Identification of Rare Damaging Missense and Loss of Function Variants in GWAS Loci Using Genome Sequencing Data from Two Cohorts of Familial Late-Onset Alzheimer's Disease

Tamil Iniyan Gunasekaran, PhD[1], Dolly Reyes-Dumeyer, BS[1], Kelley M. Faber, MS[2], Alison Goate, DPhil[3], Brad Boeve, MD[4], Carlos Cruchaga, PhD[5], Margaret Pericak-Vance, PhD[6], Jonathan L. Haines, PhD[7], Roger Rosenberg, MD[8], Debby Tsuang, MD[9], Diones Rivera Mejia, MD[10,11], Martin Medrano, MD[12], Rafael A. Lantigua, MD[1,13], Robert A. Sweet, MD[14], David A. Bennett, MD[15], Robert S. Wilson, PhD[15], Tatiana Foroud, PhD[2], Badri N. Vardarajan, PhD[1*], Richard Mayeux, MD[1*]

Affiliations:

[1]Department of Neurology, Taub Institute for Research on Alzheimer's Disease and the Aging Brain and the Gertrude H. Sergievsky Center, Columbia University in the City of New York, NY, USA

[2]Department of Medical and Molecular Genetics, National Centralized Repository for Alzheimer's Disease and Related Dementias (NCRAD), Indiana University School of Medicine, Indianapolis, IN, USA

[3]Department of Genetics & Genomic Sciences, Ronald M. Loeb Center for Alzheimer's disease, Icahn School of Medicine at Mount Sinai, New York, NY, USA

[4]Department of Neurology, Mayo Clinic, Rochester, MN, USA

[5]Department of Psychiatry, Washington University in St. Louis, St. Louis, MO, USA

[6]John P Hussman Institute for Human Genomics, Dr. John T Macdonald Foundation Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL, USA

[7]Department of Population & Quantitative Health Sciences and Cleveland Institute for Computational Biology. Case Western Reserve University, Cleveland, OH, USA

[8]Department of Neurology, University of Texas Southwestern Medical Center at Dallas, TX, USA

[9]GRECC VA Puget Sound, Department of Psychiatry and Behavioral Sciences, University of Washington Seattle, WA, USA

[10]CEDIMAT, Santo Domingo, Dominican Republic

[11]Universidad Pedro Henríquez Urena, Santo Domingo, Dominican Republic

[12]Pontíficia Universidad Católica Madre y Maestra (PUCMM), Santiago, Dominican Republic.

[13]Department of Medicine, Vagelos College of Physicians and Surgeons, Columbia University, and the New York Presbyterian Hospital, New York, NY, USA

[14]Departments of Psychiatry and Neurology, University of Pittsburgh, Pittsburgh, PA, USA

[15]Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, USA

* Contributed equally to the manuscript.

Correspondence:
Richard Mayeux, MD
Columbia University
710 West 168th Street
New York, NY 10032
(T) 212 305-2391
Email: rpm2@cumc.columbia.edu

Abstract

PURPOSE. Few rare pathogenic variants have been identified in the 70+ genetic loci from genome wide association studies of late-onset Alzheimer's disease (AD), limiting research on underlying mechanisms, risk assessment, and genetic counseling.

METHODS. Using genome sequencing data from 197 families in The National Institute on Aging Alzheimer's Disease Family Based Study (AD-FBS), and 214 families in The Estudio Familiar de la Influencia Genética en Alzheimer (EFIGA), we characterized rare coding variants predicted to highly damaging missense or loss of function variants (LoF) within known GWAS loci.

RESULTS. Eight coding and one LoF variant segregated in 10 (5.1%) AD-FBS families and 16 coding and two LoF variants segregated in 18 (8.4%) EFIGA families. *ABCA7 and AKAP9* contained the most damaging variants. In 51 (25.9%) of the AD-FBS and in 26 (12.1%) of the EFIGA families, *APOE-ε4* was the only variant segregating with familial AD (fAD). Neither *APOE-ε4* nor missense or LoF variants were found in 44.1% of the AD-FBS and 62.1% of the EFIGA families.

CONCLUSIONS. Although rare variants were found in both family groups, many families had no gene variant segregating within the family, indicating that the genetic basis for AD has yet to be fully defined.

INTRODUCTION

In the United States (US) in 2023, Alzheimer's disease (AD) was present in 6.7 million people, representing 10.8% of the population over age 65 years, and is projected to grow to 13.8 million people, nearly 3.3% of the total US population by 2060.[1] The cause(s) of AD remain elusive, but there is strong evidence that it is a polygenic disorder. Highly penetrant pathogenic variants in AD: *APP,*[2,3] *PSEN1,*[4]and *PSEN2*[5] explain less than 5% of early-onset familial AD (fAD). The *APOE-ε4* allele is associated with the risk of developing late onset, sporadic AD,[6,7] and has been replicated in most ethnic groups.

Late-onset AD is highly heritable (~70%), implying that genetic factors contribute to its etiology. Large-scale genome wide association studies (GWAS) of thousands of individuals with AD and non-demented controls led by the Alzheimer's Disease Genetic Consortium, the Alzheimer's Disease Sequencing Project, and the International Genomics of Alzheimer's Project, identified more than 70 genetic risk loci.[8,9] None of these risk loci reaches the effect size of *APOE-ε4*, nor do they fully explain fAD.

