imputed vs Measured *APOE* Genotyping

We observed discordance between the observed  $\epsilon 2$  and  $\epsilon 4$  genotypes and the imputed genotypes at the SNVs used to define them. Both rs429358 ( $\epsilon 4$ ) and rs7412 ( $\epsilon 2$ ) were polymorphic in the imputed data in which they should not have been observed, that is, among  $\epsilon 3$  homozygotes (173 of 17 276 and 79 of 17 052 alleles, respectively) and  $\epsilon 4$  homozygotes (2314 of 2492 and 41 of 2664 alleles, respectively). Within the ADGC data, the  $\epsilon 2$  and rs7412 genotypes were correlated ( $r^2 = 0.77$ ;  $P < .001$ ), with a 1.3% mismatch between observed and imputed genotypes, and both the correlation ( $r^2 = 0.88$ ;  $P < .001$ ) and mismatch (2.3%) between the  $\epsilon 4$  and rs429358 genotypes were higher. Both the correlation between observed and imputed genotypes and the mismatch between them varied by *APOE* genotyping strategies (eTable 4 in the Supplement). The SNV-based genotyping had the highest correlation with imputed  $\epsilon 2$  ( $r^2 = 0.81$ ) and  $\epsilon 4$  ( $r^2 = 0.90$ ), and high-throughput sequencing had the lowest ( $r^2 = 0.47$  and  $r^2 = 0.78$ , respectively). This discordance between

Table 1. Association Between the *TOMM40*, *APOC1*, and *APOE* SNVs and AD With and Without *APOE* Adjustment or Stratification

Model <sup>a</sup>	SNV	Nearest gene	No. of Participants	AAC	AAF	OR (95% CI)	P value
1	rs2075650	<i>TOMM40</i>	18 211	8108	0.2226	2.59 (2.45-2.75)	$3.19 \times 10^{-228b}$
2	rs2075650	<i>TOMM40</i>	18 211	8108	0.2226	1.09 (0.99-1.19)	.07
3	rs2075650	<i>TOMM40</i>	8642	746	0.0432	1.16 (0.98-1.38)	.09
4	rs2075650	<i>TOMM40</i>	1426	2106	0.7400	1.33 (1.00-1.77)	.047 <sup>c</sup>
1	rs4420638	<i>APOC1</i>	15 894	7967	0.2506	2.77 (2.62-2.94)	$2.99 \times 10^{-254b}$
2	rs4420638	<i>APOC1</i>	15 894	7967	0.2506	1.06 (0.96-1.18)	.24
3	rs4420638	<i>APOC1</i>	7821	674	0.0431	1.13 (0.95-1.34)	.18
4	rs4420638	<i>APOC1</i>	1058	1893	0.8900	0.90 (0.56-1.45)	.66

Abbreviations: AAC, alternate allele count; AAF, alternate allele frequency; AD, Alzheimer disease; *APOE*, apolipoprotein E; OR, odds ratio; SNV, single-nucleotide variant.

<sup>b</sup> Indicates passing the model-specific significance threshold.

<sup>c</sup> Indicates nominally significant.

<sup>a</sup> Model 1 included all samples, no *APOE* adjustment; model 2, all samples, adjusted for *APOE*  $\epsilon 2$  and  $\epsilon 4$  allele counts; model 3, restricted to  $\epsilon 3$  homozygotes; and model 4, restricted to  $\epsilon 4$  homozygotes.

observed and imputed  $\epsilon 2$  and  $\epsilon 4$  genotypes may have led to spurious associations with AD; there was a nominal association between imputed genotypes at rs429358 and AD after *APOE* adjustment (model 2 OR, 1.16; 95% CI, 1.00-1.34;  $P = .04$ ) and among  $\epsilon 3$  homozygotes (model 3 OR, 1.73; 95% CI, 1.26-2.38;  $P = 6.32 \times 10^{-4}$ ). Imputation accuracy varies based on both the observed marker panel and the reference data set; older arrays performed worse with the 1KGP reference panel used by the ADGC (rs7412,  $r^2 = 0.75$ ; rs429358,  $r^2 = 0.82$ ) than newer arrays (rs7412,  $r^2 = 0.95$ ; rs429358,  $r^2 = 0.95$ ), and both performed better when using the Haplotype Reference Consortium reference panel ( $r^2 > 0.98$ ).<sup>37</sup>

### Additional Associations of *APOE* Region SNVs and AD Risk

Among the 14 415 SNVs in the *APOE* region, we identified 1 significant association across models after correcting for the effective number of independent tests. The **Figure** provides Manhattan plots of the associations with AD across models 2 to 4, and **Table 2** summarizes the 12 strongest associations with AD across these 3 models. One SNV (rs192879175) was significantly associated with AD among  $\epsilon 3$  homozygotes (model 3 OR, 0.50; 95% CI, 0.37-0.68;  $P = 8.30 \times 10^{-6}$ ). No other SNVs were significantly associated with AD after multiple testing correction. None of these 12 SNVs were common in the ADGC data set (MAF > 0.10) or in LD with either rs429358 or rs7412 (maximum  $r^2 = 0.006$ ), and the *BCAM* missense variant rs117737673 represented the only coding change.

### Evidence for Replication

Limited evidence for replication of the significant associations presented in Tables 1 and 2 was available and was derived from 2 GWAS of AD in European ancestry samples. The family-based GWAS of the National Institute of Aging-Late Onset Alzheimer Disease Family Study (NIA-LOAD<sup>38</sup>) included association tests within *APOE* strata. That analysis of 1421  $\epsilon 3$  homozygotes did not provide evidence for an association between rs2968180 and AD, whereas the analysis of 408  $\epsilon 4$  homozygotes supported the association between rs2075650 and AD. This evidence was not independent of the ADGC, because the NIA-LOAD sample was represented in the ADGC LOAD cohort (eTable 1 in the [Supplement](#)). The stage 1 meta-analysis of the International Genomics Alzheimer Project data<sup>39</sup> represented 53 711 participants, including 10 273 from the ADGC.<sup>40</sup> We compared results from the International Genomics Alzheimer Project analysis of 34 152  $\epsilon 4$ -negative participants with our analysis of  $\epsilon 3$  homozygotes. Results were available for 4 SNVs from Table 2; the associations between AD and rs145414981 and rs1979377 were nominally significant, whereas the associations between AD and rs143695016 and rs73572003 were not.<sup>39</sup>

