Association of Rare *APOE* Missense Variants V236E and R251G With Risk of Alzheimer Disease

Authors: Yann Le Guen , PhD [1,2,*,#], Michael E. Belloy , PhD [1,*], Benjamin Grenier-Boley , Msc [3], Itziar de Rojas , MSc [4,5], Atahualpa Castillo-Morales , PhD [6], Iris Jansen , PhD [7,8], Aude Nicolas , PhD [3], Céline Bellenguez , PhD [3], Carolina Dalmasso , PhD [9,10], Fahri Küçükali , PhD [11,12,13], Sarah J. Eger [1], Katrine Laura Rasmussen , MD,PhD [14,15], Jesper Qvist Thomassen , PhD [14], Jean-François Deleuze , PhD [16], Zihuai He , PhD [1,17], Valerio Napolioni , PhD [18], Philippe Amouyel , MD,PhD [19], Frank Jessen , MD [20,21,22], Patrick G. Kehoe , PhD [23], Cornelia Van Duin , PhD [24,25], Magda Tsolaki , MD,PhD [26], Pascual Sánchez-Juan , MD,PhD [27,5], Kristel Sleegers , MD,PhD [11,12,13], Martin Ingelsson , MD,PhD [28,29,30], Giacomina Rossi , PhD [31], Mikko Hiltunen , PhD [32], Rebecca Sims , PhD [33], Wiesje M. van der Flier , PhD [7], Alfredo Ramirez , MD,PhD [9,34,35,21,22], Ole A. Andreassen , MD,PhD [36,37], Ruth Frikke-Schmidt , MD,PhD [14,15], Julie Williams , PhD [6,33], Agustín Ruiz , MD,PhD [4,5], Jean-Charles Lambert , PhD [3], Michael D. Greicius , MD [1], and the EADB group†, the GR@ACE group†, the DEGESCO group†, the EADI group†, the GERAD group†, for the EADB collaborators†, the GR@ACE collaborators†, the DEGESCO collaborators†, the DemGene collaborators†, the GERAD collaborators†.

* Co-first authors

†Data used in the preparation of this article were obtained from the EADB, GR@ACE, DEGESCO, DemGene, EADI, GERAD. As such, the investigators within these consortia contributed to the design and implementation of their respective consortium and/or provided data. Group authors who participated in analysis or writing of this report are listed at the end of this article. A full list of collaborators, including those who did not participate in analysis or writing of this report, is provided in the supplementary materials.

[#] Correspondence should be addressed to:
Yann Le Guen
Department of Neurology and Neurological Sciences – Greicius lab
Stanford University
290 Jane Stanford Way, E265, CA 94305-5090
Tel: 650 666 2696
Email: yleguen@stanford.edu

Word count: 3257 words

Group authors *(authors may be listed in two groups, authors already in the byline are not listed)*

EADB group authors

Beatrice Arosio, PhD 38, Luisa Benussi, PhD 39, Anne Boland, PharmD,PhD 16, Barbara Borroni, MD 40, Paolo Caffarra, PhD 41, Delphine Daian 16, Antonio Daniele, MD,PhD 42,43, Stéphanie Debette, MD,PhD 44,45, Martin Dichgans, MD 46,47,48, Carole Dufouil, PhD 49,50, Emrah Düzel, MD,PhD 51,52, Daniela Galimberti, PhD 53,54, Vilmantas Giedraitis, PhD 28, Timo Grimmer, PhD 55, Caroline Graff, MD 56, Edna Grünblatt, PhD 57,58,59, Olivier Hanon, MD 60, Lucrezia Hausner, PhD 61, Stefanie Heilmann-Heimbach, PhD 62, Henne Holstege, PhD 7,63, Jakub Hort, MD,PhD 64,65, Deckert Jürgen, MD 66, Teemu Kuulasmaa 32, Aad van der Lugt, PhD 67, Carlo Masullo, MD 68, Patrizia Mecocci, MD,PhD 69, Shima Mehrabian, PhD 70, Alexandre de Mendonça, PhD 71, Susanne Moebus, PhD 72, Benedetta Nacmias, PhD 73,74, Gael Nicolas, MD,PhD 75, Robert Olaso, PhD 16, Goran Papenberg, PhD 76, Lucilla Parnetti, MD,PhD 77, Florence Pasquier, MD,PhD 78, Oliver Peters, MD 79,80, Yolande A.L. Pijnenburg, MD 7, Julius Popp, MD 81,82,83, Innocenzo Rainero, MD,PhD 84, Inez Ramakers, PhD 85, Steffi Riedel-Heller, MD 86, Dan Rujescu, MD 87, Nikolaos Scarmeas, MD,PhD 88,89, Philip Scheltens, MD,PhD 7, Norbert Scherbaum, MD 90, Anja Schneider, MD 21,91, Davide Seripa, PhD 92, Hilkka Soininen, MD,PhD 93, Vincenzo Solfrizzi, MD,PhD 94, Gianfranco Spalletta, MD,PhD 95,96, Alessio Squassina, PhD 97, John van Swieten, MD 98, Thomas J Tegos, MD,PhD 26, Lucio Tremolizzo, MD,PhD 99, Frans Verhey, MD,PhD 100, Martin Vyhnalek, MD,PhD 64,65, Jens Wiltfang, MD 101,102,103

GRA@CE group authors

Mercè Boada, MD,PhD 4,5, Pablo García-González, MSc 4,5, Raquel Puerta, MSc 4, Luis M Real, PhD 104,105

DEGESCO group authors

Victoria Álvarez, PhD 106,107, María J. Bullido, PhD 108,5,109, Jordi Clarimon, PhD 110,5, José María García-Alberca, MD,PhD 111,5, Pablo Mir, MD,PhD 112,5, Fermin Moreno, MD,PhD 113,5,114, Pau Pastor, MD,PhD 115,116, Gerard Piñol-Ripoll, MD,PhD 117,118, Laura Molina-Porcel, MD,PhD 119,120, Jordi Pérez-Tur, PhD 121,5,122, Eloy Rodriguez-Rodriguez, MD,PhD 123,5, Jose Luís Royo, PhD 124, Raquel Sánchez-Valle, MD,PhD 125

GERAD authors

Martin Dichgans, MD 46,47,48, Dan Rujescu, MD 87

Affiliations

1. Department of Neurology and Neurological Sciences, Stanford University, Palo Alto, California, United States of America.
2. Institut du Cerveau - Paris Brain Institute - ICM, Paris, France.
3. Univ. Lille, Inserm, CHU Lille, Institut Pasteur de Lille, U1167-RID-AGE Facteurs de risque et déterminants moléculaires des maladies liées au vieillissement, F-59000 Lille, France.
4. Research Center and Memory clinic Fundació ACE, Institut Català de Neurociències Aplicades, Universitat Internacional de Catalunya, Barcelona, Spain.
5. Networking Research Center on Neurodegenerative Diseases (CIBERNED), Instituto de Salud Carlos III, Madrid, Spain.
6. UKDRI@Cardiff, School of Medicine, Cardiff University, Wales, United Kingdom.
7. Alzheimer Center Amsterdam, Department of Neurology, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam UMC, Amsterdam, The Netherlands.
8. Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, Vrije University, Amsterdam, The Netherlands.
9. Division of Neurogenetics and Molecular Psychiatry, Department of Psychiatry and Psychotherapy, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany.
10. Estudios en Neurociencias y Sistemas Complejos (ENyS) CONICET-HEC-UNAJ.
11. Complex Genetics of Alzheimer's Disease Group, VIB Center for Molecular Neurology, VIB, Antwerp, Belgium.
12. Laboratory of Neurogenetics, Institute Born - Bunge, Antwerp, Belgium.
13. Department of Biomedical Sciences, University of Antwerp, Antwerp, Belgium.
14. Department of Clinical Biochemistry, Copenhagen University Hospital - Rigshospitalet, Copenhagen, Denmark.
15. Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark.
16. Université Paris-Saclay, CEA, Centre National de Recherche en Génomique Humaine, 91057, Evry, France.
17. Quantitative Sciences Unit, Department of Medicine, Stanford University, Stanford, CA, USA.
18. School of Biosciences and Veterinary Medicine, University of Camerino, Camerino, Italy.
19. Univ. Lille, Inserm, CHU Lille, Institut Pasteur de Lille, U1167-RID-AGE LabEX DISTALZ Risk actors and molecular determinants of ageing diseases, F-59000 Lille, France.
20. Department of Psychiatry and Psychotherapy, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany.
21. German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany.
22. Cluster of Excellence Cellular Stress Responses in Aging-associated Diseases (CECAD), University of Cologne, Cologne, Germany.
23. Translational Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom.
24. Department of Epidemiology, ErasmusMC, Rotterdam, The Netherlands.
25. Nuffield Department of Population Health Oxford University, Oxford, United Kingdom.
26. 1st Department of Neurology, Medical school, Aristotle University of Thessaloniki, Thessaloniki, Makedonia, Greece.
27. Alzheimer's Centre Reina Sofia-CIEN Foundation, Madrid, Spain.
28. Department of Public Health and Caring Sciences / Geriatrics, Uppsala University, Uppsala, Sweden.
29. Krembil Brain Institute, University Health Network, Toronto, Canada.
30. Department of Medicine and Tanz Centre for Research in Neurodegenerative Diseases, University of Toronto, Toronto, Canada.
31. Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy.
32. Institute of Biomedicine, University of Eastern Finland, Joensuu, Kuopio, Eastern Finland, Finland.
33. Division of Psychological Medicine and Clinical Neuroscience, School of Medicine, Cardiff University, Wales, United Kingdom.
34. Department of Neurodegenerative diseases and Geriatric Psychiatry, University Hospital Bonn, Medical Faculty, Bonn, Germany.
35. Department of Psychiatry & Glenn Biggs Institute for Alzheimer's and Neurodegenerative Diseases, San Antonio, TX, USA.
36. NORMENT Centre, Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway.
37. Institute of Clinical Medicine, University of Oslo, Oslo, Norway.
38. Department of Clinical Sciences and Community Health, University of Milan, 20122 Milan, Italy.
39. Molecular Markers Laboratory, IRCCS Istituto Centro San Giovanni di Dio Fatebenefratelli, Brescia, Italy.