One of the most frequent questions posed by family members of patients with fAD concerns familial risk. Polygenic risk scores have recently been used to assess recurrence risk and penetrance in families.[10] However, identification of pathogenic variants in fAD, would advance mechanistic and therapeutic research efforts and facilitate genetic counseling. To identify and determine the frequency of potentially pathogenic variants in fAD, we used genome sequencing data from families recruited in the US and the Dominican Republic to identify

damaging missense and loss of function (LoF) variants in candidate genes nominated from one of the largest GWAS analyses of clinically and pathologically diagnosed AD.[8]

METHODS

National Institute on Aging AD Family-Based Study (AD-FBS). Beginning in 2002, families were recruited from Alzheimer's Disease Centers across the United States. Inclusion criteria for required one member to have a diagnosis of definite or probable AD[11] with onset after age 60 and a sibling with definite, probable, or possible AD and a similar age at onset. Another relative of the affected sibling pair, 60 years of age or older if unaffected, or if affected diagnosed with AD, was required. Any relatives over age 50 years were also recruited, regardless of cognitive status, but individuals deemed unaffected were required to have had documented cognitive testing and clinical examination. Every participant provided informed consent. Demographic information, previous clinical diagnoses, age at onset for patients with AD, method of diagnosis, Clinical Dementia Rating Scale[12] was ascertained. The age at onset for participants was defined as the age at which the family first observed signs of impaired cognition. For unaffected family members, we used their age at the time of their latest examination without impairment. Participants with advanced disease or those living in a remote location that could not complete a detailed in-person evaluation, contributed a blood sample, and the site investigator conducted a detailed review of medical records or an assessment by telephone or videoconference to document the presence or absence of AD. We used the NINCDS-ADRDA criteria[11] for definite AD or CERAD pathological criteria[13] for AD when based on postmortem information alone. Individuals with other forms of dementia, mild cognitive impairment or uncorroborated family reports of AD were considered as unaffected. From a pool of 251 families with 1,055 individuals,

we restricted the selection to families with two or more affected family members by clinical or pathological diagnosis, which yielded 197 families with 926 individuals and WGS data. There were 545 (58.8%) individuals with fAD and 381 individuals without dementia. The mean age at onset was 75.2 (s.d.8.04) years of age and they were white, non-Hispanic. Those affected were significantly older than those unaffected (Table 1a).

Estudio Familiar de la Influencia Genética en Alzheimer (EFIGA) cohort. We acquired data from a second set of families with late onset fAD from the EFIGA cohort to identify of variants segregating among families of Caribbean Hispanic ancestry. Study design, adjudication, and clinical assessment of AD in this cohort was previously described[14] and similar to that used in the AD-FBS. All individuals provided informed consent. Participants were followed every two years and evaluated using a neuropsychological battery, a structured medical and neurological examination, an assessment of depression,[15] The Clinical Dementia Rating Scale,[12] and the research clinical diagnosis of AD was determined. There were 271 families with 1,481 individuals but restricting the selection of families to those with two or more affected family members, 214 (78.9%) families remained with 1,340 individuals with available genome sequencing. There were 793 (59.2%) individuals who met criteria for clinical fAD and 547 individuals without dementia. The mean age of onset was 71.2 years (s.d.9.9). Those affected were significantly older that those unaffected (Table 1b).

Genome Sequencing. 2,535 individuals from 522 families underwent genome sequencing at the New York Genome Center using one microgram of DNA, an Illumina PCR-free library protocol, and sequenced on the Illumina HiSeq platform. We assessed sex mismatches and

unexpected familial relationships prior to analyses. We harmonized the WGS from the EFIGA and AD-FBS families, and jointly called variants to create a uniform, analysis set. Genomes were sequenced to a mean coverage of 30x. Sequence data analysis was performed using an automated analysis pipeline that matched the CCDG and TOPMed recommended best practices.[16] Briefly, sequencing reads were aligned to the human reference, hs38DH, using BWA-MEM v0.7.15. Variant calling was performed using the GATK best-practices. Variant filtration was performed using Variant Quality Score Recalibration (VQSR at tranche 99.6%) which identified annotation profiles of variants that were likely to be real and assigns a score (VQSLOD) to each variant. Variants passing the VQSR threshold were further filtered out for sample missingness (>2%), depth of coverage-DP<10 and genotype quality-GQ>20. We then annotated high quality variants using ANNOVAR. Specifically, variants were annotated for population level frequency using Genome Aggregation Database (gnomAD), *in-silico* function using Variant Effect Predictor (VEP) and variant conservation using Combined Annotation Dependent Depletion score (CADD).

Segregation analysis and identification of damaging missense and loss of function variants. The search was limited to variants in genetic loci identified in the largest GWAS where individuals were clinically or pathologically examined.[8] We also searched genes involved in early onset fAD: *APP, PSEN1 and PSEN2*, and variants within genes identified from sequencing studies: *TREM2*[17]and *AKAP9.*[18] The total number of genes or loci queried was 37 (Supplementary Table 1). Segregation of variants was defined by presence in all affected family members and in an unaffected family member if they were at least five years younger than the mean age of onset of AD in the family. Damaging coding variants were prioritized: loss of function (LoF) variants

(stoploss, stopgain, frameshift or splice variants) or damaging missense with a high CADD >20. Only variants considered rare in the general population (less that 1% in gnomAD) were included. The high allelic frequency of *APOE-ε4* in the general population and in AD, required examination of the co-segregation within families with and without a segregation of damaging missense or LoF variants.