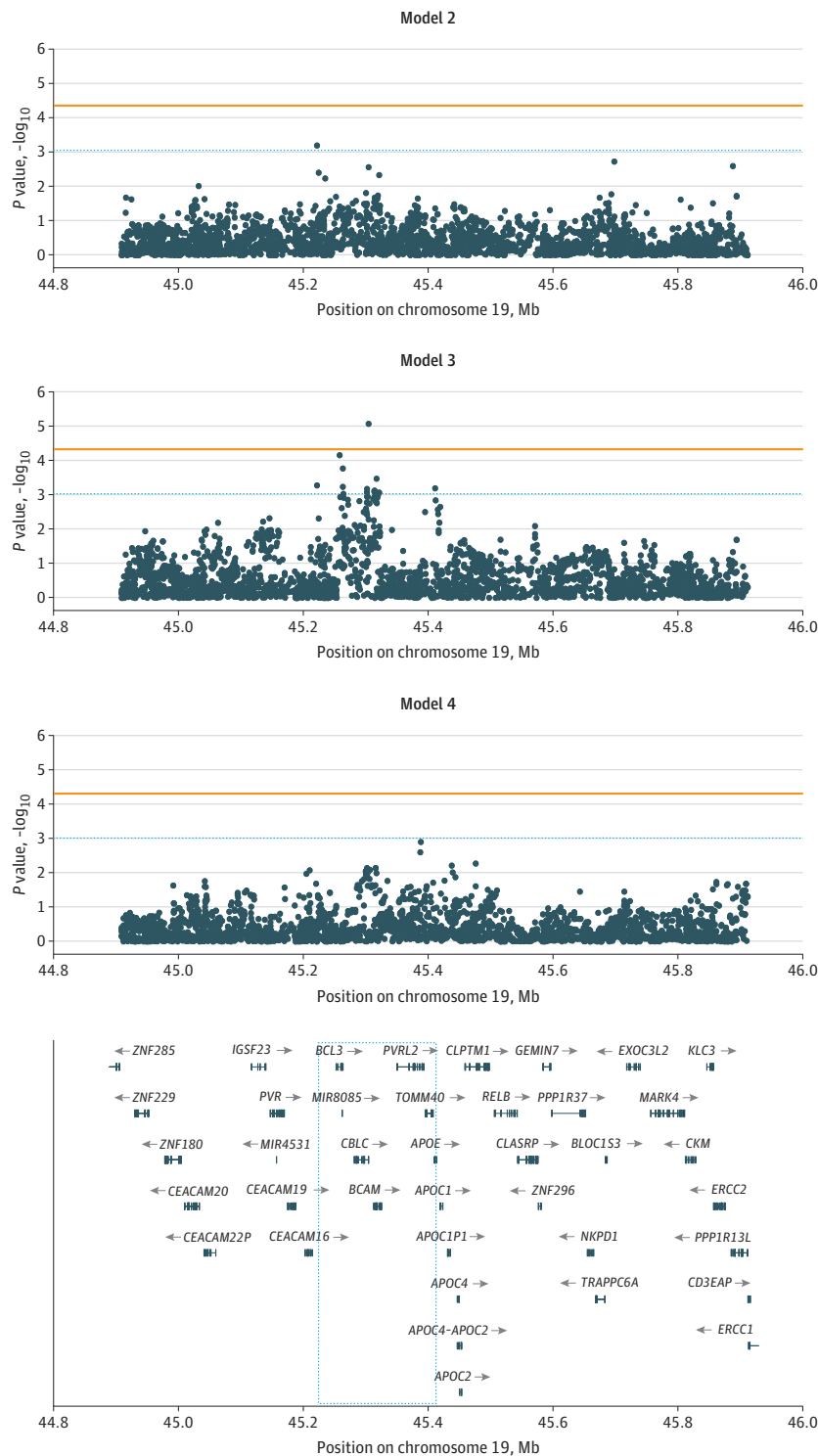
## Discussion

This study found an association between several SNVs for *APOE* and AD risk. Among these, rs192879175 was significantly associated with risk of AD among  $\epsilon 3$  homozygotes, rs143764218 was nominally associated with AD after *APOE* adjustment and among  $\epsilon 3$  homozygotes, and rs2075650 was nominally associated with AD among  $\epsilon 4$  homozygotes. There was a stronger association between SNVs near *APOE* and AD status in the *APOE*-stratified vs the *APOE*-adjusted models. This finding was likely because these strata were restricted to individuals who shared 2 copies of an *APOE* allele identical by state and were therefore more likely to share recent common ancestry.

The *TOMM40* SNV rs2075650 has a long history of an association with AD and related traits, including 8 GWAS for AD risk<sup>41-48</sup> and several studies of healthy aging and longevity.<sup>12,13,49,50</sup> It is a common variant located within intron 2 of *TOMM40* (European MAF = 0.14). rs2075650 overlaps with promoter/enhancer histone marks in immune cells and brain tissues, is predicted to alter 8 transcription factor binding site motifs,<sup>51</sup> and is significantly associated with *TOMM40*, *PVRL2*, and *HIF3A* expression levels.<sup>11,52,53</sup> Both rs192879175 and rs143764218 are uncommon (rs192879175: European MAF, 0.01; rs143764218: European MAF, 0.05), are located between genes, and bear features consistent with regulatory variants. rs192879175 is 1.5 kb 3' of *CBLC*, sits within enhancer

histone marks and a DNase I hypersensitivity site in liver, and is predicted to alter a transcription factor binding site motif.<sup>51</sup> Similarly, rs143764218 is 8.7 kb 3' of *CEACAM16*, sits within promoter/enhancer histone marks and DNase I hypersensitivity site across multiple tissues, and is predicted to alter 7 transcription factor binding site motifs.<sup>51</sup> Neither rs192879175 nor rs143764218 has previously

