

RESEARCH ARTICLE

A locus at 19q13.31 significantly reduces the *ApoE* ϵ 4 risk for Alzheimer's Disease in African Ancestry

Farid Rajabli¹, Gary W. Beecham^{1,2}, Hugh C. Hendrie³, Olusegun Baiyewu⁴, Adesola Ogunniyi⁴, Sujuan Gao⁵, Nicholas A. Kushch¹, Marina Lipkin-Vasquez¹, Kara L. Hamilton-Nelson¹, Juan I. Young^{1,2}, Derek M. Dykxhoorn^{1,2}, Karen Nuytemans^{1,2}, Brian W. Kunkle^{1,2}, Liyong Wang^{1,2}, Fulai Jin⁶, Xiaoxiao Liu⁶, Briseida E. Feliciano-Astacio⁷, Alzheimer's Disease Sequencing Project, Alzheimer's Disease Genetic Consortium, Gerard D. Schellenberg⁸, Clifton L. Dalgard⁹, Anthony J. Griswold^{1,2}, Goldie S. Byrd¹⁰, Christiane Reitz¹¹, Michael L. Cuccaro^{1,2}, Jonathan L. Haines¹², Margaret A. Pericak-Vance^{1,2}, Jeffery M. Vance^{1,2*}



OPEN ACCESS

Citation: Rajabli F, Beecham GW, Hendrie HC, Baiyewu O, Ogunniyi A, Gao S, et al. (2022) A locus at 19q13.31 significantly reduces the *ApoE* ϵ 4 risk for Alzheimer's Disease in African Ancestry. *PLoS Genet* 18(7): e1009977. <https://doi.org/10.1371/journal.pgen.1009977>

Editor: Gregory S. Barsh, HudsonAlpha Institute for Biotechnology, UNITED STATES

Received: December 4, 2021

Accepted: May 27, 2022

Published: July 5, 2022

Copyright: © 2022 Rajabli et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Data are available through the National Institute on Aging Genetics of Alzheimer's Disease Data Storage Site (NIAGADS) Data Sharing Service (DSS): <https://dss.niagads.org/datasets/ng00067/>.

Funding: This investigation was supported by grant AG16002 (GDS), AG032984 (GDS), AG052410 (GWB, GSB, MAPV, CR), AG072547 (MAPV), AG057659 (MAPV), AG009956 (HCH), AG059018 (JMV), and AG058654 (JLH, MAPV) from the National Institutes on Aging of NIH, ZEN-

1 John P. Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, Florida, United States of America, 2 Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miller School of Medicine, Miami, Florida, United States of America, 3 Department of Psychiatry, Indiana University School of Medicine, Indianapolis, Indiana, United States of America, 4 College of Medicine, University of Ibadan, Ibadan, Nigeria, 5 Department of Biostatistics and Health Data Science, Indiana University School of Medicine, Indianapolis, Indiana, United States of America, 6 Department of Genetics and Genome Sciences, Case Western Reserve University, Cleveland, Ohio, United States of America, 7 Universidad Central del Caribe, Bayamón, Puerto Rico, United States of America, 8 Penn Neurodegeneration Genomics Center, Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, United States of America, 9 Department of Anatomy, Physiology & Genetics, Uniformed Services University of the Health Sciences, Bethesda, Maryland, United States of America, 10 Maya Angelou Center for Health Equity, Wake Forest School of Medicine, Winston-Salem, North Carolina, United States of America, 11 Gertrude H. Sergievsky Center, Taub Institute for Research on the Aging Brain, Departments of Neurology, Psychiatry, and Epidemiology, College of Physicians and Surgeons, Columbia University, New York, New York State, United States of America, 12 Department of Population & Quantitative Health Sciences, Cleveland Institute for Computational Biology, Case Western Reserve University School of Medicine, Cleveland, Ohio, United States of America

* jvance@med.miami.edu

Abstract

African descent populations have a lower Alzheimer disease risk from *ApoE* ϵ 4 compared to other populations. Ancestry analysis showed that the difference in risk between African and European populations lies in the ancestral genomic background surrounding the *ApoE* locus (local ancestry). Identifying the mechanism(s) of this protection could lead to greater insight into the etiology of Alzheimer disease and more personalized therapeutic intervention. Our objective is to follow up the local ancestry finding and identify the genetic variants that drive this risk difference and result in a lower risk for developing Alzheimer disease in African ancestry populations. We performed association analyses using a logistic regression model with the *ApoE* ϵ 4 allele as an interaction term and adjusted for genome-wide ancestry, age, and sex. Discovery analysis included imputed SNP data of 1,850 Alzheimer disease and 4,331 cognitively intact African American individuals. We performed replication analyses on 63 whole genome sequenced Alzheimer disease and 648 cognitively intact

19-591586 (JMV) grant from Alzheimer Association and the A2018425S (JMV), A2018556F (FR) grants from the BrightFocus Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Ibadan individuals. Additionally, we reproduced results using whole-genome sequencing of 273 Alzheimer disease and 275 cognitively intact admixed Puerto Rican individuals. A further comparison was done with SNP imputation from an additional 8,463 Alzheimer disease and 11,365 cognitively intact non-Hispanic White individuals. We identified a significant interaction between the *ApoE* ϵ 4 allele and the SNP rs10423769_A allele, ($\beta = -0.54$, SE = 0.12, p-value = 7.50×10^{-6}) in the discovery data set, and replicated this finding in Ibadan ($\beta = -1.32$, SE = 0.52, p-value = 1.15×10^{-2}) and Puerto Rican ($\beta = -1.27$, SE = 0.64, p-value = 4.91×10^{-2}) individuals. The non-Hispanic Whites analyses showed an interaction trending in the “protective” direction but failing to pass a 0.05 significance threshold ($\beta = -1.51$, SE = 0.84, p-value = 7.26×10^{-2}). The presence of the rs10423769_A allele reduces the odds ratio for Alzheimer disease risk from 7.2 for *ApoE* ϵ 4/ ϵ 4 carriers lacking the A allele to 2.1 for *ApoE* ϵ 4/ ϵ 4 carriers with at least one A allele. This locus is located approximately 2 mB upstream of the *ApoE* locus, in a large cluster of pregnancy specific beta-1 glycoproteins on chromosome 19 and lies within a long noncoding RNA, ENSG00000282943. This study identified a new African-ancestry specific locus that reduces the risk effect of *ApoE* ϵ 4 for developing Alzheimer disease. The mechanism of the interaction with *ApoE* ϵ 4 is not known but suggests a novel mechanism for reducing the risk for ϵ 4 carriers opening the possibility for potential ancestry-specific therapeutic intervention.

Author summary

Strong associations between *ApoE* ϵ 4 and Alzheimer disease risk have been confirmed worldwide, but there is variability in the effect size across populations. African-descent populations have a lower risk from *ApoE* ϵ 4 compared to other populations. Studies in admixed populations have shown that the African ancestral background surrounding the *ApoE* gene reduces the ϵ 4 risk allele effect. Our objective in this study was to identify areas of the genome that interact with *ApoE* ϵ 4 in African ancestry and result in a lower risk for developing Alzheimer disease. In this study we identify a protective locus for the *ApoE* ϵ 4 allele that lowers the risk for African carriers of the *ApoE* ϵ 4 allele to get Alzheimer disease from an odds ratio of 7.2 to 2.1. This protective haplotype has a frequency of 12% in the African ancestry, but only 0.003 in Europeans. This has been replicated in three independent African ancestry datasets and is trending in a much larger European dataset.

Introduction

The apolipoprotein E (*ApoE*) gene (19q13.32) is the strongest genetic risk factor for late-onset Alzheimer disease (AD) and is associated with an earlier age-of-onset [1,2]. Compared to the common ϵ 3 allele, the *ApoE* ϵ 4 allele increases AD risk, while the ϵ 2 allele decreases AD risk (e.g. provides a protective effect) relative to the other two alleles [1–4].