40. Centre for Neurodegenerative Disorders, Neurology Unit, Department of Clinical and Experimental Sciences, University of Brescia, Brescia, Italy.
41. Unit of Neurology, University of Parma and AOU, Parma, Italy.
42. Department of Neuroscience, Università Cattolica del Sacro Cuore, Rome, Italy.
43. Neurology Unit, IRCCS Fondazione Policlinico Universitario A. Gemelli, Rome, Italy.
44. University Bordeaux, Inserm, Bordeaux Population Health Research Center,Bordeaux, France.
45. Department of Neurology, Bordeaux University Hospital, Bordeaux, France.
46. Institute for Stroke and Dementia Research (ISD), University Hospital, LMU Munich, Munich, Germany.
47. German Center for Neurodegenerative Diseases (DZNE), Munich, Germany.
48. Munich Cluster for Systems Neurology (SyNergy), Munich, Germany.
49. Inserm, Bordeaux Population Health Research Center, UMR 1219, Univ. Bordeaux, ISPED, CIC 1401-EC, Université de Bordeaux, Bordeaux, France.
50. CHU de Bordeaux, Pole santé publique, Bordeaux, France.
51. German Center for Neurodegenerative Diseases (DZNE), Magdeburg, Germany.
52. Institute of Cognitive Neurology and Dementia Research (IKND), Otto-von-Guericke University, Magdeburg, Germany.
53. Neurodegenerative Diseases Unit, Fondazione IRCCS Ca' Granda, Ospedale Policlinico, Milan, Italy.
54. Department of Biomedical, Surgical and Dental Sciences, University of Milan, Milan, Italy.
55. Technical University of Munich, School of Medicine, Klinikum rechts der Isar, Department of Psychiatry and Psychotherapy, Munich , Germany.
56. Unit for Hereditary Dementias, Theme Aging, Karolinska University Hospital-Solna, 171 64 Stockholm, Sweden.
57. Department of Child and Adolescent Psychiatry and Psychotherapy, Psychiatric University Hospital Zurich, University of Zurich, Zurich, Switzerland.
58. Neuroscience Center Zurich, University of Zurich and ETH Zurich, Switzerland.
59. Zurich Center for Integrative Human Physiology, University of Zurich, Switzerland.
60. Université de Paris, EA 4468, APHP, Hôpital Broca, Paris, France.
61. Department of Geriatric Psychiatry, Central Institute of Mental Health Mannheim, Faculty Mannheim, University of Heidelberg, Germany.
62. Institute of Human Genetics, University of Bonn, School of Medicine & University Hospital Bonn, Bonn, Germany.
63. Department of Clinical Genetics, VU University Medical Centre, Amsterdam, The Netherlands.
64. Memory Clinic, Department of Neurology, Charles University, 2nd Faculty of Medicine and Motol University Hospital, Czech Republic.
65. International Clinical Research Center, St. Anne's University Hospital Brno, Brno, Czech Republic.
66. Department of Psychiatry, Psychosomatics and Psychotherapy, Center of Mental Health, University Hospital of Würzburg, Würzburg, Germany
67. Department of Radiology&Nuclear medicine, ErasmusMC, Rotterdam, The Netherlands.
68. Institute of Neurology, Catholic University of the Sacred Heart, Rome, Italy.
69. Institute of Gerontology and Geriatrics,Department of Medicine and Surgery, University of Perugia, Italy.
70. Clinic of Neurology, UH "Alexandrovska", Medical University - Sofia, Sofia, Bulgaria.
71. Faculty of Medicine, University of Lisbon, Portugal.
72. Institute for Urban Public Health, University Hospital of University Duisburg-Essen, Essen, Germany.
73. Department of Neuroscience, Psychology, Drug Research and Child Health University of Florence, Florence, Italy.
74. IRCCS Fondazione Don Carlo Gnocchi, Florence, Italy.
75. Normandie Univ, UNIROUEN, Inserm U1245 and CHU Rouen, Department of Genetics and CNR-MAJ, F-76000 Rouen, France.
76. Aging Research Center, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet and Stockholm University, Stockholm, Sweden.
77. Centre for Memory Disturbances, Lab of Clinical Neurochemistry, Section of Neurology, University of Perugia, Perugia, Italy.
78. Université de Lille, Inserm 1172, CHU Clinical and Research Memory Research Centre (CMRR) of Distalz, Lille, France.
79. German Center for Neurodegenerative Diseases (DZNE), Berlin, Germany.

80. Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Institute of Psychiatry and Psychotherapy, Hindenburgdamm 30, 12203 Berlin, Germany.
81. Old Age Psychiatry, Department of Psychiatry, Lausanne University Hospital, Lausanne, Switzerland.
82. Department of Geriatric Psychiatry, University Hospital of Psychiatry Zürich, Zürich, Switzerland.
83. Institute for Regenerative Medicine, University of Zürich, Switzerland.
84. Department of Neuroscience “Rita Levi Montalcini”, University of Torino, Torino, Italy.
85. Maastricht University, Department of Psychiatry & Neuropsychologie, Alzheimer Center Limburg, Maastricht, The Netherlands.
86. Institute of Social Medicine, Occupational Health and Public Health, University of Leipzig, Leipzig, Germany.
87. Medical University of Vienna, Department of Psychiatry and Psychotherapy, Vienna, Austria
88. 1st Department of Neurology, Aiginition Hospital, National and Kapodistrian University of Athens, Medical School, Greece.
89. Taub Institute for Research in Alzheimer’s Disease and the Aging Brain, The Gertrude H. Sergievsky Center, Department of Neurology, Columbia University, New York, New York, United States of America.
90. LVR-Hospital Essen, Department of Psychiatry and Psychotherapy, Medical Faculty, University of Duisburg-Essen, Essen, Germany.
91. Department for Neurodegenerative Diseases and Geriatric Psychiatry, University Hospital Bonn, Venusberg-Campus 1, 53127 Bonn, Germany.
92. Laboratory for Advanced Hematological Diagnostics, Department of Hematology and Stem Cell Transplant, "Vito Fazzi"Hospital, Lecce (LE), Italy.
93. Institute of Clinical Medicine - Neurology, University of Eastern Finland, Finland.
94. Interdisciplinary Department of Medicine, Geriatric Medicine and Memory Unit, University of Bari “A. Moro, Bari, Italy.
95. Laboratory of Neuropsychiatry, IRCCS Santa Lucia Foundation, Rome, Italy.
96. Department of Psychiatry and Behavioral Sciences, Baylor College of Medicine, Houston, Texas, United States of America.
97. Department of Biomedical Sciences, University of Cagliari, Italy.
98. Department of Neurology, ErasmusMC, Rotterdam, The Netherlands.
99. Neurology, "San Gerardo" hospital, Monza and University of Milano-Bicocca, Italy.
100. Maastricht University, Department of Psychiatry & Neuropsychologie, Alzheimer Center Limburg, Maastricht, Netherlands.
101. Department of Psychiatry and Psychotherapy, University Medical Center Goettingen, Goettingen, Germany.
102. German Center for Neurodegenerative Diseases (DZNE), Goettingen, Germany.
103. Neurosciences and Signaling Group, Institute of Biomedicine (iBiMED), Department of Medical Sciences, University of Aveiro, Aveiro, Portugal.
104. Unidad Clínica de Enfermedades Infecciosas y Microbiología. Hospital Universitario de Valme, Sevilla, Spain.
105. Depatamento de Especialidades Quirúrgicas, Bioquímica e Inmunología. Facultad de Medicina. Universidad de Málaga. Málaga, Spain
106. Laboratorio de Genética. Hospital Universitario Central de Asturias, Oviedo, Spain.
107. Instituto de Investigación Sanitaria del Principado de Asturias (ISPA), Spain.
108. Centro de Biología Molecular Severo Ochoa (UAM-CSIC), Universidad Autónoma de Madrid, Madrid, Spain.
109. Instituto de Investigacion Sanitaria ‘Hospital la Paz’ (IdIPaz), Madrid, Spain.
110. Department of Neurology, II B Sant Pau, Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona, Barcelona, Spain.
111. Alzheimer Research Center & Memory Clinic, Andalusian Institute for Neuroscience, Málaga, Spain.
112. Unidad de Trastornos del Movimiento, Servicio de Neurología y Neurofisiología, Instituto de Biomedicina de Sevilla (IBiS), Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Seville, Spain.
113. Department of Neurology. Hospital Universitario Donostia. San Sebastian, Spain.
114. Neurosciences Area. Instituto Biodonostia. San Sebastian, Spain.
115. Fundació Docència i Recerca MútuaTerrassa, Terrassa, Barcelona, Spain.
116. Memory Disorders Unit, Department of Neurology, Hospital Universitari Mutua de Terrassa, Terrassa, Barcelona, Spain.
117. Unitat Trastorns Cognitius, Hospital Universitari Santa Maria de Lleida, Lleida, Spain.
118. Institut de Recerca Biomedica de Lleida (IRBLLeida), Lleida, Spain.
119. Neurological Tissue Bank - Biobanc- Hospital Clinic -IDIBAPS, Barcelona, Spain.
120. Alzheimer’s disease and other cognitive disorders Unit. Neurology Department, Hospital Clinic , Barcelona, Spain.