RESULTS

In the AD-FBS, 926 individuals (58.9%) met criteria for definite or probable fAD (19.6% vs 80.4%, respectively) and the remaining family members were unaffected. In EFIGA, there were 1,340 individuals in this family group, of which 59.1% met criteria for probable fAD and the remaining individuals were considered unaffected. Other demographics provide age at recruitment, age at onset and sex (Tables 1a & 1b).

Presence of damaging missense and LoF variants. We identified 1,599 total coding variants in the two cohorts. We restricted the search to damaging potentially pathogenic coding variants using LoF variants or missense variants with CADD>20 (Figure 1). Nine damaging coding variants or loss of function variants were identified in seven genes among nine white, non-Hispanic families in AD-FBS. In the EFIGA there were 18 coding variants in 13 genes. *ABCA7* and *AKAP9* contained the highest number of variants. No damaging missense variant was present in both family datasets, but a rare LoF frameshift variant in *ABCA7*, (Ch19:1044708, GGGGCACCTGGT:G) was present in both family datasets. A full list of the coding variants in both family datasets are provided in Table 2. Ten damaging missense variants in AD-FBS and 25 in EFIGA also showed incomplete segregation in families (Supplementary Table 2).

APOE co-segregation. *APOE-ε4* co-segregated with damaging missense variants in one of the white, non-Hispanic and three (1.4%) of Hispanic families, but with none of the LoF variants. The presence of *APOE ε4* was associated with an earlier age at onset by 5 years or more (Table 3). Among the AD-FBS cohort, 79 (40.1%) families had only segregation with *APOE-ε4*, and in the EFIGA cohort, 35 (16.4%) families had only *APOE-ε4* segregation.

Linkage disequilibrium (LD) between common and rare variants. None of the damaging missense variants or LoF variants were linkage disequilibrium with the most significant common variant from the GWAS[8] or with any of the five other candidate genes. Because LD is sensitive to allele frequency, we anticipated that the likelihood of a common and rare variant being in LD would be very low. Thus, we also investigated whether the most significant common variant from the GWAS data co-segregated with any rare variant we identified. Within AD-FBS families, we identified three damaging missense variants: *ABCA7* (chr19:1056493:G:C), *ADAMTS1 (chr21:26784537:C:A),* and *SPPL2A* (chr15:50716490:T:C) that completely co-segregated with the common variant (Table 3). Within EFIGA families, five damaging missense variants and one LoF variant completely co-segregated with the common variants: *ADAM10* (chr15:58753575:A:G), *ADAMTS1* (chr21:26784537:C:A), *ADAMTS20* (chr12:43573874:T:G), *CR1* (chr1:207629207:A:C), *NYAP1* (chr7:100494172:T:C), *ECHDC3* (chr10:11678309:A:C).

Among the AD-FBS cohort, 106 (53.8%) families had no rare variants identified and were not found to have *APOE-ε4* accounting for the segregation within the families. Similarly, 160

(74.8%) of the EFIGA families had no rare variants identified and did not have *APOE-ε4* segregation.

DISCUSSION

Data from genome sequencing in two groups of families multiply affected by late onset fAD were used to identify rare potentially pathogenic, segregating variants embedded in or near risk loci reported in a large GWAS of unrelated individuals[8], present in Mendelian forms of early onset fAD[2-5] or identified by small sequencing investigations (*TREM2* and *AKAP9).*[17,18] We found numerous coding variants prior to restricting the analyses to damaging missense variants and LoF variants at these loci. In 10 AD-FBS families, variants in seven genes were identified as fully segregating in all affected family members and younger relatives at risk (Table 3). Similarly, in 18 EFIGA families, variants in 13 genes were observed segregating within those affected. The highest frequency of segregating variants was found in *ABCA7* and *AKAP9* in both groups of families.

In 25.9% of the AD-FBS families and 12% of the EFIGA families, there were no segregating variants in any of the GWAS loci and *APOE-ε4* was the only risk allele segregating in these families. Remarkably, 53.8% of AD-FBS families and 74.8% of the EFIGA families had no fully segregating variants in the AD GWAS loci nor were they found to have *APOE-ε4* segregating. This clearly indicates that there are addition pathogenic variants that have not been identified accounting for fAD.

GWAS were designed to investigate genetic variation across the genome associated with a common trait or disease. Most are common variants that confer small effects regarding the proportion of disease risk. Rare variants are defined by allele frequency of less than 0.01. They generally have a larger effect size compared with common variants. Rare variants can be the target for functional genomic analyses to understand disease mechanisms and analyses to investigate them as potential drug targets. They can also be used for genetic counseling or as putative biomarkers to support a personalized medicine.[19] While none of the rare variants identified in these two cohorts were in LD with the common variants, the most significant common variant in three genes (*ABCA7, ADAMTS1* and *SPPL2A*) co-segregated with rare damaging variants. The most significant common variant in six genes (*ABCA7, ADAMTS1, ADAMTS20, CR1, NYAP1,* and *ECHDC3*) co-segregated with rare damaging variants in the EFIGA families.