Identifying protective variants against the development of AD has been a key goal of different research groups, including the AD Sequencing Project [5]. The identification of these natural protections may provide insights into disease mechanisms driving AD development as well as potential therapeutic avenues for AD treatment. Indeed, the *ApoE* ϵ 4 allele has a heterogeneous AD risk effect across diverse ancestral populations [3] (Fig 1). The strongest risk effect from *ApoE* ϵ 4 for AD is in East-Asian populations, with the lowest risk from *ApoE* ϵ 4 in

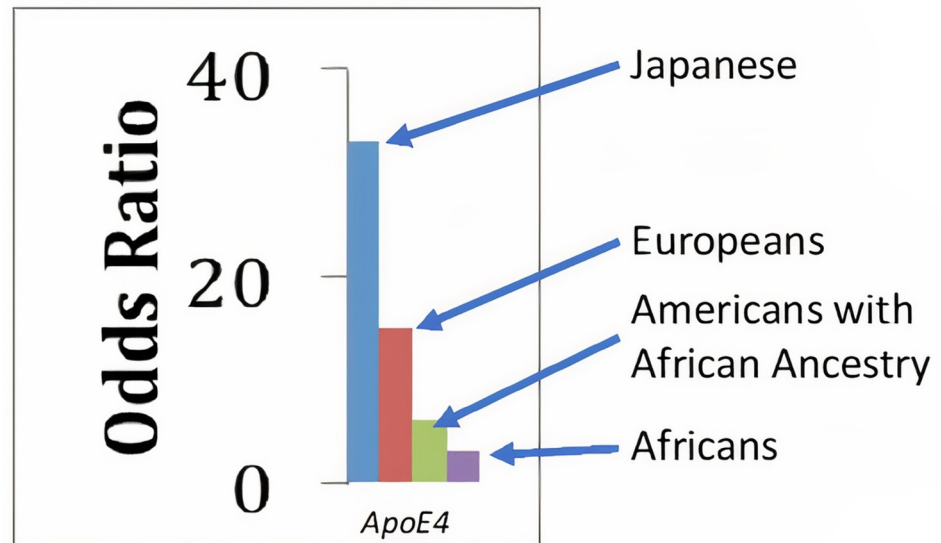


Fig 1. Odds ratios for developing Alzheimer disease according to *ApoE* $\epsilon 4/\epsilon 4$ genotypes carriers relative to the $\epsilon 3/\epsilon 3$ carriers across the multiple ancestries.

<https://doi.org/10.1371/journal.pgen.1009977.g001>

African (AF)-Ancestry populations (such as Ibadan individuals from Nigeria and African Americans (AA)) [3,6–10]. This finding suggested the presence of protective genetic loci that modify AD risk associated with the *ApoE* $\epsilon 4$ allele contributing to this difference in population risk. Using admixed populations with the substantial proportion of AF ancestral genetic background (AA, Puerto Rico (PR) and the Dominican Republic), two independent studies [11,12] demonstrated that the difference in risk between AF and European (EU) populations lies in the ancestral genomic background surrounding the *ApoE* locus (local ancestry, or LA). Specifically, when the *ApoE* $\epsilon 4$ allele lies on an AF-originated haplotype the AD risk is significantly lower than if it lies on EU-originating haplotypes. Simply put, an individual who has inherited their *ApoE* $\epsilon 4$ allele from an AF ancestor has the lower *ApoE* $\epsilon 4$ -associated AD risk observed in AF populations, while an individual who has inherited their *ApoE* $\epsilon 4$ allele from an EU ancestor has the AD risk observed in EU.

Our objective is to follow up the local ancestry finding and identify the genetic variants that lower the risk for *ApoE* $\epsilon 4$ in African ancestry. We have assessed the *ApoE* $\epsilon 4$ haplotypes of both EU and AF local ancestry using several genomic approaches [13,14]. We report here results of a genetic interaction study that identified an AF-specific haplotype that is associated with a substantially reduced risk for AD in African *ApoE* $\epsilon 4$ carriers. This locus lies in the pregnancy specific beta-1 glycoproteins (*PSG*) gene cluster on chromosome 19, approximately two megabases (mB) upstream of the *ApoE* locus.

Results

We identified a locus (rs10423769-allele A) that lies within a large cluster of *PSG* genes on chromosome 19 approximately 2 mB from the *ApoE* gene. This locus has a significant interaction with the *ApoE* $\epsilon 4$ allele ($\beta = -0.54$, $SE = 0.12$, $p\text{-value} = 7.5 \times 10^{-6}$) in the AA samples, meeting an FDR correction threshold for multiple testing ($p\text{-value} = 0.014$). Individuals carrying the minor allele “A” at rs10423769 showed a reduction in AD risk due to *ApoE* $\epsilon 4$. Fig 2 illustrates the logistic regression model results across the *ApoE* region for the interaction term

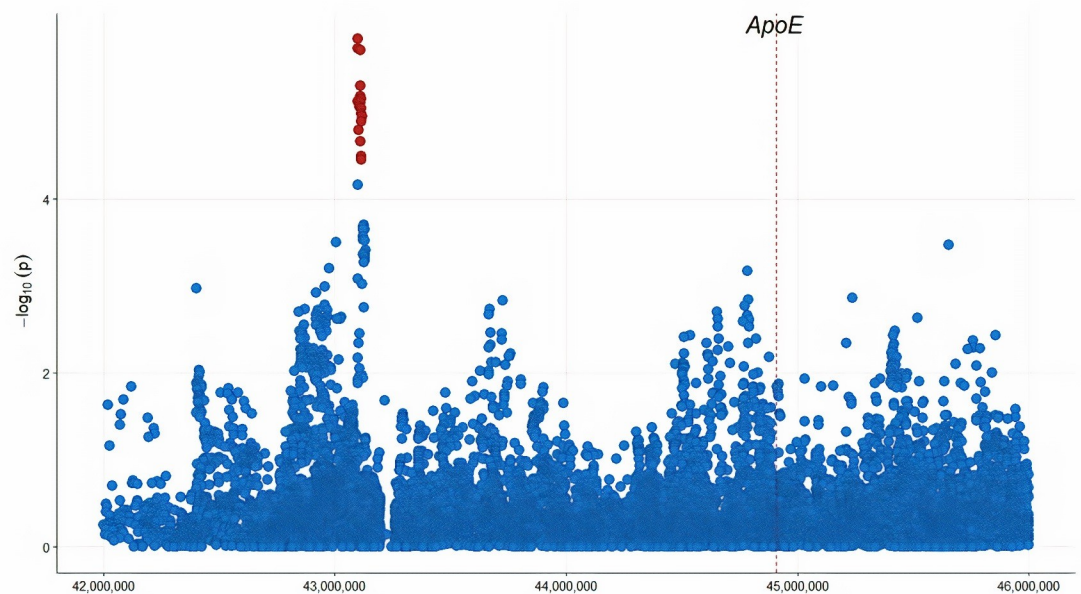


Fig 2. The regional plot of logistic regression analysis across the *ApoE* region for the interaction term. Genomic coordinates are displayed along the X-axis. Negative logarithm of the association p-value for the interaction term is displayed on the Y-axis, with the red color representing markers that have FDR adjusted p-value < 0.05.

<https://doi.org/10.1371/journal.pgen.1009977.g002>

(QQ plot illustrated in S1 Fig). In the replication phase, we performed the epistatic interaction model in an independent cohort of Ibadan individuals from Nigeria. Additionally, we used two diverse datasets to reproduce the effect: a cohort of admixed Hispanic individuals of PR ancestry from the mainland United States and Puerto Rico, and a large collection of non-Hispanic Whites (NHW), primarily from the United States. Results showed a significant interaction between *ApoE* $\epsilon 4$ and rs10423769 in the Ibadan and PR datasets (Ibadan: $\beta = -1.32$, $SE = 0.52$, $p\text{-value} = 1.15 \times 10^{-2}$; PR: $\beta = -1.27$, $SE = 0.64$, $p\text{-value} = 4.91 \times 10^{-2}$). The NHW analyses showed an interaction trending in the “protective” direction but failing to pass a 0.05 significance threshold ($\beta = -1.51$, $SE = 0.84$, $p\text{-value} = 7.26 \times 10^{-2}$) (Table 1). The main effect of the rs10423769 marker was not by itself significantly associated with AD ($p\text{-value} = 0.46$) in the reduced logistic regression model (no interaction term).