121. Unitat de Genètica Molecular, Institut de Biomedicina de València-CSIC, Valencia, Spain.
122. Unidad Mixta de Neurologia Genètica, Instituto de Investigación Sanitaria La Fe, Valencia, Spain.
123. Neurology Service, Marqués de Valdecilla University Hospital (University of Cantabria and IDIVAL), Santander, Spain.
124. Depatamento de Especialidades Quirúrgicas, Bioquímica e Inmunología. Facultad de Medicina. Universidad de Málaga. Málaga, Spain.
125. Alzheimer's disease and other cognitive disorders unit. Service of Neurology. Hospital Clínic of Barcelona. Institut d'Investigacions Biomèdiques August Pi i Sunyer, University of Barcelona, Barcelona, Spain.

Key Points (75-100 word or less)

Question: Are *APOE* missense variants, other than the common *APOE* alleles ε2 and ε4, associated with AD risk?

Findings: We meta-analyzed multiple studies including 67,896 Alzheimer's disease (AD) cases, 28,484 proxy-AD cases and 340,306 healthy controls. Two rare missense variants substantially reduced the risk of AD. *APOE*-ε3[V236E] reported previously but lacking large-scale validation, reduced risk by more than 60%. *APOE*-ε4[R251G], not previously associated with AD, reduced risk by more than 50% and reached genome-wide significance.

Meaning: Single amino acid alterations of the *APOE*-ε3 and *APOE*-ε4 isoforms can result in substantial risk reduction for AD. Functional studies examining these variants should elucidate the role of apoE in AD pathogenesis.

Abstract (350-word limit)

Importance: The *APOE*-ε2 and *APOE*-ε4 alleles are, respectively, the strongest protective and risk-increasing genetic variants for late-onset Alzheimer's disease (AD). However, the mechanisms linking *APOE* to (AD)—particularly the apoE protein's role in AD pathogenesis and how this is affected by *APOE* variants—remain poorly understood. Identifying missense variants in addition to *APOE*-ε2 and *APOE*-ε4 could provide critical new insights.

Objective: To determine whether rare missense variants on *APOE* are associated with AD risk.

Design: Association with case-control status was tested in a sequenced discovery sample (Stage 1) and followed-up in several microarray imputed cohorts as well as the UK Biobank whole-exome sequencing resource using a proxy-AD phenotype (Stages 2+3). All data were retrieved between September 2015 and November 2021 and analyzed between April 2021 and November 2021.

Setting: This study combined case-control, family-based, population-based, and longitudinal AD-related cohorts that recruited referred and volunteer participants.

Participants: Stage 1 included 37,409 non-unique participants of European or Admixed-European ancestry, with 11,868 cases and 11,934 controls passing analysis inclusion criteria. In Stages 2+3, 475,473 participants were considered across 8 cohorts, of which 84,513 cases and proxy-AD cases, and 328,372 controls passed inclusion criteria, and were of European ancestry. Selection criteria were cohort specific, and this study was performed a posteriori on individuals who were genotyped. Among the available genotypes 76,195 were excluded. The number who declined to participate in the original studies was not available.

Main Outcome(s) and Measure(s): In primary analyses, the AD risk associated with each missense variant was estimated, as appropriate, with either linear-mixed-model regression or logistic regression. In secondary analyses, we estimated associations with age-at-onset using linear-mixed-model regression, and risk of conversion to AD using competing risk regression.

Results: A total of 544,384 participants (57.4% females, age range 40-110 years old) were analyzed in the primary case-control analysis. Two missense variants were associated with a two to three-fold decreased AD risk: *APOE*-ε4[R251G] (odds ratio, 0.44; 95% confidence interval [CI], 0.33-0.59; $P=4.7\times10^{-8}$) and *APOE*-ε3[V236E] (odds ratio, 0.37; 95% CI, 0.25-0.56; $P=1.9\times10^{-6}$). Additionally, the cumulative incidence of AD in carriers of these variants was found to grow more slowly with age compared to non-carriers.

Conclusions and Relevance: We identified a novel variant associated with AD, R251G always co-inherited with ε4, which mitigates the ε4 associated AD risk, and confirmed

the protective effect of the V236E variant, always co-inherited with ε3. The location of these variants confirms that the carboxyl-terminal portion of apoE plays an important role in AD pathogenesis. The large risk reductions reported here, suggest that protein chemistry and functional assays of these variants should be pursued as they have the potential to guide drug development targeting *APOE*.

Introduction

Late-onset Alzheimer's disease (AD) is a highly polygenic neurodegenerative disorder with, to date, 75 risk loci associated with AD risk[1]. Most of the common single nucleotide polymorphisms (SNPs) at these loci only contribute a small amount to an individual's risk of AD[2], with the exception of the *APOE*-ε2 and ε4 missense variants that are associated with substantially decreased[3] and increased AD risk[4], respectively. It is estimated that 25% of the genetic variance of AD can be attributed to *APOE*-ε2 and *APOE*-ε4[5]. Despite the outsized role of these two common *APOE* alleles, more than 25 years after the initial studies linking them to AD their role in pathogenesis remains ill-defined. Human studies have shown that ε4 speeds, and ε2 slows, the age-related misprocessing of beta-amyloid, though how this occurs at the molecular level remains uncertain[6,7]. Even the most basic question, does ε4 act via a loss-of-function or gain-of-function mechanism, remains a point of contention[8]. Loss-of-function mutations on *APOE* are exceedingly rare and the sole case report describing a compound heterozygote with two loss-of-function mutations involved a patient who was too young to be informative[9]. The study of additional missense variants on *APOE* may also help to answer this critical question and further elucidate the role of *APOE* in AD. In addition to ε2 and ε4, the only common missense variant (with a minor allele frequency (MAF) > 1%) is Arg145Cys (R145C) an African-ancestry variant always found co-inherited with *APOE*-ε3, which we have shown increases risk for AD[10]. The Arg136Ser (R136S) Christchurch variant has recently been posited to play a protective role in early-onset AD related to *PSEN1* mutations, but this study had no statistical genetics support as it was based on data from a single patient[11]. Finally, strong functional evidence has been marshalled recently to support a protective role for the Val236Glu (V236E) variant, though this was based on data from an earlier case-control study with only ~9,000 subjects[12,13], likely underpowered to provide firm estimates of disease risk.

On this background, we aimed to investigate, at large scale, the association of rare missense variants on *APOE* with AD risk. We used the Alzheimer's Disease Sequencing Project (ADSP) whole-genome (WGS) and whole-exome sequencing (WES) data as our discovery sample (Stage 1), and sought to replicate significant variants (Stages 2+3) in

multiple cohorts using micro-array data imputed on the TOPMed reference panel[14], or by using directly sequenced and genotyped variants from a large Danish general prospective population cohort[15], as well as using the proxy-AD phenotype[1] in the UK Biobank WES data. After filtering, three variants, Leu28Pro (L28P), Val236Glu (V236E), and Arg251Gly (R251G), were tested for their association with AD risk after adjusting for ε2 and ε4 dosages. In complementary analyses, we assessed these associations in an *APOE*-stratified approach to account for the complete linkage disequilibrium of these variants with either the ε2, ε3, or ε4 allele. In secondary analyses, combining Stages 1 and 2 datasets, we tested their association with age-at-onset in AD cases and with risk of conversion to AD using competing risk regression.

Methods

PARTICIPANTS AND SOURCES OF DATA

Participants or their caregivers provided written informed consent in the original studies. The current study protocol was granted an exemption by the Stanford University institutional review board because the analyses were carried out on deidentified, off-the-shelf data; therefore, additional informed consent was not required. For Stage 1 and Stage 2, phenotypic information and genotypes were obtained from publicly released genome-wide association study datasets assembled by the Alzheimer's Disease Genetics Consortium (ADGC) and derived from WES and WGS data generated by the Alzheimer Disease Sequencing Project (ADSP), with phenotype and genotype ascertainment described elsewhere[16–20]. The cohorts' queried accession numbers, as well as the sequencing technology or single nucleotide polymorphism (SNP) genotyping platforms are described in **eTables 1 and 2**. Information about Stage 3, which included external replication cohorts and UK Biobank, is provided as a **Supplementary Note**. Briefly, these included EABD-core, EADI, GERAD, DemGene, and GR@ACE/DEGESCO for which phenotype, genotype quality control and imputation have already been described in Bellenguez et al.[1]; and CCHS & CGPS *APOE* sequencing and genotyping were described in Rasmussen et al.[15]. The following sections describe quality control procedures and ancestry determination applied to the ADSP and ADGC samples respectively used as Stage 1 and Stage 2. The STREGA reporting guidelines were followed.