Figure. Manhattan Plot of Association Results Between Single-Nucleotide Variations in the Apolipoprotein E (APOE) Gene Region and Risk for Alzheimer Disease Across Analysis Models



Variant positions on chromosome 19 are relative to the hg19/GRCh37 reference genome. For model 2, the analysis included all samples adjusted for *APOE*  $\epsilon 2$  and  $\epsilon 4$  allele counts, with 1128 effective independent tests. For model 3, the analysis was restricted to *APOE*  $\epsilon 3$  homozygotes, with 1055 effective independent tests. For model 4, the analysis was restricted to *APOE*  $\epsilon 4$  homozygotes, with 1013 effective independent tests. The horizontal orange line denotes the statistical significance threshold per model ( $P < .05/\text{number of independent tests}$ ), whereas the blue dotted line denotes  $P = 1/\text{number of independent tests}$ . The blue dotted square highlights the genes falling within the region harboring variants with  $P < 1/\text{effective number of tests}$ . Mb indicates megabase.



been shown to be associated with AD or other traits by GWAS,<sup>54</sup> perhaps owing to their uncommon allele frequencies.

We identified associations between noncoding variants in the *APOE* region and risk of AD. Haplotypic differences among participants sharing the same *APOE* genotype are associated with risk of AD.<sup>55-57</sup> Haplotypes derived from rs429358, rs7412, and neighboring noncoding SNVs that vary in frequency across populations are associated with increased risk of AD.<sup>55</sup> Admixture analyses in Puerto Rican, African American, and Caribbean Hispanic data sets have shown that  $\epsilon 4$  alleles inherited on an African background are associated with reduced risk of AD compared with those inherited on a European background, again suggesting that haplotype structures correlated with  $\epsilon 4$  vary between populations and are associated with AD risk.<sup>56,57</sup> All SNVs with significant associations with AD were located within a 186-kb region immediately 5' of *APOE*. All 5 genes in this region share the same transcriptional orientation as *APOE*, suggesting synchronized *cis* regulation might exist. Regulatory variants could modify this transcriptional pathway and subsequently change the gene expression profiles within this entire region.

Few AD genetics studies have accounted for *APOE* genotype, hampering replication efforts. As summarized above, 2 studies<sup>38,39</sup> offered limited support for the 5 SNVs with evidence for association with AD in our study. However, both studies included a subset of the ADGC data analyzed herein and were not truly independent. Larger data sets with high-quality *APOE* genotype data are needed to replicate the results of the present study, particularly for the associations identified among  $\epsilon 4$  homozygotes, including 1326 cases and 177 controls. Laboratory-based procedures such as molecular haplotyping, haplotype-based fine mapping,<sup>58</sup> and reporter assays are needed to investigate the potential functional consequences of SNVs and how those consequences may influence AD pathogenesis.

## Limitations

This study has limitations. Imputed genotype data are not without error. Most of the discordant genotypes we observed involved  $\epsilon 2$  or  $\epsilon 4$  alleles being imputed as  $\epsilon 3$  alleles, consistent with prior work<sup>59</sup>; this likely contributed to the spurious association between rs429358 and AD among the  $\epsilon 3$  homozygotes. The ADGC data were collected on a mixture of older and newer arrays, which may explain some of the discordance we observed between the observed and imputed *APOE* genotypes. We observed lower mismatch rates at  $\epsilon 2$  and  $\epsilon 4$  among those genotyped by an SNV-based approach

Table 2. Additional SNVs Within the *APOE* Region With an Association With AD Status Across Models 2, 3, and 4

Model <sup>a</sup>	SNV	BP37	ALT	No. of participants	AAC	AAF	OR (95% CI)	P value
2	rs143764218	45222739	AC	16 714	915	0.03	0.76 (0.64-0.89)	$6.26 \times 10^{-4}$
3	rs143764218	45222739	AC	7794	507	0.03	0.69 (0.56-0.85)	$5.20 \times 10^{-4}$
3	rs1979377	45259002	C	7396	801	0.05	0.71 (0.59-0.84)	$6.84 \times 10^{-5}$
3	chr19:45264102:1	45264102	TG	7518	555	0.04	0.68 (0.56-0.83)	$1.67 \times 10^{-4}$
3	rs10416720	45264110	T	7491	846	0.06	0.75 (0.63-0.88)	$5.72 \times 10^{-4}$
3	rs145414981	45265003	C	7355	718	0.05	0.74 (0.62-0.88)	$9.18 \times 10^{-4}$
3	rs73572003	45302665	G	7982	1250	0.08	0.79 (0.69-0.91)	$8.32 \times 10^{-4}$
3	rs143695016	45302840	T	8003	1251	0.08	0.79 (0.68-0.90)	$6.59 \times 10^{-4}$
3	rs192879175	45305363	T	8635	256	0.01	0.50 (0.37-0.68)	$8.30 \times 10^{-6b}$
3	rs28399650	45314364	A	8633	433	0.03	0.68 (0.54-0.85)	$7.80 \times 10^{-4}$
3	rs28399652	45314975	G	8640	434	0.03	0.67 (0.54-0.85)	$7.36 \times 10^{-4}$
3	rs2968180	45318153	T	8218	1542	0.09	0.79 (0.70-0.90)	$3.31 \times 10^{-4}$
3	rs117737673	45322316	T	8489	546	0.03	0.70 (0.57-0.86)	$8.42 \times 10^{-4}$

Abbreviations: AAC, alternate allele count; AAF, alternate allele frequency; AD, Alzheimer disease; ALT, alternate allele; *APOE*, apolipoprotein E; BP37, position on the hg19 map; OR, odds ratio; SNV, single-nucleotide variant.

<sup>a</sup> Model 2 included all samples, adjusted for *APOE*  $\epsilon 2$  and  $\epsilon 4$  allele counts; model 3, restricted to  $\epsilon 3$  homozygotes; and model 4, restricted to  $\epsilon 4$  homozygotes. The

effective number of tests under model 2 was 1128 of 3408 SNVs; under model 3, 1055 of 3346 SNVs; and under model 4, 1013 of 3238 SNVs. All variants are on chromosome 19.