Rare variants were found in *ABCA7* and *AKAP9,* which are well established as genes involved in AD, although only *ABCA7* was first identified using common variants in a GWAS.[20] The ATP-binding cassette, sub-family A, member 7 (*ABCA7*) gene, has both common risk variants with functional consequences and rare, damaging coding variants. *ABCA7* is a large gene with 47 exons encoding a transmembrane transporter protein. *ABCA7* is expressed in brain, especially in microglia, neurons, endothelial cells, pericytes where it may help degrade Aβ.[21] It may also have an important role in lipid metabolism. Common variants in *ABCA7* associated with AD have been observed in black and white, non-Hispanic individuals as well as in Hispanics.[20,21] There have also been family data where premature termination codons have been found in *ABCA7*.[22]

Rare variants in A-kinase anchoring protein 9, *AKAP9*, were described in a small cohort of black, non-Hispanic individuals.[18] Subsequently, a rare variant (p.Arg434Trp) was observed in two large families from the Dominican Republic.[23] *AKAP9* contains 51 exons and is a binding protein that connects protein kinase to intracellular locations promoting cAMP related activation of PKA, which, in turn, enhances tau phosphorylation.[24] This process is one of the major findings in postmortem AD brains.

Variants in several other AD genetic loci were less frequent. Several lines of evidence have implicated the angiotensin converting enzyme, *ACE* in AD.[8,25] *ACE* encodes ACE-1, a rate-limiting enzyme in the cRAS pathway which generates disease-associated angiotensin II (Ang II) but also degrades and facilitates the clearance of Aβ. The gene encoding disintegrin-like and metalloproteinase with thrombospondin type 1 (*ADAMTS1*) is a member of the ADAMTS protein family. This family of genes are distinguished from *ADAM10* based on the number of copies of the thrombospondin 1-like repeats. Both are involved in extracellular matrix damage and repair.[26] *ADAMTS1* is thought to be involved in hydrolysis of APP and consequently decreases Aβ generation through inhibiting β-secretase-mediated cleavage. Clinically the gene may also affect cognitive ability by its effect on hippocampal amyloid processing. *CR1*, which encodes a complement receptor, regulating the complement cascade to clear cellular debris, including Aβ.[27] *ECHDC3* encodes an enzyme, enoyl CoA hydratase domain containing, and is a gene known to be involved in plasma lipids and other lipid related traits. Its expression is altered in the brains of patients with AD.[28] *EPHA1* belongs to the ephrin receptor subfamily of the protein-kinase family, and these receptors have been involved in development of the nervous system. In an early GWAS of AD, the Alzheimer's Disease Genetics Consortium

identified a common variant in *EPHA1* associated with ADGene Ontology annotations related to this gene include transferase activity, transferring phosphorus-containing groups and protein tyrosine kinase activity.[29] Variants in Neuronal Tyrosine Phosphorylated Phosphoinositide-3-Kinase Adaptor 1, *NYAP1*, appear to regulate immunoglobulin-like receptors beta and alpha (PILRB and PILRA) [30]. but it is unclear how these variants contribute to AD pathogenesis. Presenilin 1, *PSEN1,* a gene associated with early onset AD, has been previously observed in late onset, fAD[31,32], and it encodes the catalytic sub-unit of α-secretase involved in Aβ processing.

Variants in Protein Tyrosine Kinase 2 Beta**,** *PTK2B*, are thought to disrupt the normal suppression of tau phosphorylation and the effects of tauopathy thereby adding to the pathogenesis of AD.[33-35] Sortilin-related receptor with A-type repeats (SORLA or SORL1) is a sorting receptor in which variants impair processing of APP to soluble (s)APP and to the amyloid beta-peptide. SORL1 is also involved in endosomal degradation of Aβ and recycling in neurons.[36,37] Finally, a common variant in signal peptide peptidase-like 2A*, SPPL2A,* was identified in a large genome wide array, and the gene is thought to be part of the immune response.[38,39]

Rare variants have distinctive features, low likelihood of linkage disequilibrium with other variants in the region, more likely to affect gene function or expression, and a more direct role in diseases.[19] Efforts to discover associations driven by low-frequency, rare variants through genome sequencing efforts provide refinements to estimates of disease heritability.[40] Identifying rare variants in protein-coding regions through genome sequencing would also

greatly facilitate annotating genes associated with complex disease and describing the functional consequences of human sequence variation.

The reduction in costs for genetic testing will also make it more feasible to include in genome or exome sequencing in clinical practice. Strategies to identify damaging variants will soon be in place and clinicians and counselors will need to be able to identify and explain these results to patients and families. Families generally want to know the diagnostic relevance of our findings and the potential recurrence risk to other family. Therefore, a compelling argument can be made to accelerate efforts to identify rare variants in fAD because of the relative importance of turning these discoveries into meaningful clinical and biological progress.

Data Availability. All clinical data are available upon request by specific projects included in this manuscript: https://www.neurology.columbia.edu/research/research-centers-and-programs/alzheimers-disease-research-center-adrc/investigators/investigator-resources. All genetic data are available from the NIA Genetics of Alzheimer's Disease (**NIAGADS**) Data Sharing Service (DSS) hosts and shares data generated from the **ADSP** with the research community: https://dss.niagads.org/documentation/data-application-and-submission/application-instructions

Acknowledgements. The authors would like to thank the project managers at our respective sites for their efforts in recruitment and retention of these families.