QUALITY CONTROL PROCEDURES

Prior to ancestry, principal components and relatedness determination, in each cohort-platform, variants were excluded based on genotyping rate (< 95%), MAF < 1%, and Hardy-Weinberg equilibrium in controls ($p < 10^{-6}$) using PLINK v1.9[21]. gnomAD[22] database-derived information was used to filter out SNPs that met one of the following exclusion criteria[23,24]: (i) located in a low complexity region, (ii) located within common structural variants (MAF > 1%), (iii) multiallelic SNPs with MAF > 1% for at least two alternate alleles, (iv) located within a common insertion/deletion, (v) having any flag

different than PASS in gnomADv.3, (vi) having potential probe polymorphisms. The latter are defined as SNPs for which the probe may have variable affinity due to the presence of other SNP(s) within 20 bp and with MAF > 1%. Individuals with more than 5% genotype missingness were excluded. Duplicate individuals were identified with KING[25] and their clinical, diagnostic and pathological data (including age-at-onset of cognitive symptoms, age-at-examination for clinical diagnosis, age-at-last exam, age-at-death), as well as sex, race, and *APOE* genotype were cross-referenced across cohorts. Duplicate entries with irreconcilable phenotype or discordant sex were flagged for exclusion. For individuals with duplicated genotype in sequencing and imputed data, the sequencing entry was used in the Stage 1 discovery set and the imputed entry was not included in the Stage 2 replication set. To apply the *PCAir* and *PCRelate* methods described in the statistical analysis section, we simply considered the intersection of the variants passing quality control in both ADSP WES and ADSP WGS in the discovery, and similarly the intersection of the variants across cohorts genotyping platform in the replication.

ANCESTRY DETERMINATION

For each cohort, we first determined the ancestry of each individual with SNPWeights v2[26] using reference populations from the 1000 Genomes Consortium[27]. By applying an ancestry percentage cut-off > 75%, the samples were stratified into five super populations: South-Asians, East-Asians, Americans, Africans, and Europeans, and an Admixed group composed of individuals not passing the 75% cut-off in any single ancestry (**eTable 3**)[10,23]. Since the *APOE* missense variants of interest L28P, V236E, and R251G are too rare to assess reliably in non-European ancestry populations (**eTable 4**), we restricted our analysis to European and Admixed-European individuals. Admixed-European individuals were also included in the main analysis and were part of the Admixed group defined above and had at least 15% European ancestry. We performed sensitivity analyses in increments of 30%, including Admixed-European individuals at 45% and 75% cutoffs. The latter corresponding to the super population threshold.

IMPUTATION

Each cohort-genotyping platform was imputed on the TOPMed imputation server per ancestry group to obtain an imputation quality (R^2) per ancestry group. We retained cohorts with $R^2 > 0.70$ at rs199768005 for the V236E analyses, and at rs26760661 for the R251G analyses. As there was no significant association signal for rs769452 (L28P) in the Stage 1 primary analysis, we did not check its imputation quality in Stage 2 samples.

APOE GENOTYPE ASCERTAINMENT

We directed specific attention to the genotyping of the SNPs determining the main *APOE* genotype (rs429358 and rs7412), rs769452-C (*APOE*[L28P]), rs199768005-A (*APOE*[V236E]), and rs267606661-G (*APOE*[R251G]) and follow the procedure described in [10]. Note that Leu28Pro (L28P), Val236Glu (V236E), and Arg251Gly (R251G) are also sometimes respectively referred to as L46P, V254E, and R269G, when the first 18 codons of *APOE* encoding a signal peptide are included.

SAMPLES ANALYZED

Our discovery sample (Stage 1) was composed of European and Admixed-European ancestry individuals from the ADSP WES and WGS, corresponding to 11,868 AD cases and 11,934 cognitively normal controls (**Table 1**). **eFigure 1** provides a flowchart of the filtering steps leading to the inclusion of these individuals and describes how these datasets were combined. To build a replication sample (Stage 2) for V236E and R251G, we queried for individuals of European and Admixed-European ancestry in all the publicly available microarray genetic datasets that we had access to at the time of the study in July 2021 (**Table 1**). These datasets are largely part of the ADGC and as such this replication will be referred to hereafter as the ADGC replication in Stage 2. After quality control and duplicate removal, 7,768 AD cases and 8,059 controls remained in the ADGC replication sample. **eTable 5** presents the demographics of the remaining AD cases and cognitively unimpaired controls. In Stage 3, we pursued additional replication in external datasets (not publicly available) and in the UK Biobank WES using the proxy-

AD phenotype (**Table 1, Supplementary Note**). Overall, the external replications included 36,393 cases and 150,943 controls, and the UK Biobank replication included 28,484 proxy-AD cases and 157,436 controls. Across cohorts reported in Table 1, the *APOE* genotype were split as follows: ε2/ε2: 0.5%, ε2/ε3: 10.4%, ε3/ε3: 54.5%, ε2/ε4: 2.5%, ε3/ε4: 27.7%, ε4/ε4: 4.4%.

STUDY DESIGN & STATISTICAL ANALYSES

In our analysis, we only considered missense variants with a minor allele count above 10 in any *APOE* main genotype groups in our next generation sequencing discovery (Stage 1) to avoid outlier-confounded effect size estimates[28]. Three *APOE* missense variants were retained for further analyses: L28P, V236E, and R251G (**eTable 4**). The V236E variant is always co-inherited with *APOE*-ε3, and the L28P and R251G are always co-inherited with *APOE*-ε4 (**eTable 6**). Two variants are co-inherited when they are on the same chromosome copy and close enough to each other that a meiotic crossover event never occurs between them. We thus developed two complementary approaches to take into account these linkage disequilibrium structures. In primary analyses, we estimated the AD risk associated with L28P, V236E, and R251G on case-control diagnoses using linear-mixed-model regression (Stages 1+2, and UK Biobank) and logistic regression (Stage 3) model, adjusted for ε2 and ε4 dosages, in addition to the covariates described below for all analyses. The adjustment by the common ε3 and ε4 *APOE* alleles is necessary because the rare variants tested here are always co-inherited with either the ε3 or ε4 *APOE* allele. In complementary analyses, we also estimated the AD risk associated with V236E and R251G stratified by their associated common *APOE* allele genotype. V236E was assessed in *APOE*-ε3/ε3 and R251G was assessed in the *APOE*-ε3/ε4 stratum. An association was considered significant in Stage 1, if it reached a Bonferroni-corrected p-value threshold of 0.017 (≈ 0.05/3) in the model adjusted for ε2 and ε4 dosages. L28P was not associated with AD risk in this model and was not studied further.

Sample sizes and demographics for the stratified analyses are shown in **eTable 6**. In sensitivity analyses, we estimated AD risk associations for different European

ancestry inclusion thresholds. In secondary analyses, combining Stages 1 and 2 datasets, we estimated the influence of significant Stage 1 variants on age-at-onset (AAO) in AD cases using linear mixed model regression, and risk of conversion to AD using competing risk regression. In secondary analyses, associations were considered significant when passing the nominal p-value threshold of 0.05. The case-control and age-at-onset analyses used linear mixed model regression available through the *GENESIS* package (v3.12)[29]. Multivariate competing risk regression and cumulative incidence estimation were implemented using the *cmprsk* package (v2.2)[30]. In this time-to-event analysis, failure events were defined as age-at-onset for cases (conversion to AD) and age-at-death for controls. Controls without reported death were right censored at age-at-last-visit. Left censoring was set at 50 years old, and younger individuals were excluded from the analysis. All statistical analyses were adjusted for sex and four genetic principal components estimated with the *PC-Air* method[31] implemented in *GENESIS*. Linear mixed model analyses were additionally covaried by a sparse genetic relationship matrix estimated with the *PC-Relate* method[32] implemented in *GENESIS*. Case-control analyses were not adjusted for age given that correcting for age when cases are younger than controls leads to the model incorrectly inferring the age effect on AD risk, resulting in statistical power loss[23].

Case-control analyses in Stage 3, external replication cohorts and proxy-AD phenotype in UK Biobank, were implemented to be consistent with the Stage 1 primary analyses. Exact model/analysis details are described in a **Supplementary Note**. For the ADSP/ADGC cohorts, all statistical analyses were performed in R (v4.0.2). All meta-analyses were implemented with a fixed-effect inverse variance weighted design implemented in the *metafor* R package (v.3.0.2)[33].

RESULTS

In Stage 1 primary analyses, V236E (rs199768005-A) and R251G (rs267606661-G) were associated with a four to five-fold decreased AD risk in non-stratified analyses adjusted for ε2 and ε4 dosages (V236E: OR = 0.23; 95% CI; 0.09-0.56; P = $1.4x10^{-3}$; R251G: OR = 0.20; 95% CI; 0.08-0.49; P = $3.7x10^{-4}$, **Figure 1**, **Table 2**). Similarly, in *APOE*-stratified analyses, V236E was associated with a threefold decreased AD risk in ε3/ε3 individuals (OR = 0.31; 95% CI; 0.12-0.82; P = 0.02) and R251G was associated with a fivefold decreased AD risk in ε3/ε4 individuals (OR = 0.17; 95% CI; 0.06-0.48; P = $7.8x10^{-4}$, **Table 2**). The L28P variant (rs769452-C) was not associated with AD risk in the non-stratified analyses (odds ratio (OR) = 1.12; 95% confidence interval [CI]; 0.77-1.62; P = 0.56). As such, it was not investigated further.