<sup>b</sup> Indicates passing the model-specific significance threshold.

with high accuracy (error rate, 0.002<sup>37</sup>) compared with those genotyped by next-generation sequencing, suggesting that genotyping error may explain these differences. The stronger correlation between the APOE region genotypes in the ADGC compared with the 1KGP Europeans (consistent with previous reports of differing LD patterns between AD cases and controls<sup>60</sup>) suggests that using sequence data generated on a large and diverse sample set ascertained for AD status as a reference may improve the quality of imputed genotypes in AD GWAS. Our data represent only those with European ancestry; thus, our results may not apply to other populations.

---

## Conclusions

This genetic association study found that  $\epsilon 2/\epsilon 3/\epsilon 4$  alleles as well as other variants in the APOE region were associated with AD risk. Although future work in independent data are needed to replicate these results, our findings appear to provide valuable new candidate sites for targeted genetic analyses on larger sample sets representing diverse ethnic groups. The findings suggest that increased LD between SNVs within the APOE region in samples ascertained for AD vs population samples may influence the accuracy of imputation within AD-related data sets. The correlation between imputed vs measured  $\epsilon 2$  and  $\epsilon 4$  genotypes within the ADGC varied by genotyping platform, suggesting next-generation sequencing at rs7412 and rs429358 may not be as accurate as alternative approaches. Association testing results in the APOE region varies between models adjusting for or stratifying by  $\epsilon 2/\epsilon 3/\epsilon 4$  genotype; future GWAS using these alternative approaches may yield novel results in existing data sets.

---

## ARTICLE INFORMATION

**Accepted for Publication:** July 13, 2020.

**Published:** October 22, 2020. doi:[10.1001/jamanetworkopen.2020.17666](https://doi.org/10.1001/jamanetworkopen.2020.17666)

**Open Access:** This is an open access article distributed under the terms of the [CC-BY License](https://creativecommons.org/licenses/by/4.0/). © 2020 Blue EE et al. JAMA Network Open.

**Corresponding Author:** Elizabeth E. Blue, PhD, Division of Medical Genetics, Department of Medicine, University of Washington School of Medicine, 1959 NE Pacific St, PO Box 357720, Seattle, WA 98195 ([em27@uw.edu](mailto:em27@uw.edu)).

**Author Affiliations:** Division of Medical Genetics, Department of Medicine, University of Washington School of Medicine, Seattle (Blue); Department of Biostatistics, University of Washington, Seattle (Cheng); Geriatric Research, Education, and Clinical Center, Veterans Affairs Puget Sound Health Care System, Seattle, Washington (Chen, Yu); Division of Gerontology and Geriatric Medicine, Department of Medicine, University of Washington, Seattle (Yu).

**Author Contributions:** Dr Blue had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

*Concept and design:* Blue, Yu.

*Acquisition, analysis, or interpretation of data:* All authors.

*Drafting of the manuscript:* Blue, Cheng.

*Critical revision of the manuscript for important intellectual content:* Blue, Chen, Yu.

*Statistical analysis:* Blue, Cheng, Chen.

*Obtained funding:* Blue.

*Administrative, technical, or material support:* Blue, Yu.

*Supervision:* Blue, Yu.

**Conflict of Interest Disclosures:** Dr Blue reported receiving grants from the National Institute on Aging (NIA), National Institutes of Health (NIH), during the conduct of the study. No other disclosures were reported.

**Funding/Support:** This work was supported in part by the US Department of Veterans Affairs Office of Research and Development Biomedical Laboratory Research Program and grants IO1BX004823, P5OAG05136, and RO1AG059737 from the NIH/NIA. The Alzheimer's Disease Genetics Consortium (ADGC) and its contributors were supported by grants U01AG032984, RC2AG036528, and U24AG21886 from the NIH/NIA (for the collection,