Next, we investigated the modifier effect of the *ApoE* $\epsilon 4$ risk allele for AD in individuals with the homozygous AF local ancestry within two subgroups stratified by the rs10423769_A allele. In one subgroup, we included individuals homozygous for the “G” allele and in the other we included individuals with at least one alternative “A” allele. To assess the AD risk effect of *ApoE* genotypes $\epsilon 3/\epsilon 4$ and $\epsilon 4/\epsilon 4$ relative to the $\epsilon 3/\epsilon 3$ genotype, we restricted the

Table 1. Effects of the rs10423769 genotype and its *ApoE* $\epsilon 4$ allele interaction terms in the logistic regression model across studies.

	β	SE	$p\text{-value}$
African American	-0.54	0.12	7.5×10^{-6}
Ibadan	-1.32	0.52	1.2×10^{-2}
Puerto Rican	-1.27	0.64	4.9×10^{-2}
non-Hispanic White	-1.51	0.84	7.3×10^{-2}

<https://doi.org/10.1371/journal.pgen.1009977.t001>

Table 2. Odds Ratios for developing Alzheimer disease according to ApoE genotypes $\epsilon 3/\epsilon 4$ and $\epsilon 4/\epsilon 4$ relative to the $\epsilon 3/\epsilon 3$ stratified by the rs10423769_A allele and studies.

		African American			Ibadan			African American + Ibadan		
		N	OR (95% CI)	p-value	N	OR (95% CI)	p-value	N	OR (95% CI)	p-value
rs10423769_A carriers	$\epsilon 3/\epsilon 3$	300	1(Referent)		92	1(Referent)		392	1(Referent)	
	$\epsilon 3/\epsilon 4$	250	1.06(0.74–1.51)	0.757	60	0.49(0.10–1.91)	0.337	310	1.07(0.86–1.61)	0.681
	$\epsilon 4/\epsilon 4$	55	2.41(1.34–4.35)	0.003	12	...*	...	67	2.06(1.21–3.52)	0.008
rs10423769_A non-carriers	$\epsilon 3/\epsilon 3$	1139	1(Referent)		244	1(Referent)		1383	1(Referent)	
	$\epsilon 3/\epsilon 4$	934	2.61(2.15–3.17)	3.59×10^{-22}	157	2.83(1.40–5.87)	0.004	1091	2.59(2.16–3.12)	2.37×10^{-24}
	$\epsilon 4/\epsilon 4$	187	7.58(5.43–10.69)	8.97×10^{-32}	20	5.02(1.26–17.0)	0.013	207	7.24(5.31–9.94)	3.12×10^{-35}

* $\epsilon 4/\epsilon 4$ genotype data were absent in cases

<https://doi.org/10.1371/journal.pgen.1009977.t002>

sample set to those that were not $\epsilon 2$ allele carriers. Only four PR individuals were identified with the rs10423769_A allele and homozygous AF local ancestry, so we did not include individuals from PR for further analysis. We performed logistic regression analysis within AA, Ibadan populations, and in combined AA and Ibadan individuals separately. Then, we tested the risk effect size differences of $\epsilon 3/\epsilon 4$ and $\epsilon 4/\epsilon 4$ genotypes using two-sample z-test. We found that in the subgroup of rs10423769_A allele carriers the effect sizes of $\epsilon 3/\epsilon 4$ and $\epsilon 4/\epsilon 4$ genotypes were significantly lower than in the non-carriers (AA: $\epsilon 3/\epsilon 4$: p -value = 1.43×10^{-5} ; $\epsilon 4/\epsilon 4$: p -value = 8.79×10^{-4} ; Ibadan: $\epsilon 3/\epsilon 4$: p -value = 0.033; $\epsilon 4/\epsilon 4$ were absent in cases of rs10423769_A allele carriers; AA and Ibadan individuals combined: $\epsilon 3/\epsilon 4$: p -value = 5.70×10^{-6} ; $\epsilon 4/\epsilon 4$: p -value = 7.11×10^{-5}). Odds Ratios for developing AD according to ApoE genotypes stratified by the rs10423769_A allele in AA and Ibadan populations are shown in Table 2. Fig 3 illustrates the AD risk effect of $\epsilon 3/\epsilon 4$ and $\epsilon 4/\epsilon 4$ genotypes relative to the $\epsilon 3/\epsilon 3$ in combined AA and Ibadan individuals across the strata of rs10423769 genotypes.

The AF haplotype associated with the “A” allele for rs10423769 is shown in Fig 4. The haplotype lies 18kb upstream of the PSG2 gene. This haplotype lies within the long noncoding gene ENSG00000282943 (also identified as AC004784.1 and CTC-490G23.6), primarily expressed in the cerebellum and fibroblasts based on data from GTEx [15].

Splicing quantitative trait loci (sQTL) analysis using the GTEx database shows that rs10423769 is a significant sQTL for the TMEM145 gene in the cerebellum with the “A” allele having a 1.6-fold increase in splicing levels between chr19:42320437 and chr19:42320653 relative to the “G” allele (p -value = 2.7×10^{-6}).

We used Hi-C analysis to investigate if the locus represented by rs10423769 directly interacts with ApoE gene locus via cis 3D chromatin looping. As ApoE is primarily expressed in astrocytes and microglia, we performed Hi-C analysis in iPSC-derived astrocytes from ApoE $\epsilon 4$ homozygotes who were homozygous for either African or European local ancestry surrounding ApoE $\epsilon 4$. As shown in Fig 5 PSG2 and ApoE reside in separate topologically associated domains (TADs) and no cis enhancer-promoter loop was detected between the two loci in either local ancestry. However, cerebellum data were not available.

Discussion

This study identified a new AF ancestry-specific haplotype that reduces the AD risk effect of ApoE $\epsilon 4$ homozygotes in AF ancestry by approximately 75%. Previous studies have shown that the African local ancestral background of the ApoE gene reduces the AD risk due to the $\epsilon 4$ allele, with individuals inheriting the ApoE $\epsilon 4$ allele from African ancestors having a lower risk of AD than individuals inheriting the ApoE $\epsilon 4$ allele from European ancestry [11,12]. Our