In Stages 2+3, across multiple replication cohorts, the effects of V236E and R251G in non-stratified analyses were concordant and both were significantly associated with AD risk: V236E (OR = 0.42; 95% CI, 0.27-0.66; P=$2.0x10^{-4}$) and R251G (OR = 0.48; 95% CI, 0.35-0.66; P= $5.8x10^{-6}$). The overall meta-analysis (**Figure 1, Table 2**) provides robust effect size estimate for these two variants and confirmed their association with a two to three-fold decreased AD risk: V236E (OR = 0.37; 95% CI, 0.25-0.56; P=$1.9x10^{-6}$) and R251G (OR = 0.44; 95% CI, 0.33-0.59; P=$4.7x10^{-8}$). Similar results were obtained in *APOE*-stratified meta-analyses (**Table 2, eFigure 1**). We further estimated the odds per *APOE* genotype group, using ε3/ε3 individuals that did not carry V236E as the reference (i.e., odds ratio of *APOE*-ε3/ε3 individuals equals 1), by meta-analyzing the ADSP discovery and ADGC replication cohorts. Compared to the reference ε3/ε3 group, ε3/ε3[V236E] and ε3/ε4[R251G] individuals had AD risk lower than or similar to ε2/ε3 (**Figure 2**).

Results of sensitivity analyses evaluating different European ancestry cutoffs are shown in (**eTable 8, eFigure 2**). Briefly, the results remained unchanged when selecting admixed ancestry individuals with at least 45% European ancestry, or when restricting the analysis to European ancestry individuals (75% cutoff). We note that the odds ratio in the combined ADSP/ADGC datasets for V236E and R251G remain unchanged at

different ancestry cutoffs. For example, using an ancestry cutoff at 75%, the non-stratified meta-analysis yielded an odds ratio of 0.27 (95% CI, 0.12 to 0.58; P = 8.6×10^{-4}) for V236E as compared to an odds ratio of 0.26 using a cutoff of 15%. Similar observations were made for the R251G variant. As additional supplementary analyses, we assessed the effect of the inclusion of "all dementia" (rather than AD specifically) in the CCHS & CGPS dataset and we estimated the significance without including UK Biobank. Overall, the significance of the results slightly improved when including a broader dementia category (e.g. R251G, OR= 0.44; 95% CI, 0.33-0.59; P=3.5×10^{-8}, **eTable 9**). While removing UK Biobank proxy-AD phenotype samples reduced the significance of our results slightly, the ORs became slightly more protective (e.g. R251G, OR= 0.39; 95% CI, 0.27-0.56; P=1.2×10^{-7}, **eTable 10**).

In secondary analyses, including data from Stages 1+2, we considered the meta-analysis of ADSP/ADGC samples (**eTable 5**). In non-*APOE* stratified analyses adjusted for ε2 and ε4 dosages (**eTable 7**), V236E carriers had an age-at-AD-onset on average 10.5 years older than non-carriers (β = 10.64; 95% CI, 1.78 to 19.49; P = 0.02) and slower incidence with age (HR = 0.30; 95% CI; 0.12-0.76; P = 0.01). While R251G's association with age-at-onset was not significant (β = 0.97; 95% CI, -2.96 to 4.91; P = 0.63) and its association with reduced AD incidence with age was just nominally significant (HR = 0.67; 95% CI; 0.46-0.97; P = 0.04). In *APOE*-stratified analyses (**eTable 7**), a similar effect of V236E on age-at-AD-onset was observed in ε3/ε3 (β = 10.93; 95% CI, 1.06 to 20.81; P = 0.03). R251G carriers had an age-at-AD-onset on average 6 years older than non-carriers in ε3/ε4 but this association was only trending towards significance (β = 6.04; 95% CI, -0.71 to 12.79; P = 0.08). The competing risk results emphasized that the cumulative incidence of AD in ε3/ε3 participants grows slower with age in individuals carrying the V236E variant (hazard ratio [HR] = 0.40; 95% CI; 0.17-0.97; P = 0.04), and similarly in ε3/ε4 participants carrying the R251G variant (HR = 0.26; 95% CI; 0.13-0.54; P = 2.9×10^{-4}).

DISCUSSION

We have shown that two missense variants V236E and R251G are each associated with a more than 2-fold reduction in AD risk (**Figure 2**). These variants have an allele frequency of less than 0.1% in gnomAD v3.1, even when restricting this frequency estimate to Europeans (**eTable 4**). Due to their rarity and linkage disequilibrium with the common *APOE* -ε3 and -ε4 alleles, they have not been identified in prior genome-wide association studies[1]. The protective effect of V236E has already been reported in a smaller prior study focused on *APOE*[13] and was suggestive in a population-based study[15], but we validated this finding here in a large-scale genomic study and provide an improved estimate of its effect size. The association of R251G with AD risk has not been previously reported. This variant, carried on the same haplotype as ε4, is the first *APOE* variant found to mitigate the AD risk attributable to the ε4 isoform of the apoE protein. Notably, having R251G in association with *APOE*-ε4 results in a risk estimate similar to *APOE*-ε2, as shown in **Figure 2** where *APOE-ε3*/ε4[R251G] and *APOE-ε2*/ε3 have an equivalent odds ratio. Our study has several limitations (i) the V236E association was not genome-wide significant, (ii) we included the UKB dataset that does not include a direct clinical diagnosis of AD, (iii) due to the paucity of variant carriers in non-European ancestries we did not assess these variants in other ancestries (although they can be found in African-Americans and Admixed-Latinos based on gnomAD estimates (**eTable 4**)). These three caveats point to the need for further confirmation of these variants as available AD datasets grow and become more ancestrally diverse.

Regarding potential mechanisms driving these associations, it is notable that these two variants are on apoE's C-terminal domain. The common *APOE-ε2* and *APOE-ε4* alleles are located on the N-terminal domain of the protein near the receptor-binding region. Their outsized role in AD risk has, understandably, focused attention on the N-terminal domain and the differential capacity of these alleles to, for example, bind apoE's receptors[34,35]. The current results add support to studies suggesting that the C-terminal domain is also of critical importance for AD pathogenesis[36–38]. R251G is located within apoE's lipid-binding region (amino acid residues 244 to 272), while V236E is adjacent to

this region[8]. A recent publication provided evidence for the protectiveness of V236E against AD pathology and explored the functional mechanism supporting its protective role[12]. The lipid-binding region, with its abundance of non-polar residues, is thought to be a region that can foster oligomerization[39–41]. Switching a non-polar valine for an acidic glutamic acid might be predicted to reduce the hydrophobicity of this region and reduce its tendency to oligomerize. Notably, the authors showed reduced levels of insoluble Aβ and apoE aggregates in the brain of V236E carriers compared to non-carriers[12]. In 5xFAD mice, they observed that *APOE-ε3*[V236E] reduced Aβ deposition, plaque-associated immune response, and neuritic dystrophy around amyloid plaques[12]. Chemically, they noted that *APOE-ε3*[V236E] primarily remains as a monomer and is less likely to form oligomers compared to the canonical *APOE-ε3* allele[12]. This propensity of V236E to reduce apoE aggregation was also observed when this variant was introduced on an *APOE-ε4* allele. It is worth noting, however, that V236E also appears to increase dimerization (see their Figure S10[12]), which may impact apoE's ability to bind to its receptors[42–44].

Given that R251G is located squarely in the lipid-binding region of the protein, it is possible that R251G confers a protective effect by reducing apoE's ability to form insoluble oligomers. The switch from a charged arginine amino acid to a non-polar glycine might, however, be expected to increase rather than decrease oligomerization. Changes in this region could also enhance apoE-ε4's ability to bind lipids rendering it more like ε3 or ε2 in this capacity[45]. Alternatively, the introduction of glycine could disrupt the alpha-helix structure of the C-terminal impacting apoE-ε4's hypothesized N-terminal/C-terminal domain interaction[34,35]. In any case, pending protein chemistry experiments exploring potential structural and functional changes, the mechanism underlying the substantial protective effect of R251G remains to be elucidated.

Our work, performed on the largest available sample to-date for these two variants, has validated the protective effect of the V236E variant and has uncovered a novel protective missense variant on *APOE*-ε4. Each variant has a substantial effect on reducing the risk of AD. While some compelling functional data suggest that V236E confers protection by reducing oligomerization of apoE, there are alternative

mechanisms that merit consideration (increasing dimerization, for one). The protective mechanism of R251G remains unexplored but finding a single amino acid substitution that renders the *APOE-ε4* allele protective supports the idea that *APOE-ε4*-specific treatments are worth exploring[46,47]. We anticipate that the findings reported here will spark additional mechanistic work on apoE's role in AD pathogenesis.

FUNDING AND ACKNOWLEDGMENTS

This work was supported by the National Institute of Health and National Institute of Aging grants AG060747 (MDG), AG066206 (ZH), AG066515 (ZH, MDG), the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie (grant agreement No. 890650, YLG), the Alzheimer's Association (AARF-20-683984, MEB), and the Iqbal Farrukh and Asad Jamal Fund, a grant from the EU Joint Programme – Neurodegenerative Disease Research (European Alzheimer DNA BioBank, EADB; JPND). Inserm UMR1167 is also funded by the Inserm, Institut Pasteur de Lille, Lille Métropole Communauté Urbaine, and the French government's LABEX DISTALZ program (development of innovative strategies for a transdisciplinary approach to Alzheimer's disease). EADB thank the study participants, researchers, and staff for collecting and contributing to the data, the high-performance computing service at the University of Lille, and the staff at CEA-CNRGH for their help with sample preparation and genotyping, and technical assistance. Additional funders of individual investigators and institutions who contributed to data collection and genotyping are provided in the **Supplemental Online Content**.