management, analysis, and interpretation of the data). Data for this study were prepared, archived, and distributed by the NIA Alzheimer's Disease Data Storage Site (NIAGADS) at the University of Pennsylvania, which received funding support for the collection, management, analysis, and interpretation of the data from grant U24AG041689-01 from the NIH/NIA. Additional ADGC funding includes grant U01AG016976 from the National Alzheimer's Coordinating Center; grants U24AG026395 and R01AG041797 from NIA-Late Onset Alzheimer Disease Family Study; grant P30AG019610 from Banner Sun Health Research Institute; grants P30AG013846, U01AG10483, R01CA129769, R01MH080295, R01AG017173, R01AG025259, and R01AG33193 from Boston University; grants P50AG008702 and R37AG015473 from Columbia University; grants P30AG028377 and AG05128 from Duke University; grant AG025688 from Emory University; grants U01AG006781, U01HG004610, and U01HG006375 from Group Health Research Institute; grant P30AG10133 from Indiana University; grants P50AG005146 and R01AG020688 from The Johns Hopkins University; grant 50AG005134 from Massachusetts General Hospital; grant P50AG016574 from the Mayo Clinic; grants P50AG005138 and P01AG002219 from Mount Sinai School of Medicine; grants P30AG08051, UL1RR029893, 5R01AG012101, 5R01AG022374, 5R01AG013616, 1RC2AG036502, and 1R01AG035137 from New York University; grant P30AG013854 from Northwestern University; grants P30AG008017 and R01AG026916 from Oregon Health & Science University; grants P30AG010161, R01AG019085, R01AG15819, R01AG17917, and R01AG30146 from Rush University; grant R01NS059873 from the Translational Genomics Research Institute (TGen); grant P50AG016582 from University of Alabama at Birmingham; grant R01AG031581 from University of Arizona; grant P30AG010129 from University of California, Davis; grant P50AG016573 from University of California, Irvine; grant P50AG016570 from University of California, Los Angeles; grant P50AG005131 from University of California, San Diego; grants P50AG023501 and P01AG019724 from University of California, San Francisco; grants P30AG028383 and AG05144 from University of Kentucky; grant P50AG008671 from University of Michigan; grant P30AG010124 from University of Pennsylvania; grants P50AG005133, AG030653, AG041718, AG07562, and AG02365 from University of Pittsburgh; grant P50AG005142 from University of Southern California; grant P30AG012300 from University of Texas Southwestern Medical Center; grants R01AG027944, AG010491, AG027944, AG021547, and AG019757 from University of Miami; grant P50AG005136 from University of Washington; grant P50AG033514 from University of Wisconsin; grant R01AG019085 from Vanderbilt University; and grants P50AG005681 and P01AG03991 from Washington University. The Kathleen Price Bryan Brain Bank at Duke University Medical Center is supported by grant NS39764 from the National Institute of Neurological Disorders and Stroke, grant MH60451 from the National Institute of Mental Health, and GlaxoSmithKline. Genotyping of the TGen2 cohort was supported by Kronos Science. The TGen series was also supported by grant AG041232 from the NIA (Amanda J. Myers and Matthew J. Huentelman), The Banner Alzheimer's Foundation, The Johnnie B. Byrd Sr Alzheimer's Institute, the Medical Research Council (MRC), and the state of Arizona and also includes samples from the following sites: Newcastle Brain Tissue Resource (funded via the MRC, local National Health Service [NHS] trusts, and Newcastle University), MRC London Brain Bank for Neurodegenerative Diseases (funding via the MRC), South West Dementia Brain Bank (funding via numerous sources including the Higher Education Funding Council for England, Alzheimer's Research Trust, BRACE, and North Bristol NHS Trust Research and Innovation Department and Dementias & Neurodegenerative Diseases Research Network), The Netherlands Brain Bank (funding via numerous sources including Stichting MS Research, Brain Net Europe, Hersenstichting Nederland Breinbrekend Werk, International Parkinson Fonds, and Internationale Stichting Alzheimer Onderzoek), and Institut de Neuropatologia, Servei Anatomia Patologica, Universitat de Barcelona. Alzheimer's Disease Neuroimaging Initiative (ADNI) data collection and sharing was supported by grant U01AG024904 from the NIH and grant W81XWH-12-2-0012 from the Department of Defense. ADNI is supported by the NIA, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from AbbVie Alzheimer's Association, Alzheimer's Drug Discovery Foundation, Araclon Biotech, BioClinica, Inc, Biogen, Bristol-Myers Squibb Company, CereSpir, Inc, Eisai Co, Ltd, Elan Pharmaceuticals, LLC, Eli Lilly and Company, EuroImmun, F. Hoffmann-La Roche, Ltd, and its affiliated company Genentech, Inc, Fujirebio, GE Healthcare, IXICO plc, Janssen Alzheimer Immunotherapy Research & Development, LLC, Johnson & Johnson Pharmaceutical Research & Development, LLC, Lumosity, H Lundbeck A/S, Merck & Co, Meso Scale Diagnostics, LLC, NeuroRx Research, Neurotrack Technologies, Inc, Novartis Pharmaceuticals Corporation, Pfizer, Inc, Piramal Imaging, Servier, Takeda Pharmaceutical Company, and Transition Therapeutics, Inc. The Canadian Institutes of Health Research supported ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the NIH ([www.fnih.org](http://www.fnih.org)). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. The ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. This study was also supported by the Alzheimer's Association (Lindsay A. Farrer, IIRG-08-89720; Margaret Pericak-Vance, IIRG-05-14147); by the US Department of Veterans Affairs Administration, Office of Research and Development, Biomedical Laboratory Research Program; and by Wellcome Trust, Howard Hughes Medical Institute, and the Canadian Institute of Health Research (Peter St George-Hyslop, MD).