Funding Statement. NIA AD-FBS: The National Institute on Aging Family Based Study supported the collection of samples used in this study through National Institute on Aging (NIA) grants U24AG026395, U24AG056270, and U24AG021886 (National Centralized Repository for Alzheimer's Disease). **EFIGA:** Data collection for this project was supported by the Genetic Studies of Alzheimer's disease in Caribbean Hispanics (EFIGA) funded by the National Institute on Aging (NIA) and by the National Institutes of Health (NIH) (5R37AG015473, RF1AG015473, R56AG051876, R01 AG067501).

Author Contributions. Conceptualization: B.N.V, T.F., R.M.; **Methodology:** B.N.V, R.M.; **Formal Analysis**: T.I.G., B.N.V.; **Investigation and Resources**: D.R-D., K.M.F, B.N.V, T.F., R.M.., A.G., B.B., C.C., M.P-V., J.L.H., R.R., D.T., D.R.M., M.M., R.A.L. R.A.S., D.A.B., R.S.W., T.F., R.M.; **Data Curation**: D.R-D., K.M.F, T.I.G., B.N.V.; **Writing Original Draft**: R.M., B.N.V.; **Writing Review and Editing:** B.N.V, T.F., R.M., D.R-D., K.M.F, R.A.S., D.A.B.; **Project Administration**: D.R-D., K.M.F, T.F., R.M.; **Funding Acquisition**: B.N.V, T.F., R.M.

Ethics Declaration. The authors assure that for the manuscript, “Identification of Rare Damaging Missense and Loss of Function Variants in GWAS Loci Using Genome Sequencing Data from Two Cohorts of Familial Late-Onset Alzheimer's Disease” the following is fulfilled: 1) This material is the authors' own original work, which has not been previously published elsewhere. 2) The paper is not currently being considered for publication elsewhere.

References

1. 2023 Alzheimer's disease facts and figures. Alzheimers Dement. 2023;19(4):1598-1695.
2. Chartier-Harlin MC, Crawford F, Houlden H, et al. Early-onset Alzheimer's disease caused by mutations at codon 717 of the beta-amyloid precursor protein gene. Nature. 1991;353(6347):844-846.
3. Goate A, Chartier-Harlin MC, Mullan M, et al. Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. Nature. 1991;349(6311):704-706.
4. Sherrington R, Rogaev EI, Liang Y, et al. Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. Nature. 1995;375(6534):754-760.
5. Levy-Lahad E, Wasco W, Poorkaj P, et al. Candidate gene for the chromosome 1 familial Alzheimer's disease locus. Science. 1995;269(5226):973-977.
6. 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.
7. Pericak-Vance MA, Bebout JL, Gaskell PC, Jr., et al. Linkage studies in familial Alzheimer disease: evidence for chromosome 19 linkage. Am J Hum Genet. 1991;48(6):1034-1050.
8. Kunkle BW, Grenier-Boley B, Sims R, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. Nat Genet. 2019;51(3):414-430.
9. Lambert JC, Ibrahim-Verbaas CA, Harold D, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet. 2013;45(12):1452-1458.

10. Qiao M, Lee AJ, Reyes-Dumeyer D, et al. Polygenic risk score penetrance & recurrence risk in familial Alzheimer disease. Ann Clin Transl Neurol. 2023;10(5):744-756.

11. McKhann GM, Knopman DS, Chertkow H, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 2011;7(3):263-269.

12. Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology. 1993;43(11):2412-2414.

13. Mirra SS, Heyman A, McKeel D, et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. Neurology. 1991;41(4):479-486.

14. Vardarajan BN, Faber KM, Bird TD, et al. Age-specific incidence rates for dementia and Alzheimer disease in NIA-LOAD/NCRAD and EFIGA families: National Institute on Aging Genetics Initiative for Late-Onset Alzheimer Disease/National Cell Repository for Alzheimer Disease (NIA-LOAD/NCRAD) and Estudio Familiar de Influencia Genetica en Alzheimer (EFIGA). JAMA Neurol. 2014;71(3):315-323.

15. Ruiz-Grosso P, Loret de Mola C, Vega-Dienstmaier JM, et al. Validation of the Spanish Center for Epidemiological Studies Depression and Zung Self-Rating Depression Scales: a comparative validation study. PLoS One. 2012;7(10):e45413.

16. Arora K, Shah M, Johnson M, et al. Deep whole-genome sequencing of 3 cancer cell lines on 2 sequencing platforms. Sci Rep. 2019;9(1):19123.

17. Guerreiro R, Wojtas A, Bras J, et al. TREM2 variants in Alzheimer's disease. N Engl J Med. 2013;368(2):117-127.

18. Logue MW, Schu M, Vardarajan BN, et al. Two rare AKAP9 variants are associated with Alzheimer's disease in African Americans. Alzheimers Dement. 2014;10(6):609-618.e611.

19. Momozawa Y, Mizukami K. Unique roles of rare variants in the genetics of complex diseases in humans. J Hum Genet. 2021;66(1):11-23.