References

1. Bellenguez C, Küçükali F, Jansen I, et al. New insights on the genetic etiology of Alzheimer's and related dementia. *medRxiv*. Published online January 1, 2020:2020.10.01.20200659. doi:10.1101/2020.10.01.20200659

2. de Rojas I, Moreno-Grau S, Tesi N, et al. Common variants in Alzheimer's disease and risk stratification by polygenic risk scores. *Nat Commun*. 2021;12(1):3417. doi:10.1038/s41467-021-22491-8

3. Corder EH, Saunders AM, Risch NJ, et al. Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. *Nature Genetics*. 1994;7(2):180-184. doi:10.1038/ng0694-180

4. 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

5. Ridge PG, Hoyt KB, Boehme K, et al. Assessment of the genetic variance of late-onset Alzheimer's disease. *Neurobiol Aging*. 2016;41:200.e13-200.e20. doi:10.1016/j.neurobiolaging.2016.02.024

6. Morris JC, Roe CM, Xiong C, et al. APOE predicts amyloid-beta but not tau Alzheimer pathology in cognitively normal aging. *Annals of Neurology*. 2010;67(1):122-131. doi:10.1002/ana.21843

7. Castellano JM, Kim J, Stewart FR, et al. Human apoE isoforms differentially regulate brain amyloid-β peptide clearance. *Sci Transl Med*. 2011;3(89):89ra57. doi:10.1126/scitranslmed.3002156

8. Belloy ME, Napolioni V, Greicius MD. A Quarter Century of APOE and Alzheimer's Disease: Progress to Date and the Path Forward. *Neuron*. 2019;101(5):820-838. doi:10.1016/j.neuron.2019.01.056

9. Mak ACY, Pullinger CR, Tang LF, et al. Effects of the absence of apolipoprotein e on lipoproteins, neurocognitive function, and retinal function. *JAMA Neurol*. 2014;71(10):1228-1236. doi:10.1001/jamaneurol.2014.2011

10. Le Guen Y, Belloy ME, Eger SJ, et al. *APOE Missense Variant R145C Is Associated with Increased Alzheimer's Disease Risk in African Ancestry Individuals with the APOE E3/E4 Genotype*.; 2021:2021.10.20.21265141. doi:10.1101/2021.10.20.21265141

11. Arboleda-Velasquez JF, Lopera F, O'Hare M, et al. Resistance to autosomal dominant Alzheimer's disease in an APOE3 Christchurch homozygote: a case report. *Nat Med*. 2019;25(11):1680-1683. doi:10.1038/s41591-019-0611-3

12. Liu CC, Murray ME, Li X, et al. APOE3-Jacksonville (V236E) variant reduces self-aggregation and risk of dementia. *Science Translational Medicine*. 2021;13(613):eabc9375. doi:10.1126/scitranslmed.abc9375

13. Medway CW, Abdul-Hay S, Mims T, et al. ApoE variant p.V236E is associated with markedly reduced risk of Alzheimer's disease. *Molecular Neurodegeneration*. 2014;9(1):11. doi:10.1186/1750-1326-9-11

14. Taliun D, Harris DN, Kessler MD, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature*. 2021;590(7845):290-299. doi:10.1038/s41586-021-03205-y

15. Rasmussen KL, Tybjærg-Hansen A, Nordestgaard BG, Frikke-Schmidt R. APOE and dementia – resequencing and genotyping in 105,597 individuals. *Alzheimer's & Dementia*. 2020;16(12):1624-1637. doi:10.1002/alz.12165

16. Beecham GW, Bis JC, Martin ER, et al. The Alzheimer's Disease Sequencing Project: Study design and sample selection. *Neurol Genet*. 2017;3(5):e194. doi:10.1212/NXG.0000000000000194

17. Weiner MW, Aisen PS, Jack CR, et al. The Alzheimer's Disease Neuroimaging Initiative: Progress report and future plans. *Alzheimer's & Dementia*. 2010;6(3):202-211.e7. doi:10.1016/j.jalz.2010.03.007

18. Bennett DA, Schneider JA, Buchman AS, Barnes LL, Boyle PA, Wilson RS. Overview and findings from the rush Memory and Aging Project. *Current Alzheimer research*. 2012;9(6):646-663. doi:10.2174/156720512801322663

19. Kunkle BW, Grenier-Boley B, Sims R, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nature Genetics*. 2019;51(3):414-430. doi:10.1038/s41588-019-0358-2

20. Kunkle BW, Schmidt M, Klein HU, 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;78(1):102. doi:10.1001/jamaneurol.2020.3536

21. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*. 2015;4(1):7. doi:10.1186/s13742-015-0047-8

22. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581(7809):434-443. doi:10.1038/s41586-020-2308-7

23. Le Guen Y, Belloy ME, Napolioni V, et al. A novel age-informed approach for genetic association analysis in Alzheimer's disease. *Alzheimer's Research & Therapy*. 2021;13(1):72. doi:10.1186/s13195-021-00808-5

24. Le Guen Y, Napolioni V, Belloy ME, et al. Common X-Chromosome Variants Are Associated with Parkinson Disease Risk. *Annals of Neurology*. 2021;90(1):22-34. doi:10.1002/ana.26051

25. 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

26. Chen CY, Pollack S, Hunter DJ, Hirschhorn JN, Kraft P, Price AL. Improved ancestry inference using weights from external reference panels. *Bioinformatics*. 2013;29(11):1399-1406. doi:10.1093/bioinformatics/btt144

27. Auton A, Abecasis GR, Altshuler DM, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74. doi:10.1038/nature15393

28. Bis JC, Jian X, Kunkle BW, et al. Whole exome sequencing study identifies novel rare and common Alzheimer's-Associated variants involved in immune response and transcriptional regulation. *Molecular Psychiatry*. 2020;25(8):1859-1875. doi:10.1038/s41380-018-0112-7

29. Gogarten SM, Sofer T, Chen H, et al. Genetic association testing using the GENESIS R/Bioconductor package. *Bioinformatics*. 2019;35(24):5346-5348. doi:10.1093/bioinformatics/btz567

30. Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. *Journal of the American Statistical Association*. 1999;94(446):496-509. doi:10.1080/01621459.1999.10474144

31. Conomos MP, Miller MB, Thornton TA. Robust Inference of Population Structure for Ancestry Prediction and Correction of Stratification in the Presence of Relatedness. *Genetic Epidemiology*. 2015;39(4):276-293. doi:https://doi.org/10.1002/gepi.21896

32. Conomos MP, Laurie CA, Stilp AM, et al. Genetic Diversity and Association Studies in US Hispanic/Latino Populations: Applications in the Hispanic Community Health Study/Study of Latinos. *The American Journal of Human Genetics*. 2016;98(1):165-184. doi:10.1016/j.ajhg.2015.12.001

33. Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software*. 2010;36(1):1-48. doi:10.18637/jss.v036.i03

34. Huang Y, Weisgraber KH, Mucke L, Mahley RW. Apolipoprotein E: diversity of cellular origins, structural and biophysical properties, and effects in Alzheimer's disease. *J Mol Neurosci*. 2004;23(3):189-204. doi:10.1385/JMN:23:3:189

35. Huang Y, Mahley RW. Apolipoprotein E: structure and function in lipid metabolism, neurobiology, and Alzheimer's diseases. *Neurobiol Dis*. 2014;72 Pt A:3-12. doi:10.1016/j.nbd.2014.08.025

36. Harris FM, Brecht WJ, Xu Q, et al. Carboxyl-terminal-truncated apolipoprotein E4 causes Alzheimer's disease-like neurodegeneration and behavioral deficits in transgenic mice. *PNAS*. 2003;100(19):10966-10971. doi:10.1073/pnas.1434398100

37. Bien-Ly N, Andrews-Zwilling Y, Xu Q, Bernardo A, Wang C, Huang Y. C-terminal-truncated apolipoprotein (apo) E4 inefficiently clears amyloid-β (Aβ) and acts in concert with Aβ to elicit neuronal and behavioral deficits in mice. *PNAS*. 2011;108(10):4236-4241. doi:10.1073/pnas.1018381108

38. Huang YWA, Zhou B, Wernig M, Südhof TC. ApoE2, ApoE3, and ApoE4 Differentially Stimulate APP Transcription and Aβ Secretion. *Cell*. 2017;168(3):427-441.e21. doi:10.1016/j.cell.2016.12.044

39. Choy N, Raussens V, Narayanaswami V. Inter-molecular coiled-coil formation in human apolipoprotein E C-terminal domain. *J Mol Biol*. 2003;334(3):527-539. doi:10.1016/j.jmb.2003.09.059

40. Westerlund JA, Weisgraber KH. Discrete carboxyl-terminal segments of apolipoprotein E mediate lipoprotein association and protein oligomerization. *J Biol Chem*. 1993;268(21):15745-15750.