**Role of the Funder/Sponsor:** The sponsors had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

**Group Information:** The following is a list of members of the ADGC: Erin Abner, PhD, Perrie M. Adams, PhD, Marilyn S. Albert, PhD, Roger L. Albin, MD, Liana G. Apostolova, MD, Steven E. Arnold, MD, Sanjay Asthana, MD, Craig S. Atwood, PhD, Clinton T. Baldwin, PhD, Cynthia M. Carlsson, MD, GRECC, Robert C. Barber, PhD, Lisa L. Barnes, PhD, Sandra Barral, PhD, Thomas G. Beach, MD, PhD, James T. Becker, PhD, Gary W. Beecham, PhD, Duane Beekly, BS, David A. Bennett, MD, Eileen H. Bigio, MD, Thomas D. Bird, MD, Deborah Blacker, MD, Bradley F. Boeve, MD, James D. Bowen, MD, Adam Boxer, MD, PhD, James R. Burke, MD, PhD, Jeffrey M. Burns, MD, MS, Joseph D. Buxbaum, PhD, Nigel J. Cairns, PhD, Laura B. Cantwell, MPH, Chuanhai Cao, PhD, Chris S. Carlson, PhD, Regina M. Carney, MD, Minerva M. Carrasquillo, PhD, Helena C. Chui, MD, Paul K. Crane, MD, PhD, Elizabeth A. Crocco, MD, David H. Cribbs, PhD, Carlos Cruchaga, PhD, Charles DeCarli, MD, Philip L. De Jager, MD, PhD, Malcolm Dick, MD, Dennis W. Dickson, MD, Rachelle S. Doody, MD, PhD, Nilufer Ertekin-Taner, MD, PhD, Denis A. Evans, MD, Kelley M. Faber, MS, Thomas J. Fairchild, PhD, Kenneth B. Fallon, MD, David W. Fardo, PhD, Martin R. Farlow, MD, Lindsay A. Farrer, PhD, Steven Ferris, PhD, Tatiana M. Foroud, PhD, Matthew P. Frosch, MD, PhD, Douglas R. Galasko, MD, Marla Gearing, PhD, Daniel H. Geschwind, MD, PhD, Bernardino Ghetti, MD, John R. Gilbert, PhD, Alison M. Goate, DPhil, Neill R. Graff-Radford, MD, Robert C. Green, MD, MPH, John H. Growdon, MD, Jonathan L. Haines, PhD, Hakon Hakonarson, MD, PhD, Kara L. Hamilton-Nelson, MPH, John Hardy, PhD, Lindy E. Harrell, MD, PhD, Lawrence S. Honig, MD, PhD, Ryan M. Huebinger, PhD, Matthew J. Huentelman, PhD, Christine M. Hulette, MD, Bradley T. Hyman, MD, PhD, Gail P. Jarvik, MD, PhD, Lee-Way Jin, MD, PhD, Gyungah Jun, PhD, M. Ilyas Kamboh, PhD, Anna Karydas, BA, Mindy J. Katz, MPH, John S. K. Kauwe, PhD, Jeffrey A. Kaye, MD, C. Dirk Keene, MD, PhD, Ronald Kim, MD, Neil W. Kowall, MD, Joel H. Kramer, PsyD, Walter A. Kukull, PhD, Brian W. Kunkle, PhD, MPH, Amanda P. Kuzma, MS, Frank M. LaFerla, PhD, James J. Lah, MD, PhD, Eric B. Larson, MD, PhD, Yuk Ye Leung, PhD, James B. Leverenz, MD, Allan I. Levey, MD, PhD, Ge Li, MD, PhD, Andrew P. Lieberman, MD, PhD, Oscar L. Lopez, MD, Richard B. Lipton, MD, Kathryn L. Lunetta, PhD, Constantine G. Lyketsos, MD, MHS, John Malamon, MSE, Daniel C. Marson, JD, PhD, Eden R. Martin, PhD, Frank Martiniuk, PhD, Deborah C. Mash, PhD, Eliezer Masliah, MD, Richard Mayeux, MD, MSc, Wayne C. McCormick, MD, PhD, Susan M. McCurry, PhD, Andrew N. McDavid, BA, Stefan McDonough, PhD, Ann C. McKee, MD, Marsel Mesulam, MD, Carol A. Miller, MD, Thomas J. Montine, MD, PhD, Shubhabrata Mukherjee, PhD, Amanda J. Myers, PhD, Bruce L. Miller, MD, Joshua W. Miller, PhD, John C. Morris, MD, Adam C. Naj, PhD, Sid O'Bryant, PhD, John M. Olichney, MD, Joseph E. Parisi, MD, Henry L. Paulson, MD, PhD, Margaret A. Pericak-Vance, PhD, Elaine Peskind, MD, Ronald C. Petersen, MD, PhD, Aimee Pierce, MD, Wayne W. Poon, PhD, Huntington Potter, PhD, Liming Qu, MS, Joseph F. Quinn, MD, Ashok Raj, MD, Murray Raskind, MD, Eric M. Reiman, MD, Barry Reisberg, MD, Joan S. Reisch, PhD, Christiane Reitz, MD, PhD, John M. Ringman, MD, Erik D. Roberson, MD, PhD, Ekaterina Rogaeva, PhD, Howard J. Rosen, MD, Roger N. Rosenberg, MD, Donald R. Royall, MD, Mark A. Sager, MD, Mary Sano, PhD, Andrew J. Saykin, PsyD, Gerard D. Schellenberg, PhD, Julie A. Schneider, MD, Lon S. Schneider, MD, William W. Seeley, MD, Amanda G. Smith, MD, Joshua A. Sonnen, MD, Salvatore Spina, MD, Peter St George-Hyslop, MD, Robert A. Stern, PhD, Russell H. Swerdlow, MD, Rudolph E. Tanzi, PhD, John Q. Trojanowski, MD, PhD, Juan C. Troncoso, MD, Debby W. Tsuang, MD, Otto Valladares, MS, Viviana M. Van Deerlin, MD, PhD, Linda J. Van Eldik, PhD, Badri N. Vardarajan, PhD, Harry V. Vinters, MD, Jean Paul Vonsattel, MD, Li-San Wang, PhD, Sandra Weintraub, PhD, Kathleen A. Welsh-Bohmer, PhD, Kirk C. Wilhelmsen, MD, PhD, Jennifer Williamson, MS, Thomas S. Wingo, MD, Randall L. Wolter, MD, PhD, Clinton B. Wright, MD, MS, Chuang-Kuo Wu, MD, PhD, Steven G. Yankin, MD, PhD, Chang-En Yu, PhD, Lei Yu, PhD, and Yi Zhao, MS.

**Additional Contributions:** We thank D. Stephen Snyder, PhD, and Marilyn Miller, PhD, from the NIA, who are ex-officio ADGC members and were not compensated for this work; the contributors who collected samples used in this study, and the patients and their families.

## REFERENCES

1. 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(5123):921-923. doi:10.1126/science.8346443
2. Strittmatter WJ, Saunders AM, Schmechel D, et al. Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. *Proc Natl Acad Sci U S A*. 1993;90(5):1977-1981. doi:10.1073/pnas.90.5.1977
3. Alzheimer's Association. 2017 Alzheimer's disease facts and figures. *Alzheimer's Dementia*. 2017;13(4):325-373. doi:10.1016/j.jalz.2017.02.001
4. Pimenova AA, Raj T, Goate AM. Untangling genetic risk for Alzheimer's disease. *Biol Psychiatry*. 2018;83(4):300-310. doi:10.1016/j.biopsych.2017.05.014