20. Hollingworth P, Harold D, Sims R, et al. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. Nat Genet. 2011;43(5):429-435.

21. Reitz C, Jun G, Naj A, et al. Variants in the ATP-binding cassette transporter (ABCA7), apolipoprotein E ϵ4,and the risk of late-onset Alzheimer disease in African Americans. JAMA. 2013;309(14):1483-1492.

22. May P, Pichler S, Hartl D, et al. Rare ABCA7 variants in 2 German families with Alzheimer disease. Neurol Genet. 2018;4(2):e224.

23. Vardarajan BN, Barral S, Jaworski J, et al. Whole genome sequencing of Caribbean Hispanic families with late-onset Alzheimer's disease. Ann Clin Transl Neurol. 2018;5(4):406-417.

24. You Y, Hersh SW, Aslebagh R, et al. Alzheimer's disease associated AKAP9 I2558M mutation alters posttranslational modification and interactome of tau and cellular functions in CRISPR-edited human neuronal cells. Aging Cell. 2022;21(6):e13617.

25. Farrer LA, Sherbatich T, Keryanov SA, et al. Association between angiotensin-converting enzyme and Alzheimer disease. Arch Neurol. 2000;57(2):210-214.

26. Qiu Y, Sha L, Zhang X, Li G, Zhu W, Xu Q. Induction of A Disintegrin and Metalloproteinase with Thrombospondin motifs 1 by a rare variant or cognitive activities

reduces hippocampal amyloid-beta and consequent Alzheimer's disease risk. Front Aging Neurosci. 2022;14:896522.

27. Fonseca MI, Chu S, Pierce AL, et al. Analysis of the Putative Role of CR1 in Alzheimer's Disease: Genetic Association, Expression and Function. PLoS One. 2016;11(2):e0149792.

28. Yin J, Feng W, Yuan H, et al. Association analysis of polymorphisms in STARD6 and near ECHDC3 in Alzheimer's disease patients carrying the APOE epsilon4 Allele. Neuropsychiatr Dis Treat. 2019;15:213-218.

29. Naj AC, Jun G, Beecham GW, et al. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. Nat Genet. 2011;43(5):436-441.

30. Heath L, Earls JC, Magis AT, et al. Manifestations of Alzheimer's disease genetic risk in the blood are evident in a multiomic analysis in healthy adults aged 18 to 90. Sci Rep. 2022;12(1):6117.

31. Cruchaga C, Haller G, Chakraverty S, et al. Rare variants in APP, PSEN1 and PSEN2 increase risk for AD in late-onset Alzheimer's disease families. PLoS One. 2012;7(2):e31039.

32. Lee JH, Kahn A, Cheng R, et al. Disease-related mutations among Caribbean Hispanics with familial dementia. Mol Genet Genomic Med. 2014;2(5):430-437.

33. Beecham GW, Hamilton K, Naj AC, et al. Genome-wide association meta-analysis of neuropathologic features of Alzheimer's disease and related dementias. PLoS Genet. 2014;10(9):e1004606.

34. Brody AH, Nies SH, Guan F, et al. Alzheimer risk gene product Pyk2 suppresses tau phosphorylation and phenotypic effects of tauopathy. Mol Neurodegener. 2022;17(1):32.

35. Fernandez MV, Budde J, Del-Aguila JL, et al. Evaluation of Gene-Based Family-Based Methods to Detect Novel Genes Associated With Familial Late Onset Alzheimer Disease. Front Neurosci. 2018;12:209.

36. Holstege H, van der Lee SJ, Hulsman M, et al. Characterization of pathogenic SORL1 genetic variants for association with Alzheimer's disease: a clinical interpretation strategy. European Journal of Human Genetics. 2017;25(8):973-981.

37. Mishra S, Knupp A, Szabo MP, et al. The Alzheimer's gene SORL1 is a regulator of endosomal traffic and recycling in human neurons. Cell Mol Life Sci. 2022;79(3):162.