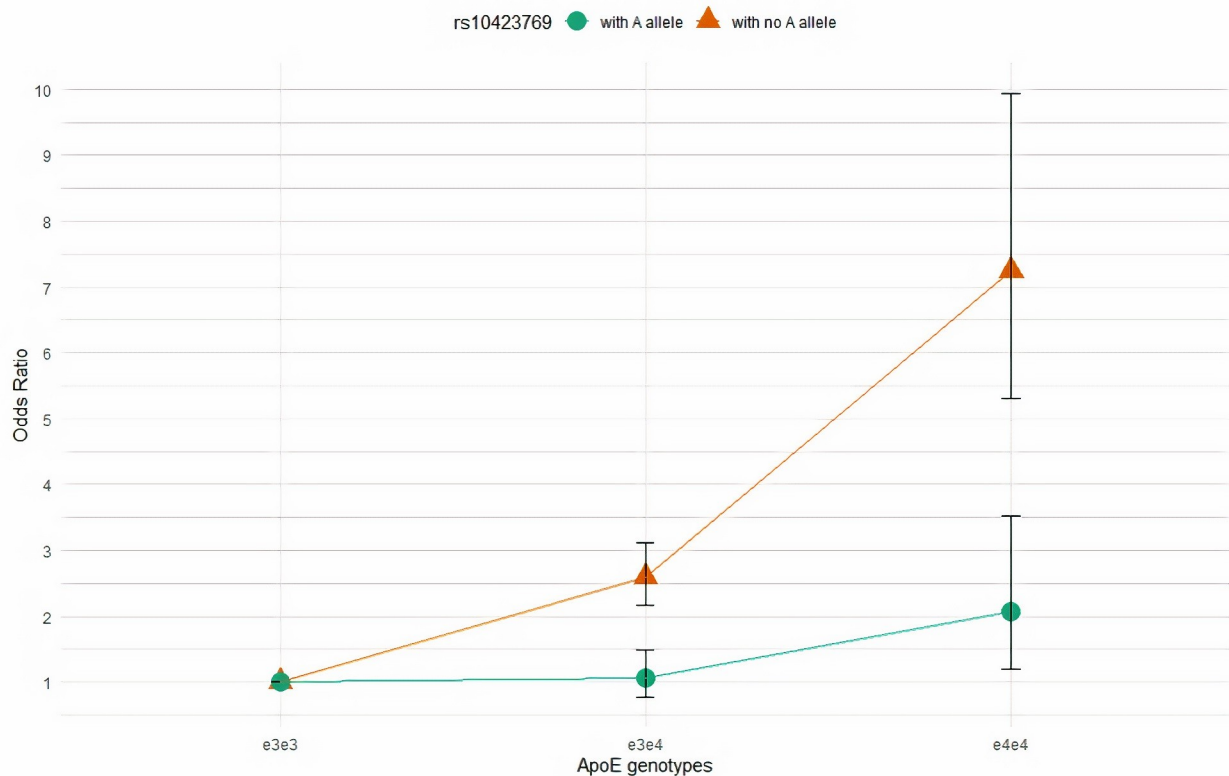


Fig 3. The plot illustrates odds ratios for $\epsilon 3/\epsilon 4$ and $\epsilon 4/\epsilon 4$ genotypes relative to $\epsilon 3/\epsilon 3$ individuals in the combined African American and Ibadan individuals stratified by the rs10423769_A allele. Subgroup odds ratios (95% CIs) are denoted by green color (circle) for rs10423769_A allele carriers and brown color (triangular) for no rs10423769_A allele carriers.

<https://doi.org/10.1371/journal.pgen.1009977.g003>

results corroborate with these findings and identify a novel African locus (19q13.31) that explains a portion of the lower risk due to *ApoE* $\epsilon 4$ in African local ancestry individuals. A recent single-nuclei RNA study showed that European local ancestry carriers had significantly higher *ApoE* expression than African local ancestry carriers, suggesting ancestral-specific regulation of *ApoE* gene expression [13] could be contributing to this risk difference as well. These two findings suggest a polygenic modulation of the $\epsilon 4$ allele risk effect among populations.

Local ancestry blocks across the genome have a wide distribution in size. In our study's largest dataset of AA individuals (3000) the mean of local ancestry block sizes across the genome was ~36Mb (S2 Fig). The local ancestry region in the Rajabli et al. study [11] has an ad-hoc definition of 1 mB on either side of the *ApoE* gene to functionally include the topological associated domains surrounding *ApoE*. The current study expands the Rajabli et al study to a wider genetic region that includes +/-3 mB around the *ApoE* gene, but still correlates with the previously identified effect of the local ancestry associated with differences in risk for AD between *ApoE* $\epsilon 4$ carriers.

Since we detected a statistical interaction between *PSG2* and *ApoE* in individuals with African ancestry, we asked if the two loci have cis interaction with each other that could explain the statistical finding. Towards this end, we constructed a chromatin 3D interaction map in iPSC-derived astrocytes with African local ancestry and European local ancestry surrounding *ApoE* using Hi-C. No evidence of cis interaction was observed in this cell type, which along

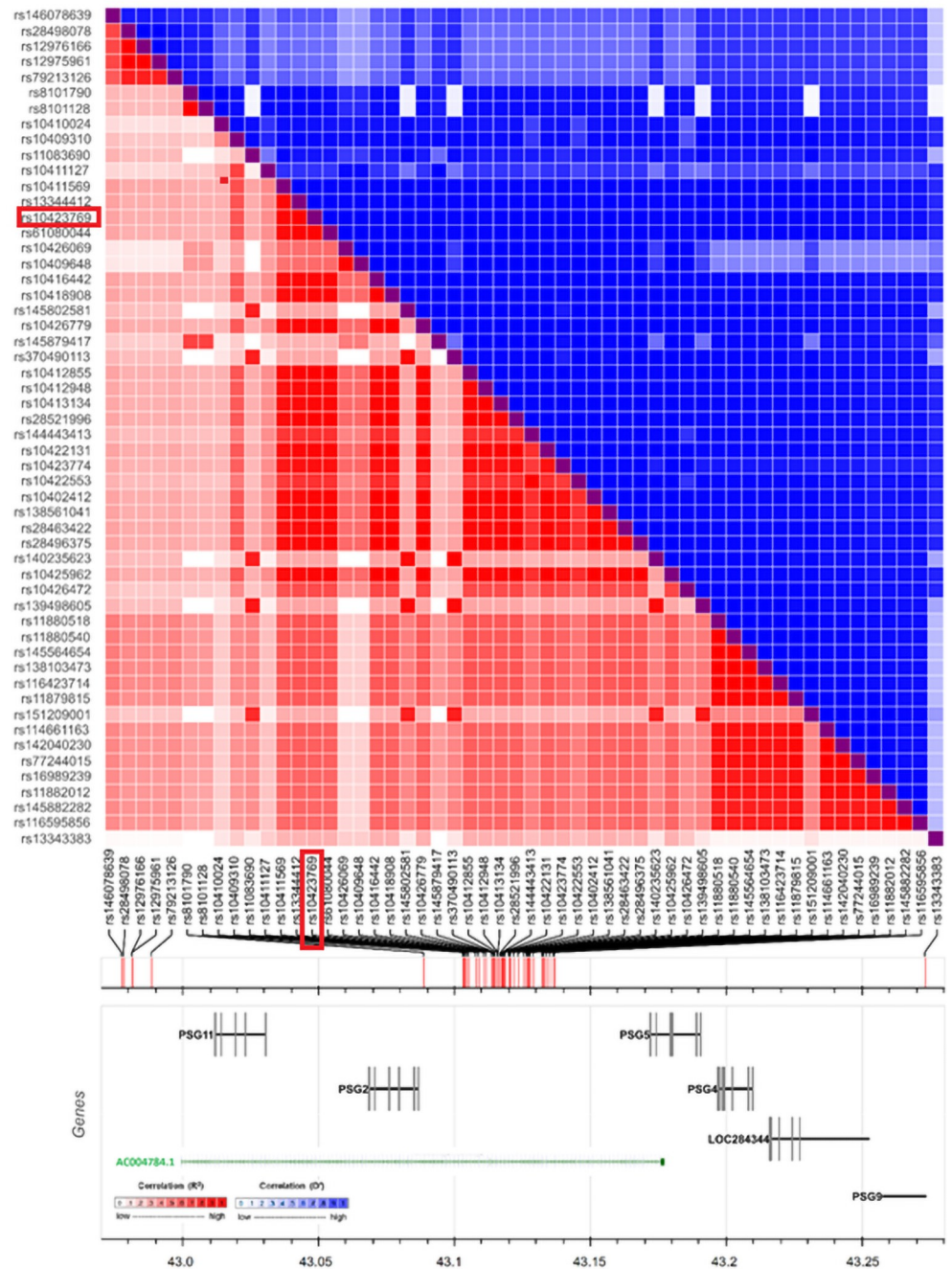


Fig 4. The plot illustrates the Linkage Disequilibrium (LD) heatmap surrounding the rs10423769 variant. The LD matrix was constructed based on r^2 (red color squares) and D' (blue color squares) measurement for all pairs of variants. “Genes” panel illustrates the genes that are overlapping with the LD pattern.

<https://doi.org/10.1371/journal.pgen.1009977.g004>

with microglia, is the major cell that expresses *ApoE*. As most enhancers interact within 1 mB [16], it is unlikely that other cell types in African ancestry would contain an interaction at the distance separating rs10423769 and *ApoE* is (~2 mB) but can’t be completely ruled out.