5. Kunkle BW, Grenier-Boley B, Sims R, et al; Alzheimer Disease Genetics Consortium (ADGC); European Alzheimer's Disease Initiative (EADI); Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium (CHARGE); Genetic and Environmental Risk in AD/Defining Genetic, Polygenic and Environmental Risk for Alzheimer's Disease Consortium (GERAD/PERADES). Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates A $\beta$ , tau, immunity and lipid processing. *Nat Genet*. 2019;51(3):414-430. doi:10.1038/s41588-019-0358-2
6. Kinoshita J, Clark T. Alzforum. *Methods Mol Biol*. 2007;401:365-381. doi:10.1007/978-1-59745-520-6\_19
7. Yashin AI, Fang F, Kovtun M, et al. Hidden heterogeneity in Alzheimer's disease: insights from genetic association studies and other analyses. *Exp Gerontol*. 2018;107:148-160. doi:10.1016/j.exger.2017.10.020
8. Nho K, Kim S, Horgusluoglu E, et al; Alzheimer's Disease Neuroimaging Initiative (ADNI). Association analysis of rare variants near the APOE region with CSF and neuroimaging biomarkers of Alzheimer's disease. *BMC Med Genomics*. 2017;10(suppl 1):29. doi:10.1186/s12920-017-0267-0
9. Prendecki M, Florzak-Wypianska J, Kowalska M, et al. Biothiols and oxidative stress markers and polymorphisms of TOMM40 and APOC1 genes in Alzheimer's disease patients. *Oncotarget*. 2018;9(81):35207-35225. doi:10.18632/oncotarget.26184
10. Huentelman M, Corneveaux J, Myers A, et al. Genome-wide association study for Alzheimer's disease risk in a large cohort of clinically characterized and neuropathologically verified subjects. *Alzheimer Dementia*. 2010;6(4):e13. doi:10.1016/j.jalz.2010.08.041
11. Liu C, Chyr J, Zhao W, et al; Alzheimer's Disease Neuroimaging Initiative. Genome-wide association and mechanistic studies indicate that immune response contributes to Alzheimer's disease development. *Front Genet*. 2018;9:410. doi:10.3389/fgene.2018.00410
12. Pilling LC, Atkins JL, Bowman K, et al. Human longevity is influenced by many genetic variants: evidence from 75,000 UK Biobank participants. *Aging (Albany NY)*. 2016;8(3):547-560. doi:10.18632/aging.100930
13. Yashin AI, Arbeevev KG, Wu D, et al. Genetics of human longevity from incomplete data: new findings from the Long Life Family Study. *J Gerontol A Biol Sci Med Sci*. 2018;73(11):1472-1481. doi:10.1093/gerona/gly057
14. Shadyab AH, Kooperberg C, Reiner AP, et al. Replication of genome-wide association study findings of longevity in White, African American, and Hispanic women: the Women's Health Initiative. *J Gerontol A Biol Sci Med Sci*. 2017;72(10):1401-1406. doi:10.1093/gerona/glw198
15. Alzheimer's Disease Genetics Consortium. Accessed May 31, 2018. <http://www.adgenetics.org/content/feedback-and-queries>
16. Naj AC, Jun G, Beecham GW, et al. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat Genet*. 2011;43(5):436-441. doi:10.1038/ng.801
17. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. *Nat Methods*. 2011;9(2):179-181. doi:10.1038/nmeth.1785
18. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*. 2009;5(6):e1000529. doi:10.1371/journal.pgen.1000529
19. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol*. 2010;34(8):816-834. doi:10.1002/gepi.20533
20. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet*. 2012;44(8):955-959. doi:10.1038/ng.2354
21. Auton A, Brooks LD, Durbin RM, et al; 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74. doi:10.1038/nature15393
22. Devlin B, Roeder K. Genomic control for association studies. *Biometrics*. 1999;55(4):997-1004. doi:10.1111/j.0006-341X.1999.00997.x
23. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*. 1984;34(7):939-944. doi:10.1212/WNL.34.7.939
24. Dubois B, Feldman HH, Jacova C, et al. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *Lancet Neurol*. 2007;6(8):734-746. doi:10.1016/S1474-4422(07)70178-3
25. GENESIS: Genetic Estimation and Inference in Structured Samples (GENESIS): statistical methods for analyzing genetic data from samples with population structure and/or relatedness. R package version 2.4.0. [computer program]. R Foundation for Statistical Computing; 2016.

26. Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*. 2012;28(24):3326-3328. doi:10.1093/bioinformatics/bts606
27. Conomos MP, Miller MB, Thornton TA. Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness. *Genet Epidemiol*. 2015;39(4):276-293. doi:10.1002/gepi.21896
28. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. *Bioinformatics*. 2010;26(22):2867-2873. doi:10.1093/bioinformatics/btq559
29. Conomos MP, Reiner AP, Weir BS, Thornton TA. Model-free estimation of recent genetic relatedness. *Am J Hum Genet*. 2016;98(1):127-148. doi:10.1016/j.ajhg.2015.11.022
30. Li MX, Yeung JM, Cherny SS, Sham PC. Evaluating the effective numbers of independent tests and significant P-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Hum Genet*. 2012;131(5):747-756. doi:10.1007/s00439-011-1118-2
31. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559-575. doi:10.1086/519795
32. R: a language and environment for statistical computing [computer program]. R Foundation for Statistical Computing; 2011.
33. Ward LD, Kellis M. HaploReg v4.1: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res*. 2016;44(D1):D877-D881. doi:10.1093/nar/gkv1340
34. Lek M, Karczewski KJ, Minikel EV, et al; Exome Aggregation Consortium. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536(7616):285-291. doi:10.1038/nature19057
35. Roses AD, Lutz MW, Amrine-Madsen H, et al. A TOMM40 variable-length polymorphism predicts the age of late-onset Alzheimer's disease. *Pharmacogenomics J*. 2010;10(5):375-384. doi:10.1038/tpj.2009.69
36. Lutz MW, Crenshaw DG, Saunders AM, Roses AD. Genetic variation at a single locus and age of onset for Alzheimer's disease. *Alzheimers Dement*. 2010;6(2):125-131. doi:10.1016/j.jalz.2010.01.011
37. Lupton MK, Medland SE, Gordon SD, et al. Accuracy of inferred APOE genotypes for a range of genotyping arrays and imputation reference panels. *J Alzheimers Dis*. 2018;64(1):49-54. doi:10.3233/JAD-171104
38. Wijsman EM, Pankratz ND, Choi Y, et al; NIA-LOAD/NCRAD Family Study Group. Genome-wide association of familial late-onset Alzheimer's disease replicates BIN1 and CLU and nominates CUGBP2 in interaction with APOE. *PLoS Genet*. 2011;7(2):e1001308. doi:10.1371/journal.pgen.1001308
39. Jun G, Ibrahim-Verbaas CA, Vronskaya M, et al; IGAP Consortium. A novel Alzheimer disease locus located near the gene encoding tau protein. *Mol Psychiatry*. 2016;21(1):108-117. doi:10.1038/mp.2015.23
40. Lambert JC, Ibrahim-Verbaas CA, Harold D, et al; European Alzheimer's Disease Initiative (EADI); Genetic and Environmental Risk in Alzheimer's Disease; Alzheimer's Disease Genetic Consortium; Cohorts for Heart and Aging Research in Genomic Epidemiology. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet*. 2013;45(12):1452-1458. doi:10.1038/ng.2802
41. Harold D, Abraham R, Hollingworth P, et al. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat Genet*. 2009;41(10):1088-1093. doi:10.1038/ng.440
42. Heinzen EL, Need AC, Hayden KM, et al. Genome-wide scan of copy number variation in late-onset Alzheimer's disease. *J Alzheimers Dis*. 2010;19(1):69-77. doi:10.3233/JAD-2010-1212
43. Lambert JC, Heath S, Even G, et al; European Alzheimer's Disease Initiative Investigators. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat Genet*. 2009;41(10):1094-1099. doi:10.1038/ng.439
44. Seshadri S, Fitzpatrick AL, Ikram MA, et al; CHARGE Consortium; GERAD1 Consortium; EADI Consortium. Genome-wide analysis of genetic loci associated with Alzheimer disease. *JAMA*. 2010;303(18):1832-1840. doi:10.1001/jama.2010.574
45. Pérez-Palma E, Bustos BI, Villamán CF, et al; Alzheimer's Disease Neuroimaging Initiative; NIA-LOAD/NCRAD Family Study Group. Overrepresentation of glutamate signaling in Alzheimer's disease: network-based pathway enrichment using meta-analysis of genome-wide association studies. *PLoS One*. 2014;9(4):e95413. doi:10.1371/journal.pone.0095413
46. Nelson PT, Estus S, Abner EL, et al; Alzheimer's Disease Genetic Consortium. ABCC9 gene polymorphism is associated with hippocampal sclerosis of aging pathology. *Acta Neuropathol*. 2014;127(6):825-843. doi:10.1007/s00401-014-1282-2