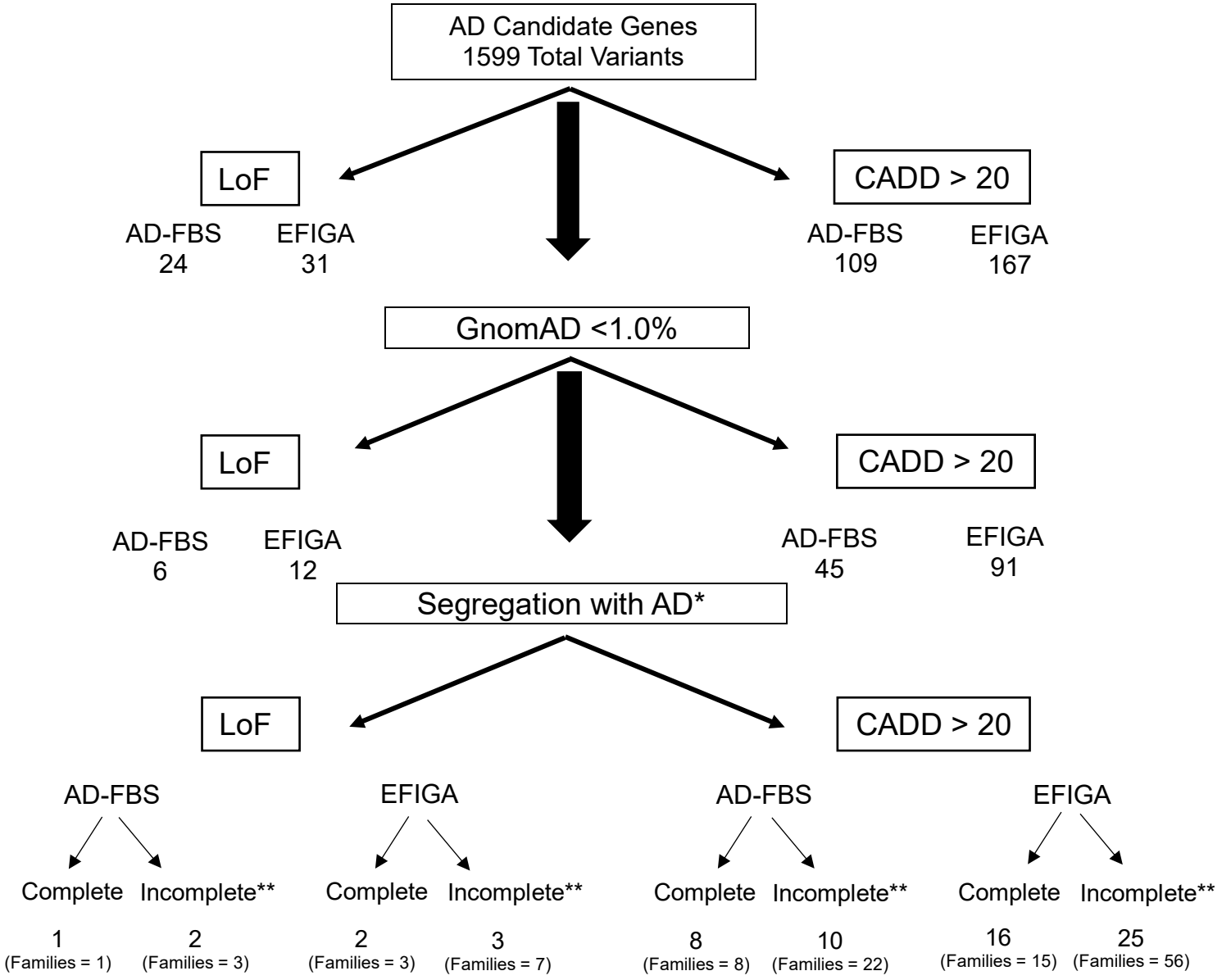
38. Bellenguez C, Kucukali F, Jansen IE, et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. Nat Genet. 2022;54(4):412-436.

39. Li QS, Tian C, andMe Research T, Hinds D, Seabrook GR. The association of clinical phenotypes to known AD/FTD genetic risk loci and their inter-relationship. PLoS One. 2020;15(11):e0241552.

40. Kryukov GV, Shpunt A, Stamatoyannopoulos JA, Sunyaev SR. Power of deep, all-exon resequencing for discovery of human trait genes. Proc Natl Acad Sci U S A. 2009;106(10):3871-3876.

Legend for figure 1. A filtering process was described in the methods. To identify rare variants (gnomAD less than 1%), degree of segregation (complete vs incomplete). Segregation was defined as being present in all affected members of a family. If a variant segregated in one family and showed incomplete segregation in another family, we required a qualifying variant to be seen only in affected. Unaffected family members with qualifying variants were accepted if their age was five or more years below the age at onset of the affected carriers. **Incomplete segregation was defined as variants present in at least four affected individuals with AD in the specific cohort (AD-FBS or EFIGA), observed in more affected than unaffected individuals, and if present in unaffected individuals if their age was five or more years below the age at onset of the affected carriers.

Figure 1. High impact (HI) coding and loss of function (LoF) variants identified in the AD-FBS and EFIGA Families.



Tables 1a and 1b. Distribution of participating families, age at onset and APOE-ε4 frequency.

1a. NIA AD-FBS	Total families	Total participants	Affected	Unaffected	Age Affected	Age Unaffected	APOE-ε4 (%)
Total Families	251	1055	594	461	75.19±8.03	71.61±9.72	0.38
<2 affected	54	79	60	19	76.18±6.17	81±2.94	0.42
>2 affected	197	926	545	381	75.21±8.04	70.56±9.41	0.38

1b. EFIGA	Total families	Total participants	Affected	Unaffected	Age Affected	Age Unaffected	APOE-ε4 (%)
Total Families	271	1480	833	647	71.97±9.85	66.54±8.77	0.33
<2 affected	57	435	145	290	80.25±8.66	81.28±6.76	0.27
>2 affected	214	1340	793	547	71.91±9.88	66.11±8.72	0.33

Table 2. Distribution of damaging missense and loss of function variants by cohort and variant type.