Indeed, the distance between the two interacting loci suggests that other mechanisms than enhancer-promoter maybe involved in this protective effect. The protective haplotype overlies

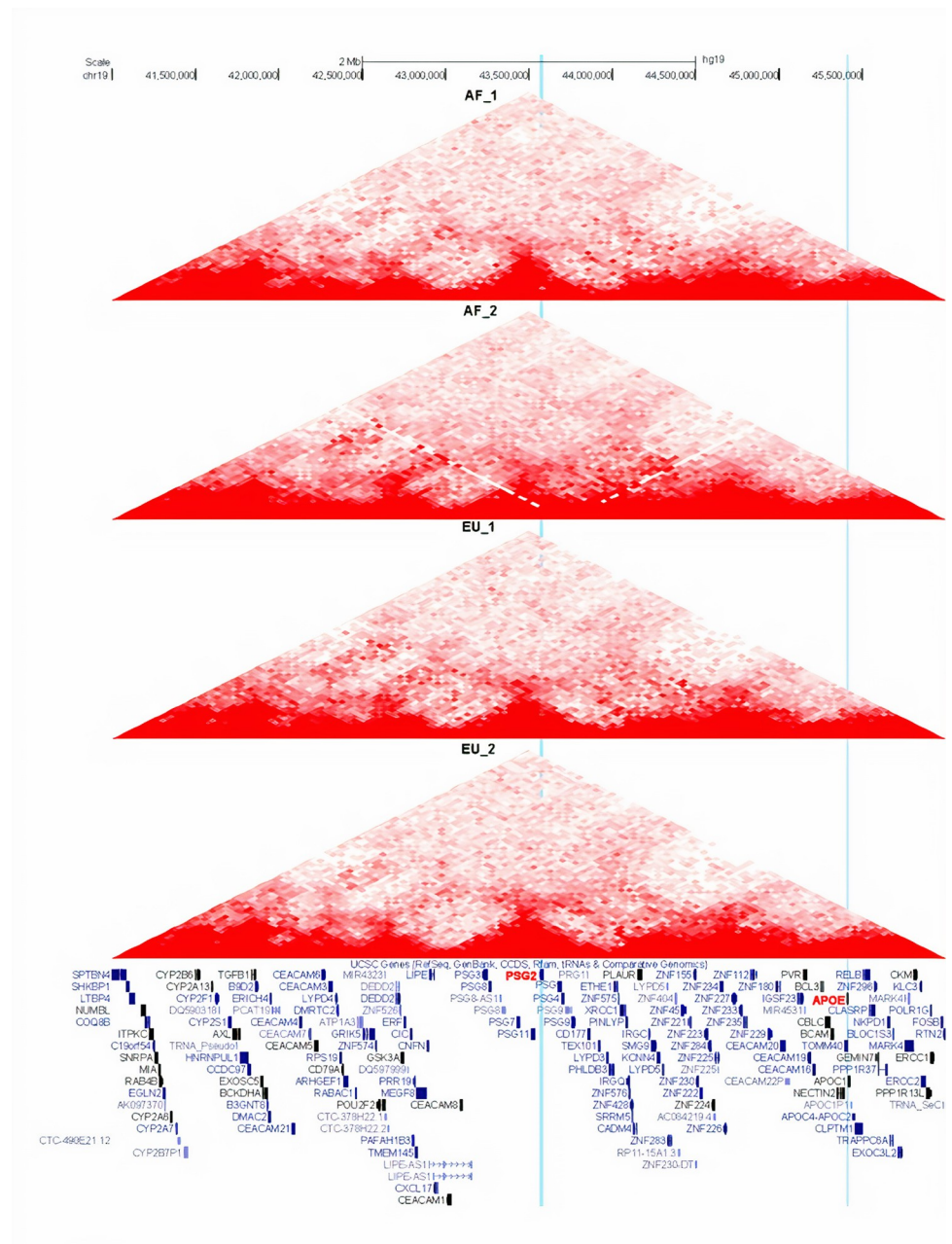


Fig 5. Chromatin Loops in chromosome 19 region centering on *PSG2* and *ApoE* genes are displayed in four tracks, each representing one astrocyte line. Induced pluripotent stem cells (iPSCs) with African (20–1611720 and LWHiC, top two tracks) or non-Hispanic White local ancestry (20–1616981 and 201–806023, bottom two tracks) were differentiated into astrocytes. Chromatin Hi-C library was constructed with 4-cutter enzyme. Blue vertical bar highlights rs10423769; yellow vertical bar highlights *ApoE*.

<https://doi.org/10.1371/journal.pgen.1009977.g005>

the long noncoding RNA ENSG00000282943, but little is known about its specific function. Interestingly, rs10423769 is reported to be a sQTL for *TMEM145*, which has not been implicated in AD, although it has been reported to be upregulated in anterior cingulate cortex in Dementia with Lewy body patients [17]. Interestingly, while both loci have low expression throughout the brain, their highest brain expression is in the cerebellum, particularly for

TMEM145. The cerebellum has had observed changes in AD, especially in early onset forms of the disease. But it has received little investigation in the pathophysiology of AD, as historically the cerebellum has been studied for its role in the regulation of motor activity. However, recent studies have shown a role for the cerebellum in cognition, including roles in working memory and executive and visuospatial functions [18–20].

The protective haplotype also lies within a large cluster of *PSG* genes, a family of glycoproteins that are primarily synthesized in the syncytiotrophoblast of the human placenta. Several case-control studies have shown that low levels of *PSG2* are associated with pre-eclampsia [21], which, in turn, has been suggested to be associated with an increased risk of dementia later in life [22]. Indeed, a recent study demonstrated that inducible pluripotent stem cell (iPSC) derived neurons made from the blood of autopsy confirmed AD patients, had abnormal tau deposition which matched their autopsy findings [23]. As these are very young neurons, it raises the possibility that the processes that evidentially lead to AD could begin very early in life. Thus, although they appear temporally distant from the onset of clinical AD in late life, their involvement cannot be entirely ruled out.

This finding is the first AF specific protective effect in AD and highlights the importance of diversity and the inclusion of all populations into research. It hopefully will encourage additional studies focused on diverse populations, where allelic frequency differences can discover information that is hidden when studying only a single population. With most clinical trials in AD ending in failure, the identification of natural protective interactions is of key importance in moving therapeutic efforts in AD forward for all ancestries. Finally, as *ApoE* ϵ 4 is the major genetic risk factor, this work further supports growing efforts to explore it as a major therapeutic target for AD.

Materials and methods

Ethics statement

Written consent was obtained from all participants and study protocols were approved by the University of Miami Institutional Review Board (IRB), the Indiana University IRB and the IRB of the University of Pennsylvania.

Study samples

Our study consisted of a discovery phase using TOPMed imputed genotype data from AA individuals in nine datasets [24] and a replication phase with whole genome sequencing (WGS) data from Ibadan individuals from Nigeria [25]. Additionally, we used PR individuals [26], as well as TOPMed imputed genotype data from non-Hispanic/Latino White (NHW) individuals in 31 datasets [27] to reproduce the findings in diverse datasets. The characteristics of each post-QC datasets are shown in Table 3. The detailed description of the datasets is provided in S1 Appendix and elsewhere [24–28].

Genotyping and sequencing. Genome-wide single-nucleotide polymorphism (SNP) genotyping was processed on multiple different platforms and *ApoE* genotyping was performed in the individual datasets as summarized in the S1 Table, S1 Methods and elsewhere [24–28].

Samples from Ibadan and PR had WGS performed at the Uniformed Services University of the Health Sciences (USUHS) using standard protocols as previously described [29]. Illumina's HiSeq Alignment Software (HAS) was used to analyze the data including alignment to the GRCh38 reference genome with the Issac aligner [30] and variant calling with Strelka [31]. Illumina's gvcfgenotyper was used to merge the resulting gvcfs into a cohort level vcf. Variant calls for the positions used in the replication phase had sequencing depth of coverage greater

Table 3. Characteristics of African American, Ibadan, Puerto Rican, and non-Hispanic White data sets.