47. Naj AC, Beecham GW, Martin ER, et al. Dementia revealed: novel chromosome 6 locus for late-onset Alzheimer disease provides genetic evidence for folate-pathway abnormalities. *PLoS Genet*. 2010;6(9):e1001130. doi:10.1371/journal.pgen.1001130
48. Kim S, Swaminathan S, Shen L, et al; Alzheimer's Disease Neuroimaging Initiative. Genome-wide association study of CSF biomarkers Abeta1-42, t-tau, and p-tau181p in the ADNI cohort. *Neurology*. 2011;76(1):69-79. doi:10.1212/WNL.0b013e318204a397
49. Deelen J, Beekman M, Uh HW, et al. Genome-wide association study identifies a single major locus contributing to survival into old age; the APOE locus revisited. *Aging Cell*. 2011;10(4):686-698. doi:10.1111/j.1474-9726.2011.00705.x
50. Sebastiani P, Solovieff N, Dewan AT, et al. Genetic signatures of exceptional longevity in humans. *PLoS One*. 2012;7(1):e29848. doi:10.1371/journal.pone.0029848
51. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res*. 2012;40(database issue):D930-D934. doi:10.1093/nar/gkr917
52. Lappalainen T, Sammeth M, Friedländer MR, et al; Geuvadis Consortium. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature*. 2013;501(7468):506-511. doi:10.1038/nature12531
53. Greenawalt DM, Dobrin R, Chudin E, et al. A survey of the genetics of stomach, liver, and adipose gene expression from a morbidly obese cohort. *Genome Res*. 2011;21(7):1008-1016. doi:10.1101/gr.112821.110
54. Buniello A, MacArthur JAL, Cerezo M, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res*. 2019;47(D1):D1005-D1012. doi:10.1093/nar/gky1120
55. Babenko VN, Afonnikov DA, Ignatieva EV, Klimov AV, Gusev FE, Rogaev EI. Haplotype analysis of APOE intragenic SNPs. *BMC Neurosci*. 2018;19(suppl 1):16. doi:10.1186/s12868-018-0413-4
56. Blue EE, Horimoto ARVR, Mukherjee S, Wijsman EM, Thornton TA. Local ancestry at APOE modifies Alzheimer's disease risk in Caribbean Hispanics. *Alzheimers Dement*. 2019;15(12):1524-1532. doi:10.1016/j.jalz.2019.07.016
57. Rajabli F, Feliciano BE, Celis K, et al. Ancestral origin of ApoE ε4 Alzheimer disease risk in Puerto Rican and African American populations. *PLoS Genet*. 2018;14(12):e1007791. doi:10.1371/journal.pgen.1007791
58. Zhou X, Chen Y, Mok KY, et al; Alzheimer's Disease Neuroimaging Initiative. Non-coding variability at the APOE locus contributes to the Alzheimer's risk. *Nat Commun*. 2019;10(1):3310. doi:10.1038/s41467-019-10945-z
59. Oldmeadow C, Holliday EG, McEvoy M, et al. Concordance between direct and imputed APOE genotypes using 1000 Genomes data. *J Alzheimers Dis*. 2014;42(2):391-393. doi:10.3233/JAD-140846
60. Kulminski AM, Huang J, Wang J, He L, Loika Y, Culminskaya I. Apolipoprotein E region molecular signatures of Alzheimer's disease. *Aging Cell*. 2018;17(4):e12779. doi:10.1111/acer.12779

#### SUPPLEMENT.

**eMethods.** Software Tools

**eTable 1.** Sample Summary Table by ADGC Cohort

**eTable 2.** Evidence for Linkage Disequilibrium Among the TOMM40, APOC1, and APOE SNVs

**eTable 3.** Evidence for Association Between rs8106922 With and Without APOE Adjustment or Stratification

**eTable 4.** Comparison of APOE ε2 and ε4 Genotypes With Imputed Genotypes at rs7412 and rs429358

**eFigure 1.** Principal Components Analysis Results

**eFigure 2.** Difference Between Kinship Estimates Based on Genotypes for All Autosomes vs All Autosomes Except Chromosome 19

**eFigure 3.** Quantile-Quantile Plots of Genome-Wide Association Tests Under Each Analysis Model

**eReferences.**