Gene		Cohort	Variant Type	Reference
ABCA7	chr19:1057344:G:A chr19:1051965:C:T chr19:1044708:GGGGCACCTGGT:G	AD-FBS EFIGA EFIGA/ AD-FBS	Missense Missense LoF	Transmembrane transporter protein [29,30]
ACE	chr17:63479054:G:C	EFIGA	Missense	Degradation and clearance of Ab [42,43]
ADAM10	chr15:58717601:C:T	EFIGA	Missense	constitutive α-secretase processing of APP[62,63]
ADAMTS1	chr21:26838049:G:A chr21:26838216:C:T	AD-FBS EFIGA	Missense Missense	APP hydrolysis decreased Aβ [9,45,64]
ADAMTS20	chr12:43356572:T:C chr12:43428519:A:G	EFIGA EFIGA	Missense Missense	A disintegrin-like and metalloproteinase with thrombospondin type 1 motifs[44]
AKAP9	chr7:91992989:G:C chr7:92097077:C:A chr7:92038422:A:G chr7:92040807:G:A chr7:92102769:G:A	AD-FBS AD-FBS EFIGA EFIGA EFIGA	Missense Missense Missense	Enhances tau phosphorylation[27,34,35]
CR1	chr1:207511589:C:T chr1:207523952:C:T	EFIGA EFIGA	Missense	Complement receptor C3b [65]
ECHDC3	chr10:11742730:C:T	EFIGA	LoF	Enoyl-CoA Hydratase Domain Containing 3[47,66]
EPHA1	chr7:143398406:G:A chr7:143391696:G:A	AD-FBS EFIGA	Missense Missense	Development of CNS[29,67]
NYAP1	chr7:100490675:C:T chr7:100488976:C:T	AD-FBS EFIGA	Missense Missense	Affects PILRB and PILRA immunoglobulin-like receptors[49]
PSEN1	chr14:73170945:C:T chr14:73170959:A:G	AD-FBS EFIGA	Missense Missense	Subunit of g-secretase[52]
PTK2B	chr8:27437198:G:A	EFIGA	Missense	Pyk2 suppresses Tau phosphorylation [53-55]
SORL1	chr11:121604205:C:T	EFIGA	Missense	Endosomal degradation and recycling in neurons[56-58,68]
SPPL2A	chr15:50722170:A:T	AD-FBS	Missense	Lysosomal membrane protein [59]

AD-FBS: Alzheimer's Disease - National Institute on Aging Family Based Study,
EFIGA: Estudio Familiar de la Genética de la Enfermedad.

Table 3: Details of variants that are completely segregating in at least one family

Variant Type	Cohort	SNP	Gene	Families variant completely segregates	AD patients in families where the variant completely segregates	Healthy members in families where the variant completely segregates*	Cosegre-gates with APOE	AD GWAS index SNP	Percentage of family members that also carry the GWAS index SNP
Frameshift variant	EFIGA	chr19:1044708:GGGGCACCTGGT:G	ABCA7	2	8	2	No	chr19:1056493:G:C	0.2
Stop Gain		chr10:11742730:C:T	ECHDC3	1	3	0	No	chr10:11678309:A:G	1
Frameshift Variant	AD-FBS	chr19:1044708:GGGGCACCTGGT:G	ABCA7	1	3	1	No	chr19:1056493:G:C	0
Damaging Missense (CADD>20)	EFIGA	chr19:1051965:C:T	ABCA7	1	2	4	No	chr19:1056493:G:C	0.17
		chr17:63479054:G:C	ACE	1	2	0	No	chr17:63460787:G:A	0
		chr15:58717601:C:T	ADAM10	1	2	2	No	chr15:58753575:A:G	1
		chr21:26838216:C:T	ADAMTS1	2	4	0	No	chr21:26784537:C:A	1
		chr12:43356572:T:C	ADAMTS20	1	5	3	No	chr12:43573874:T:G	0
		chr12:43428519:A:G	ADAMTS20	1	2	1	Yes	chr12:43573874:T:G	1
		chr7:92038422:A:G	AKAP9	1	2	0	No		
		chr7:92040807:G:A	AKAP9	1	2	0	No		
		chr7:92102769:G:A	AKAP9	1	2	0	No		

		chr1:207511589:C:T	CR1	1	2	1	Yes	chr1:207629207:A:C	1
		chr1:207523952:C:T	CR1	1	5	6	No	chr1:207629207:A:C	0.18
		chr7:143391696:G:A	EPHA1	1	2	3	No	chr7:143402040:C:A	0
		chr7:100488976:C:T	NYAP1	1	2	2	No	chr7:100494172:T:C	1
		chr14:73170959:A:G	PSEN1	1	5	1	No		
		chr8:27437198:G:A	PTK2B	1	2	0	No	chr8:27362470:C:T	0
		chr11:121604205:C:T	SORL1	1	3	0	Yes	chr11:121564878:T:C	0
	AD-FBS	chr19:1057344:G:A	ABCA7	2	5	4	No	chr19:1056493:G:C	1
		chr21:26838049:G:A	ADAMTS1	1	4	2	No	chr21:26784537:C:A	1
		chr7:91992989:G:C	AKAP9	1	2	0	Yes		
		chr7:92097077:C:A	AKAP9	1	3	2	No		
		chr7:143398406:G:A	EPHA1	1	3	1	No	chr7:143402040:C:A	0.5
		chr7:100490675:C:T	NYAP1	1	2	6	No	chr7:100494172:T:C	0.13
		chr14:73170945:C:T	PSEN1	1	2	0	No		
		chr15:50722170:A:T	SPPL2A	1	4	0	No	chr15:50716490:T:C	1