41. Flowers SA, Rebeck GW. APOE in the normal brain. *Neurobiol Dis*. 2020;136:104724. doi:10.1016/j.nbd.2019.104724

42. Dyer CA, Cistola DP, Parry GC, Curtiss LK. Structural features of synthetic peptides of apolipoprotein E that bind the LDL receptor. *J Lipid Res*. 1995;36(1):80-88.

43. Weisgraber KH, Shinto LH. Identification of the disulfide-linked homodimer of apolipoprotein E3 in plasma. Impact on receptor binding activity. *J Biol Chem*. 1991;266(18):12029-12034.

44. Minami SS, Cordova A, Cirrito JR, et al. ApoE mimetic peptide decreases Aβ production in vitro and in vivo. *Mol Neurodegener*. 2010;5:16. doi:10.1186/1750-1326-5-16

45. Minagawa H, Gong JS, Jung CG, et al. Mechanism Underlying Apolipoprotein E (ApoE) Isoform-dependent Lipid Efflux From Neural Cells in Culture. *J Neurosci Res*. 2009;87(11):2498-2508. doi:10.1002/jnr.22073

46. Zhao N, Liu CC, Qiao W, Bu G. Apolipoprotein E, Receptors, and Modulation of Alzheimer's Disease. *Biol Psychiatry*. 2018;83(4):347-357. doi:10.1016/j.biopsych.2017.03.003

47. Williams T, Borchelt DR, Chakrabarty P. Therapeutic approaches targeting Apolipoprotein E function in Alzheimer's disease. *Mol Neurodegener*. 2020;15(1):8. doi:10.1186/s13024-020-0358-9

Data availability

Data used in preparation of this manuscript can be obtained upon application at:

- dbGaP (https://www.ncbi.nlm.nih.gov/gap/advanced_search/)
- NIAGADS and NIAGADS DSS (https://www.niagads.org/)
- LONI (https://ida.loni.usc.edu/)
- Synapse (https://adknowledgeportal.synapse.org/)
- RADC Rush (https://www.radc.rush.edu/)
- NACC (https://naccdata.org/)
- UK Biobank (https://biobank.ndph.ox.ac.uk/showcase/)

eTables 1 and 2 provide the details of repositories and accession number per cohort-platform group. UK Biobank WES data were analyzed under Application Number 45420.

Table 1. Demographics per *APOE* genotype. DX: diagnosis, CN: cognitively normal, AD: Alzheimer's disease, N: number of individuals, %Females: percentage of female individuals, μ and σ: mean age and standard deviation. UK Biobank demographics are not reported in this table since cases correspond to proxy-AD phenotype mostly relying on self-report of first-degree relatives' diagnosis without age-at-onset being specified.

Sample	DX	N	*APOE ε2/ε2*		*APOE ε2/ε3*		*APOE ε3/ε3*		*APOE ε2/ε4*		*APOE ε3/ε4*		*APOE ε4/ε4*	
			N (%Females)	**Age μ(σ)**	**N (%Females)**	**Age μ(σ)**	**N (%Females)**	**Age μ(σ)**	**N (%Females)**	**Age μ(σ)**	**N (%Females)**	**Age μ(σ)**	**N (%Females)**	**Age μ(σ)**
ADSP	CN	11,934	73(54.8%)	82.6(8.3)	1481(62.4%)	83.0(8.0)	7429(62.4%)	82.3(8.1)	195(70.3%)	79.8(8.9)	2561(62.1%)	79.7(8.2)	195(63.1%)	76.6(7.5)
	AD	11,868	29(58.6%)	82.5(6.9)	583(63.3%)	80.1(9.7)	5313(60.9%)	77.0(10.1)	258(61.2%)	75.3(8.2)	4919(58.0%)	73.2(8.5)	766(53.0%)	67.9(8.1)
ADGC	CN	8,059	56(46.4%)	79.1(10.2)	978(64.3%)	76.2(9.5)	4795(61.9%)	74.5(9.4)	209(63.2%)	73.8(10.1)	1847(61.9%)	71.4(10.1)	174(60.9%)	68.7(9.3)
	AD	7,768	10(60.0%)	72.5(8.2)	323(56.0%)	75.8(10.4)	2494(63.6%)	74.7(10.5)	237(63.3%)	75.7(8.8)	3258(63.2%)	73.0(8.6)	1446(57.4%)	69.7(7.2)
EADB-core	CN	21,160	121(59.5%)	68.6(13.2)	2503(58.2%)	66.8(15.1)	13365(57.8%)	67.0(14.5)	396(55.6%)	66.7(13.3)	4390(55.7%)	66.3(13.6)	385(55.1%)	64.2(12.6)
	AD	19,873	27(51.9%)	76.4(11.7)	877(59.5%)	74.2(11.2)	8285(61.9%)	72.9(11.0)	435(66.0%)	73.2(10.7)	8003(63.0%)	71.7(9.7)	2246(57.4%)	67.6(8.8)
GR@ACE	CN	8,539	33(57.6%)	53.1(17.6)	858(52.2%)	57.5(18.7)	6005(50.1%)	56.7(18.0)	99(49.5%)	56.7(17.6)	1459(49.8%)	56.7(17.6)	85(43.5%)	54.9(14.8)
	AD	7,355	16(84.6%)	84.6(3.5)	389(70.4%)	81.4(8.1)	3840(70.4%)	80.9(7.9)	115(73.0%)	78.7(7.4)	2590(69.8%)	78.7(7.4)	405(64.7%)	74.8(7.3)
EADI	CN	6,331	38(52.6%)	82.6(7.5)	772(59.2%)	81.0(7.5)	4247(60.8%)	80.1(7.7)	109(60.6%)	78.8(7.1)	1106(59.2%)	79.0(7.6)	59(71.2%)	77.1(6.7)
	AD	2,397	7(85.7%)	79.3(6.0)	128(68.8%)	78.0(10.8)	1078(65.3%)	76.5(10.6)	71(59.2%)	73.4(8.8)	888(66.0%)	72.6(9.2)	225(64.9%)	68.1(7.0)
GERAD	CN	7,007	47(55.3%)	49.3(11.0)	853(50.1%)	51.5(12.6)	4127(51.9%)	50.9(11.9)	180(51.7%)	49.8(10.9)	1627(51.8%)	49.9(10.9)	173(49.7%)	49.9(11.0)
	AD	2,989	10(60.0%)	81.2(9.7)	140(62.9%)	79.3(11.3)	1092(62.0%)	79.3(9.6)	90(63.3%)	80.4(7.6)	1306(64.2%)	77.7(8.9)	351(62.4%)	74.2(8.4)
DemGene	CN	5,911	32(34.4%)	68.7(11.2)	685(49.1%)	69.2(12.4)	3236(47.6%)	68.9(11.0)	167(45.5%)	70.6(10.6)	1595(48.2%)	67.3(10.5)	196(44.4%)	64.7(11.0)
	AD	1,687	5(40.0%)	74.0(1.4)	72(58.3%)	71.6(10.6)	537(66.9%)	73.7(9.6)	43(72.1%)	75.4(7.0)	769(66.6%)	72.2(8.4)	261(61.7%)	69.3(8.1)
CCHS & CGPS	CN	101,995	705(54.9%)	57.0(13.2)	12818(55.1%)	57.6(13.6)	57115(54.8%)	57.5(13.4)	2936(55.4%)	56.8(13.0)	25616(54.9%)	56.7(12.8)	2778(57.6%)	55.3(12.7)
	AD	2,092	12(50.0%)	72.6(5.3)	129(53.5%)	73.3(8.4)	844(58.8%)	73.3(8.4)	70(61.4%)	71.2(8.0)	821(62.4%)	70.9(8.0)	216(56.9%)	68.8(7.9)

Table 2. V236E and R251G are associated with a decreased AD risk. The significance of their association with AD risk is equivalent in non-stratified analyses adjusted by *APOE* ε2 and ε4 dosages, and in *APOE*-stratified analysis considering the main *APOE* genotype group with the most carriers for each variant, namely ε3/ε3 and ε3/ε4 respectively for V236E and R251G.

N: Number of individuals, MAC: Minor allele count, OR: odds ratio, 95% CI: 95% confidence interval, P: p-value.