Characteristic	African American	Ibadan	Puerto Rican	non-Hispanic White
Individuals, No.				
Cases	1850	63	273	8463
Controls	4331	648	275	11365
Women, No. (%)				
Cases	1290 (69.7)	52 (82.5)	181 (66.3)	4710 (55.7)
Controls	3120 (72.0)	401 (61.9)	213 (77.5)	6678 (58.8)
Age, mean (SD)				
Cases ^a	78.6(8.1)	83.5(5.2)	76.5(8.3)	75.9(6.9)
Controls ^b	75.9(8.4)	81.1(4.5)	75.4(6.5)	77.5(7.3)
ApoE ϵ 4 frequencies, No. (%)				
-/- ^c	3553 (57.5)	432 (60.8)	383 (69.9)	12218 (61.6)
-/ ϵ 4	2280 (36.9)	246 (34.6)	144 (26.3)	6627 (33.4)
ϵ 4/ ϵ 4	348 (5.6)	33 (4.6)	23 (3.8)	983 (5.0)

^a Age on onset^b Age at last evaluation^c Containing genotypes ϵ 2/ ϵ 2, ϵ 2/ ϵ 3 and ϵ 3/ ϵ 3<https://doi.org/10.1371/journal.pgen.1009977.t003>

than 15X and an alternate allele fraction between 35% and 65% for heterozygotes and >95% for homozygotes.

Standard quality control (QC) for genotype and individual-level data were performed for each dataset using software PLINK v.2 [32]. Variants with the call score less than 98%, or not in Hardy-Weinberg equilibrium (HWE) ($p < 1.e-6$) were eliminated. Individuals with genotyping call rates less than 97% were removed. Individuals whose reported sex differed from the genotype-inferred sex by analysis of the X-chromosome SNPs were excluded. The relatedness among the individuals within and across the case/control datasets was identified by the estimated proportion of alleles (π) shared identical by descent (IBD), and one individual from relatives ($\pi > 0.4$) was included for the analysis. Population substructure was evaluated in each cohort separately using EIGENSTRAT software [33]. Population substructure in each data sets were compared with those in the 1000 Genome reference panel YRI (Yoruba from Nigeria) and CEU (Utah Residents with Northern and Western European ancestry) populations. Outliers with respect to CEU population (overlapping within the cluster of CEU) were removed from the datasets [34]. S3 Fig illustrates principal component analysis for each African American dataset.

Genotype imputation. We imputed AA and NHW genotype array datasets individually using the TOPMed R2 version panel (build of GRCh38) and TOPMed Imputation server [35]. The TOPMed R2 reference panel has 97,256 samples and provides information on 308,107,085 genetic variants [36]. Most of the samples in the TOPMed panel are non-EU and around 25% of the samples are from AF-descent populations. We kept the high-quality common variants ($R^2 > 0.8$) with a minor allele frequency (MAF) > 0.05 in the existing AA datasets.

Assessment of genetic ancestry. We calculated global ancestry (principal components (PC)) within each array dataset using the EIGENSTRAT approach using EIGENSOFT software with no reference population [33]. To estimate the local ancestry in AA and PR datasets, we first combined each of the array datasets with the Human Genome Diversity Project (HGDP) reference panel separately using PLINK v2 software [32,37]. We used 98 AF, 109 EU,

and 108 Amerindian individuals from HGDP reference populations. Then, we phased combined datasets using the SHAPEIT tool ver. 2 with default settings and the 1000 Genomes Phase 3 reference panel [34,38]. Finally, we inferred the local ancestries using the discriminative modeling approach implemented in RFMix with the PopPhased option and a minimum node size of 5 [39].

Statistical analysis

Identifying protective loci. To identify protective loci that modify the *ApoE* $\epsilon 4$ risk effect, we performed an interaction analysis in our discovery AA datasets, focusing on the broad genetic region that includes +/-3 mB around the *ApoE* gene. Imputed data were force-called to the most likely genotype (with 0.90 threshold for the probability) and then assessed using a logistic regression approach. Our primary model for the logistic regression included AD status as the outcome (dependent) variable. Independent variables included force-called genotype and *ApoE* $\epsilon 4$ main effects along with an interaction term between genotype and *ApoE* $\epsilon 4$. Additional covariates included age, sex, and genome-wide ancestry (PC1:3) (Eq 1). We coded both the variants and *ApoE* $\epsilon 4$ allele under a dosage model (0,1,2) and performed interaction analysis in each imputed AA datasets separately. Then, we meta-analyzed all terms across AA datasets by applying fixed-effect meta-analysis (assuming similar effect sizes) from META-SOFT software [40]. We used the Benjamini and Hochberg approach to control for the false discovery rate (FDR) [41]. Subsequently, replication analyses were performed using WGS datasets on Ibadan and PR individuals and meta-analysis in 31 imputed genotype datasets from NHWs.

$$AD \sim Age + Sex + PC1 : 3 + \epsilon 4 + Variant + \epsilon 4 \times Variant \quad 1$$

Assessing modifiers. To assess the influence of putative modifiers on *ApoE* $\epsilon 4$ risk effect, we compared the AD risk effect of *ApoE* $\epsilon 3/\epsilon 4$ and $\epsilon 4/\epsilon 4$ genotypes relative to the $\epsilon 3/\epsilon 3$ genotype in those that were carriers of the modifier alleles to those that were not carriers. First, we restricted the sample set to those that were homozygous for the AF genetic ancestry around the *ApoE* locus and were not *ApoE* $\epsilon 2$ allele carriers. We then stratified by carrier status at loci identified as FDR significant. Next, we performed logistic regression with *ApoE* genotypes ($\epsilon 3/\epsilon 3$ (reference), $\epsilon 3/\epsilon 4$, $\epsilon 4/\epsilon 4$), age, sex, principal components (PC1-3), and batch as covariates within each group (carriers and non-carriers of the modifier allele) for each study and across studies. Finally, we tested the difference in effect sizes of the *ApoE* genotypes between carriers and non-carriers of the putative modifiers using two-sample z-test. Statistical analyses were performed using the “GLM2” package available in R computing environment [42].

Hi-C analysis. Hi-C analysis was performed on astrocytes derived from induced pluripotent stem cells (iPSCs) derived from AD patients who were *ApoE* $\epsilon 4/\epsilon 4$ and had AF ancestry. Cells were differentiated and cultured using the StemDiff Astrocyte Differentiation and Maturation kits (StemCell Technologies) according to the manufacturer’s protocol. *In situ* Hi-C libraries were prepared using the protocol adapted from Rao et al. [43]. For each library, 450–550 million paired-end reads at 150 bp length were obtained. Sequencing data were processed using BWA to map each read end separately to GRCh38 reference genome [44]. Duplicate and non-uniquely mapped reads were removed. For each library, over 270 million of non-redundant, uniquely mapped, paired reads were used for further analysis. Contact matrices were generated at base pair delimited resolutions of 50 kb [45].

Supporting information

S1 Appendix. Description of datasets.

(DOCX)

S1 Methods. Diagnosis of AD and age of onset and Genotyping.

(DOCX)

S1 Table. Genotyping platforms used in individual datasets.

(DOCX)

S2 Table. Effects of the rs10423769 genotype and ApoE ϵ 4 allele interaction term in the logistic regression model for individual African American datasets.

(DOCX)

S1 Fig. QQ plot and the lambda value of the interaction term in the primary logistic regression model.

(DOCX)

S2 Fig. Distribution of admixture block sizes in ~3,000 African American individuals.

Admixture block sizes are displayed along the x-axis in \log_{10} scale.

(DOCX)

S3 Fig. Principal component analysis for each African American dataset combined with 1000 Genome CEU and YRI reference populations.

(DOCX)

Author Contributions

Conceptualization: Farid Rajabli, Margaret A. Pericak-Vance, Jeffery M. Vance.

Data curation: Margaret A. Pericak-Vance.

Formal analysis: Farid Rajabli, Nicholas A. Kushch, Marina Lipkin-Vasquez, Kara L. Hamilton-Nelson, Karen Nuytemans, Liyong Wang, Anthony J. Griswold.

Funding acquisition: Farid Rajabli, Hugh C. Hendrie, Gerard D. Schellenberg, Goldie S. Byrd, Jonathan L. Haines, Margaret A. Pericak-Vance, Jeffery M. Vance.

Investigation: Farid Rajabli, Jeffery M. Vance.

Methodology: Farid Rajabli, Gary W. Beecham, Liyong Wang, Fulai Jin, Xiaoxiao Liu, Jeffery M. Vance.

Project administration: Margaret A. Pericak-Vance, Jeffery M. Vance.

Resources: Margaret A. Pericak-Vance, Jeffery M. Vance.

Supervision: Gary W. Beecham, Juan I. Young, Jeffery M. Vance.

Validation: Farid Rajabli, Gary W. Beecham, Jeffery M. Vance.

Visualization: Farid Rajabli, Liyong Wang.

Writing – original draft: Farid Rajabli, Liyong Wang, Jeffery M. Vance.

Writing – review & editing: Farid Rajabli, Gary W. Beecham, Hugh C. Hendrie, Olusegun Baiyewu, Adesola Ogunniyi, Sujuan Gao, Nicholas A. Kushch, Marina Lipkin-Vasquez, Kara L. Hamilton-Nelson, Juan I. Young, Derek M. Dykxhoorn, Karen Nuytemans, Brian W. Kunkle, Liyong Wang, Fulai Jin, Xiaoxiao Liu, Briseida E. Feliciano-Astacio, Gerard D.

Schellenberg, Clifton L. Dalgard, Anthony J. Griswold, Goldie S. Byrd, Christiane Reitz, Michael L. Cuccaro, Jonathan L. Haines, Margaret A. Pericak-Vance, Jeffery M. Vance.

References

1. Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science*. 1993 Aug 13; 261(5123):921–3. <https://doi.org/10.1126/science.8346443> PMID: 8346443.
2. Corder EH, Saunders AM, Risch NJ, Strittmatter WJ, Schmechel DE, Gaskell PC Jr, et al. Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. *Nat Genet*. 1994 Jun; 7(2):180–4. <https://doi.org/10.1038/ng0694-180> PMID: 7920638.
3. Farrer LA, Cupples LA, Haines JL, Hyman B, Kukull WA, Mayeux R, et al. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. *JAMA*. 1997 Oct 22–29; 278(16):1349–56. PMID: 9343467.
4. Reiman EM, Arboleda-Velasquez JF, Quiroz YT, Huentelman MJ, Beach TG, Caselli RJ, et al. Alzheimer's Disease Genetics Consortium. Exceptionally low likelihood of Alzheimer's dementia in APOE2 homozygotes from a 5,000-person neuropathological study. *Nat Commun*. 2020 Feb 3; 11(1):667. <https://doi.org/10.1038/s41467-019-14279-8> PMID: 32015339; PMCID: PMC6997393.
5. Beecham GW, Bis JC, Martin ER, Slifer MA, Gilbert JR, Haines JL, et al. The Alzheimer's Disease Sequencing Project: Study design and sample selection. *Neurol Genet*. 2017 Oct 13; 3(5):e194. <https://doi.org/10.1212/NXG.000000000000194> PMID: 29184913; PMCID: PMC5646177.
6. Tang MX, Cross P, Andrews H, Jacobs DM, Small S, Bell K, et al. Incidence of AD in African-Americans, Caribbean Hispanics, and Caucasians in northern Manhattan. *Neurology*. 2001 Jan 9; 56(1):49–56. <https://doi.org/10.1212/wnl.56.1.49> PMID: 11148235.
7. Tang MX, Maestre G, Tsai WY, Liu XH, Feng L, Chung WY, et al. Relative risk of Alzheimer disease and age-at-onset distributions, based on APOE genotypes among elderly African Americans, Caucasians, and Hispanics in New York City. *Am J Hum Genet*. 1996 Mar; 58(3):574–84. PMID: 8644717; PMCID: PMC1914582.
8. Tang MX, Stern Y, Marder K, Bell K, Gurland B, Lantigua R, et al. The APOE-epsilon4 allele and the risk of Alzheimer disease among African Americans, whites, and Hispanics. *JAMA*. 1998 Mar 11; 279(10):751–5. <https://doi.org/10.1001/jama.279.10.751> PMID: 9508150.
9. Sahota A, Yang M, Gao S, Hui SL, Baiyewu O, Gureje O, et al. Apolipoprotein E-associated risk for Alzheimer's disease in the African-American population is genotype dependent. *Ann Neurol*. 1997 Oct; 42(4):659–61. <https://doi.org/10.1002/ana.410420418> PMID: 9382479.
10. Hendrie HC, Murrell J, Baiyewu O, Lane KA, Purnell C, Ogunniyi A, et al. APOE ϵ 4 and the risk for Alzheimer disease and cognitive decline in African Americans and Yoruba. *Int Psychogeriatr*. 2014 Jun; 26(6):977–85. <https://doi.org/10.1017/S1041610214000167> Epub 2014 Feb 24. PMID: 24565289; PMCID: PMC4012422.
11. Rajabli F, Feliciano BE, Celis K, Hamilton-Nelson KL, Whitehead PL, Adams LD, et al. Ancestral origin of ApoE ϵ 4 Alzheimer disease risk in Puerto Rican and African American populations. *PLoS Genet*. 2018 Dec 5; 14(12):e1007791. <https://doi.org/10.1371/journal.pgen.1007791> PMID: 30517106; PMCID: PMC6281216.
12. Blue EE, Horimoto ARVR, k S, Wijsman EM, Thornton TA. Local ancestry at APOE modifies Alzheimer's disease risk in Caribbean Hispanics. *Alzheimers Dement*. 2019 Dec; 15(12):1524–1532. <https://doi.org/10.1016/j.jalz.2019.07.016> Epub 2019 Oct 9. PMID: 31606368; PMCID: PMC6925639.
13. Griswold AJ, Celis K, Bussies PL, Rajabli F, Whitehead PL, Hamilton-Nelson KL, et al. Increased APOE ϵ 4 expression is associated with the difference in Alzheimer's disease risk from diverse ancestral backgrounds. *Alzheimers Dement*. 2021 Jul; 17(7):1179–1188. <https://doi.org/10.1002/alz.12287> Epub 2021 Feb 1. PMID: 33522086.
14. Nuytemans K, Lipkin M, Wang L, Van Booven D, Griswold AJ, Rajabli F, et al. Converging evidence for differential regulatory control of APOE ϵ 4 on African versus European haplotypes. *BioRxiv*. 2021. <https://doi.org/10.1101/2021.08.23.457375>.
15. Carithers LJ, Ardlie K, Barcus M, Branton PA, Britton A, Buia SA, et al. A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. *Biopreserv Biobank*. 2015 Oct; 13(5):311–9. <https://doi.org/10.1089/bio.2015.0032> PMID: 26484571
16. Andersson R, Sandelin A. Determinants of enhancer and promoter activities of regulatory elements. *Nature Reviews Genetics*, 2020; 21(2), 71–87. <https://doi.org/10.1038/s41576-019-0173-8> PMID: 31605096