	Sample	AD Case-Control Regression Non-stratified				AD Case-Control Regression *APOE*-Stratified			
		N	MAC	OR [95% CI]	P	N	MAC	OR [95% CI]	P
V236E (all APOE (left) and ε3/ε3 only (right))	*ADSP*	23,427	20	0.23 [0.09; 0.56]	1.4E-03	12,604	17	0.31 [0.12; 0.82]	0.020
	ADGC imputed	11,652	11	0.35 [0.08; 1.51]	0.16	5,741	10	0.40 [0.1; 1.57]	0.19
	EADB-core	41,033	27.17	0.59 [0.19; 1.80]	0.34	21,650	21.28	0.53 [0.15; 1.92]	0.30
	GERAD	9,996	17.72	0.37 [0.07; 1.90]	0.18	5,219	9.43	0.77 [0.10; 6.06]	0.78
	DemGene	7,598	58.68	0.21 [0.05; 0.90]	8.5E-03	3,773	35.88	0.56 [0.13; 2.46]	0.40
	CCHS & CGPS	104,084	240	0.45 [0.11; 1.84]	0.23	57,955	191	0.18 [0.01; 2.97]	0.27
	UKB proxy-AD	185,741	277	0.45 [0.23; 0.89]	0.021	109,120	219	0.47 [0.21; 1.04]	0.063
	Meta-analysis	**383,531**	**649.57**	**0.37 [0.25; 0.56]**	**1.9E-06**	**216,062**	**503.59**	**0.43 [0.27; 0.69]**	**4.4E-04**
R251G *(all APOE (left) and ε3/ε4 only (right))*	*ADSP*	23,314	26	0.20 [0.08; 0.49]	3.7E-04	7,335	18	0.17 [0.06; 0.48]	7.8E-04
	ADGC imputed	14,134	29	0.29 [0.12; 0.70]	5.8E-03	4,630	16	0.19 [0.07; 0.54]	1.7E-03
	EADB-core	41,033	59.16	0.51 [0.26; 0.99]	0.049	12,393	40.27	0.34 [0.15; 0.76]	7.8E-03
	GR@ACE	15,894	21.27	0.35 [0.12; 1.01]	0.049	4,049	17.81	0.22 [0.06; 0.77]	0.011
	EADI	8,728	19.21	0.68 [0.22; 2.09]	0.49	1,994	13.32	1.14 [0.32; 4.04]	0.84
	GERAD	9,996	23.17	0.50 [0.17; 1.47]	0.18	2,933	16.82	0.57 [0.18; 1.88]	0.34
	CCHS & CGPS	104,087	105	0.41 [0.10; 2.72]	0.23	26,437	75	0.33 [0.05; 2.43]	0.28
	UKB proxy-AD	185,735	335	0.57 [0.34; 0.98]	0.041	43,820	262	0.67 [0.36; 1.22]	0.19
	Meta-analysis	**402,921**	**617.81**	**0.44 [0.33; 0.59]**	**4.7E-08**	**103,591**	**459.22**	**0.41 [0.29; 0.57]**	**3.2E-07**

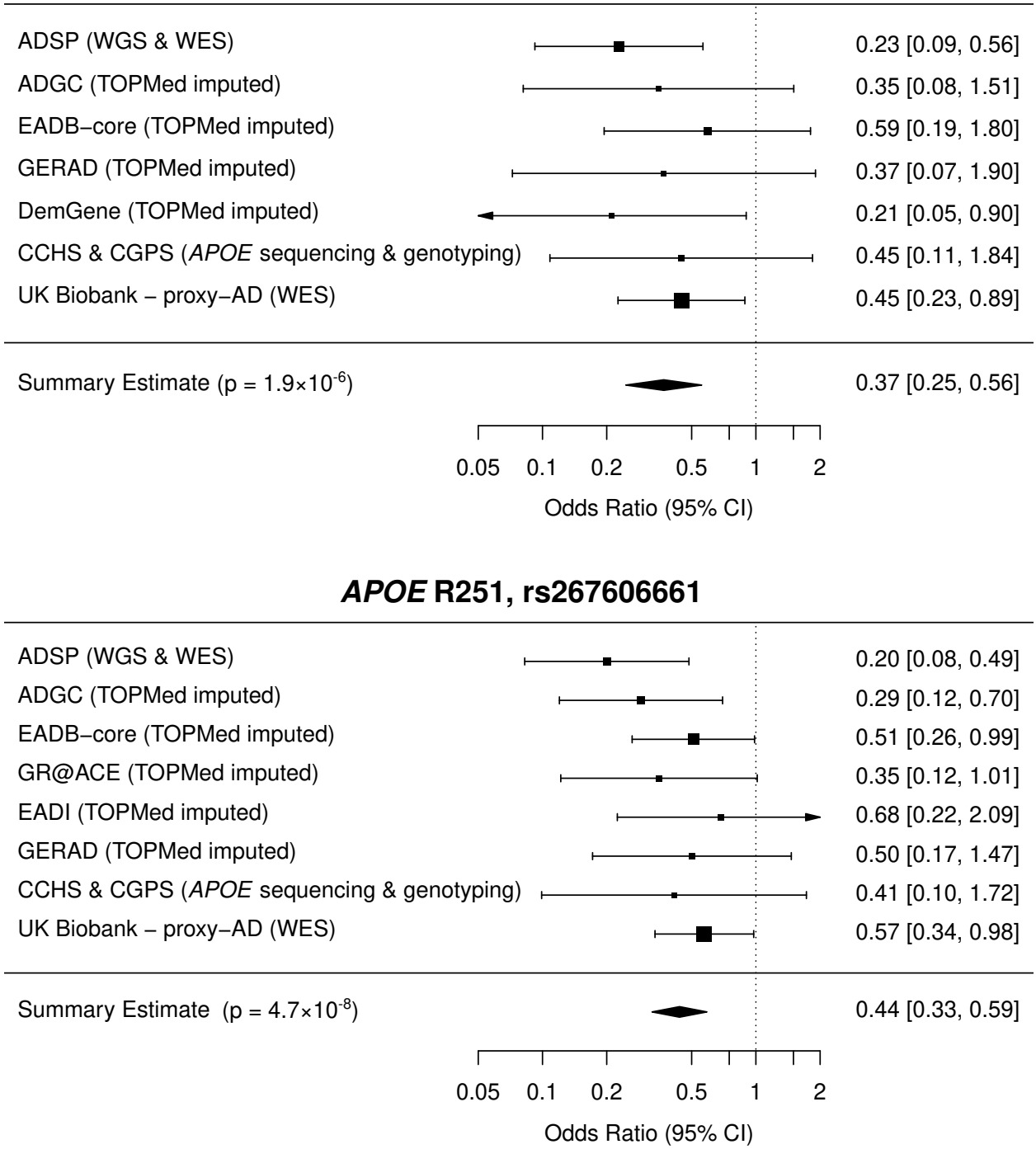

Figure 1. V236E and R251G are associated with decreased AD risk across all cohorts. Forest plots show the results for the non-*APOE* stratified analyses adjusted by ε2 and ε4 dosages. **eFigure 1** presents equivalent forest plots for these two variants in the *APOE*-stratified sensitivity analyses, showing consistent findings.

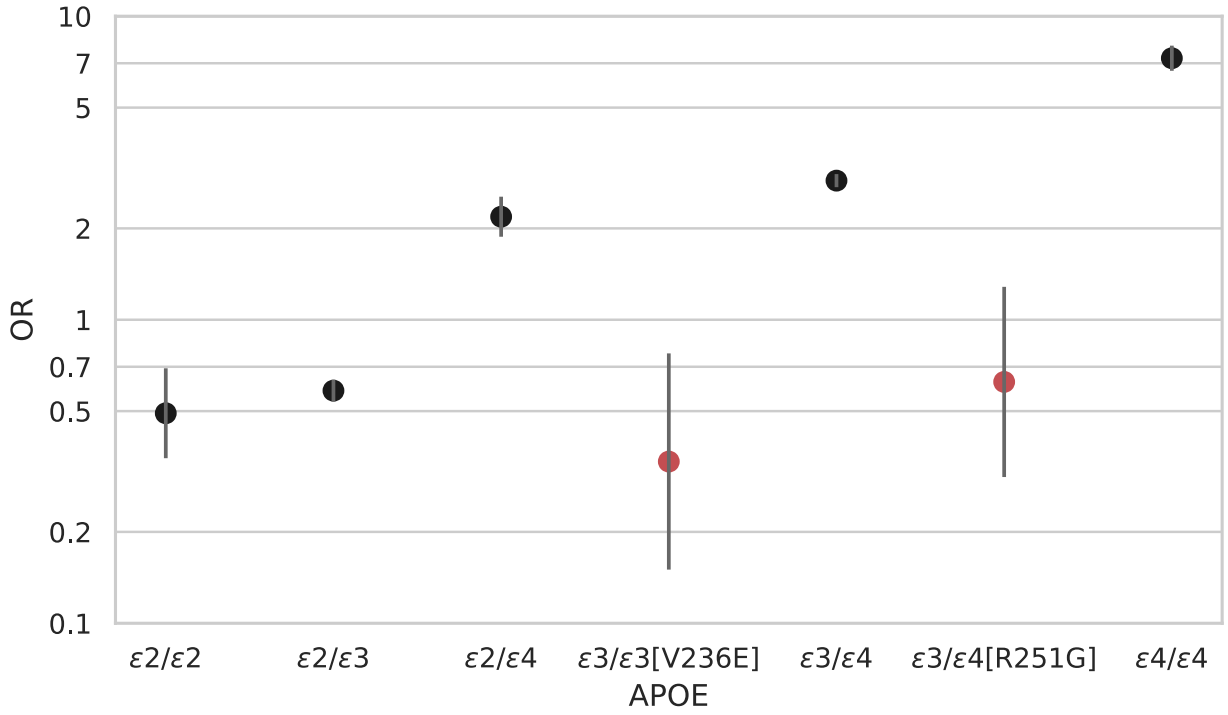

Figure 2. *APOE* ε3/ε3[V236E] and *APOE* ε3/ε4[R251G] have a risk equivalent to ε2/ε3 carriers. Alzheimer's disease (AD) risk per *APOE* genotype was compared to the *APOE* ε3/ε3 reference group (i.e., odds ratio (OR) for *APOE* ε3/ε3 equals to 1), meta-analyzing results from the ADSP and ADGC cohorts (Stages 1+2). **eFigure 2** presents equivalent results at different inclusion cutoffs for European ancestry.