17. Pietrzak M, Papp A, Curtis A, et al. Gene expression profiling of brain samples from patients with Lewy body dementia. *Biochem Biophys Res Commun*. 2016 Oct 28; 479(4):875–880. <https://doi.org/10.1016/j.bbrc.2016.09.114> Epub 2016 Sep 22. PMID: 27666482; PMCID: PMC5079284.
18. Schmahmann JD, Guell X, Stoodley CJ, Halko MA. The Theory and Neuroscience of Cerebellar Cognition. *Annu Rev Neurosci*. 2019 Jul 8; 42:337–364. <https://doi.org/10.1146/annurev-neuro-070918-050258> Epub 2019 Apr 2. PMID: 30939101.
19. Zhou Z, Zhu R, Shao W, et al. Changes in Resting-State Functional Connectivity of Cerebellum in Amnesic Mild Cognitive Impairment and Alzheimer's Disease: A Case-Control Study. *Front Syst Neurosci*. 2021 Mar 10; 15:596221. <https://doi.org/10.3389/fnsys.2021.596221> Erratum in: *Front Syst Neurosci*. 2021 May 12;15:693951. PMID: 33790747; PMCID: PMC8006280.
20. Sepulveda-Falla D, Matschke J, Bernreuther C, et al. Deposition of hyperphosphorylated tau in cerebellum of PS1 E280A Alzheimer's disease. *Brain Pathol*. 2011 Jul; 21(4):452–63. <https://doi.org/10.1111/j.1750-3639.2010.00469.x> Epub 2011 Jan 27. PMID: 21159009; PMCID: PMC8094246.
21. Temur M, Serpim G, Tuzluoğlu S, Taşgöz FN, Şahin E, Üstünyurt E. Comparison of serum human pregnancy-specific beta-1-glycoprotein 1 levels in pregnant women with or without preeclampsia. *J Obstet Gynaecol*. 2020 Nov; 40(8):1074–1078. <https://doi.org/10.1080/01443615.2019.1679734> Epub 2019 Dec 2. PMID: 31790616.
22. Basit S, Wohlfahrt J, Boyd HA. Pre-eclampsia and risk of dementia later in life: nationwide cohort study. *BMJ*. 2018 Oct 17; 363:k4109. <https://doi.org/10.1136/bmj.k4109> PMID: 30333106; PMCID: PMC6191824.
23. Lagomarsino VN, Pearse RV 2nd, Liu L, et al. Stem cell-derived neurons reflect features of protein networks, neuropathology, and cognitive outcome of their aged human donors. *Neuron*. 2021 Aug 26; S0896–6273(21)00578-X. <https://doi.org/10.1016/j.neuron.2021.08.003> Epub ahead of print. PMID: 34473944.
24. Kunkle BW, Schmidt M, Klein HU, Naj AC, Hamilton-Nelson KL, Larson EB, et al. Novel Alzheimer Disease Risk Loci and Pathways in African American Individuals Using the African Genome Resources Panel: A Meta-analysis. *JAMA Neurol*. 2021 Jan 1; 78(1):102–113. <https://doi.org/10.1001/jamaneurol.2020.3536> Erratum in: *Neurol JAMA*. 2021 May 1;78(5):620. PMID: 33074286; PMCID: PMC7573798.
25. Ogunniyi A, Baiyewu O, Gureje O, Hall KS, Unverzagt F, Siu SH, et al. Epidemiology of dementia in Nigeria: results from the Indianapolis-Ibadan study. *Eur J Neurol*. 2000 Sep; 7(5):485–90. <https://doi.org/10.1046/j.1468-1331.2000.00124.x> PMID: 11054131.
26. Feliciano-Astacio BE, Celis K, Ramos J, et al. The Puerto Rico Alzheimer Disease Initiative (PRADI): A Multisource Ascertainment Approach. *Front Genet*. 2019 Jun 19; 10:538. <https://doi.org/10.3389/fgene.2019.00538> PMID: 31275353; PMCID: PMC6593074.
27. Kunkle BW, Grenier-Boley B, Sims R, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates A β , tau, immunity and lipid processing. *Nat Genet*. 2019 Mar; 51(3):414–430. <https://doi.org/10.1038/s41588-019-0358-2> Epub 2019 Feb 28. Erratum in: *Genet Nat*. 2019 Sep;51(9):1423–1424. PMID: 30820047; PMCID: PMC6463297.
28. Kunkle BW, Carney RM, Kohli MA, Naj AC, Hamilton-Nelson KL, Whitehead PL, et al. Targeted sequencing of ABCA7 identifies splicing, stop-gain and intronic risk variants for Alzheimer disease. *Neurosci Lett*. 2017 May 10; 649:124–129. <https://doi.org/10.1016/j.neulet.2017.04.014> Epub 2017 Apr 8. PMID: 28400126.
29. Rajabli F, Feliciano-Astacio BE, Cukier HN, et al. Linkage of Alzheimer disease families with Puerto Rican ancestry identifies a chromosome 9 locus. *Neurobiol Aging*. 2021 Aug; 104:115.e1–115.e7. <https://doi.org/10.1016/j.neurobiolaging.2021.02.019> Epub 2021 Feb 28. PMID: 33902942.
30. Raczky C, Petrovski R, Saunders CT, Chorny I, Kruglyak S, Margulies EH, et al. Isaac: ultra-fast whole-genome secondary analysis on Illumina sequencing platforms. *Bioinformatics*. 2013 Aug 15; 29(16):2041–3. <https://doi.org/10.1093/bioinformatics/btt314> Epub 2013 Jun 4. PMID: 23736529.
31. Kim S, Scheffler K, Halpern AL, Bekritsky MA, Noh E, Källberg M, et al. Strelka2: fast and accurate calling of germline and somatic variants. *Nat Methods*. 2018 Aug; 15(8):591–594. <https://doi.org/10.1038/s41592-018-0051-x> Epub 2018 Jul 16. PMID: 30013048.
32. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015 Feb 25; 4:7. <https://doi.org/10.1186/s13742-015-0047-8> PMID: 25722852; PMCID: PMC4342193.
33. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006 Aug; 38(8):904–9. <https://doi.org/10.1038/ng1847> Epub 2006 Jul 23. PMID: 16862161.
34. 1000 Genomes Project Consortium, Auton A, Brooks LD, u RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. *Nature*. 2015 Oct 1; 526(7571):68–74. <https://doi.org/10.1038/nature15393> PMID: 26432245; PMCID: PMC4750478.

35. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. *Nat Genet.* 2016 Oct; 48(10):1284–1287. <https://doi.org/10.1038/ng.3656> Epub 2016 Aug 29. PMID: 27571263; PMCID: PMC5157836.
36. Taliun D, Harris DN, Kessler MD, Carlson J, Szpiech ZA, Torres R, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature.* 2021 Feb; 590(7845):290–299. <https://doi.org/10.1038/s41586-021-03205-y> Epub 2021 Feb 10. PMID: 33568819; PMCID: PMC7875770.
37. Cavalli-Sforza LL. Human evolution and its relevance for genetic epidemiology. *Annu Rev Genomics Hum Genet.* 2007; 8:1–15. <https://doi.org/10.1146/annurev.genom.8.080706.092403> PMID: 17408354.
38. Delaneau O, Marchini J; 1000 Genomes Project Consortium; 1000 Genomes Project Consortium. Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. *Nat Commun.* 2014 Jun 13; 5:3934. <https://doi.org/10.1038/ncomms4934> PMID: 25653097; PMCID: PMC4338501.
39. Maples BK, Gravel S, Kenny EE, Bustamante CD. RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. *Am J Hum Genet.* 2013 Aug 8; 93(2):278–88. <https://doi.org/10.1016/j.ajhg.2013.06.020> Epub 2013 Aug 1. PMID: 23910464; PMCID: PMC3738819.
40. Han B, Eskin E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am J Hum Genet.* 2011 May 13; 88(5):586–98. <https://doi.org/10.1016/j.ajhg.2011.04.014> PMID: 21565292; PMCID: PMC3146723.
41. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple hypothesis testing. 1995. *J R Stat Soc B* 57:289–300
42. Marschner I. glm2: Fitting Generalized Linear Models. R package version 1.1.2.2014; <https://CRAN.R-project.org/package=glm2>
43. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell.* 2014; 159(7):1665–80. <https://doi.org/10.1016/j.cell.2014.11.021> Epub 2014 Dec 11. Erratum in: *Cell.* 2015 Jul 30;162(3):687–8. PMID: 25497547
44. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2010 Mar 1; 26(5):589–95. <https://doi.org/10.1093/bioinformatics/btp698> Epub 2010 Jan 15. PMID: 20080505; PMCID: PMC2828108.
45. Lu L, Liu X, Huang WK, Giusti-Rodríguez P, Cui J, Zhang S, et al. Robust Hi-C Maps of Enhancer-Promoter Interactions Reveal the Function of Non-coding Genome in Neural Development and Diseases. *Mol Cell.* 2020 Aug 6; 79(3):521–534.e15. <https://doi.org/10.1016/j.molcel.2020.06.007> PMID: 32592681