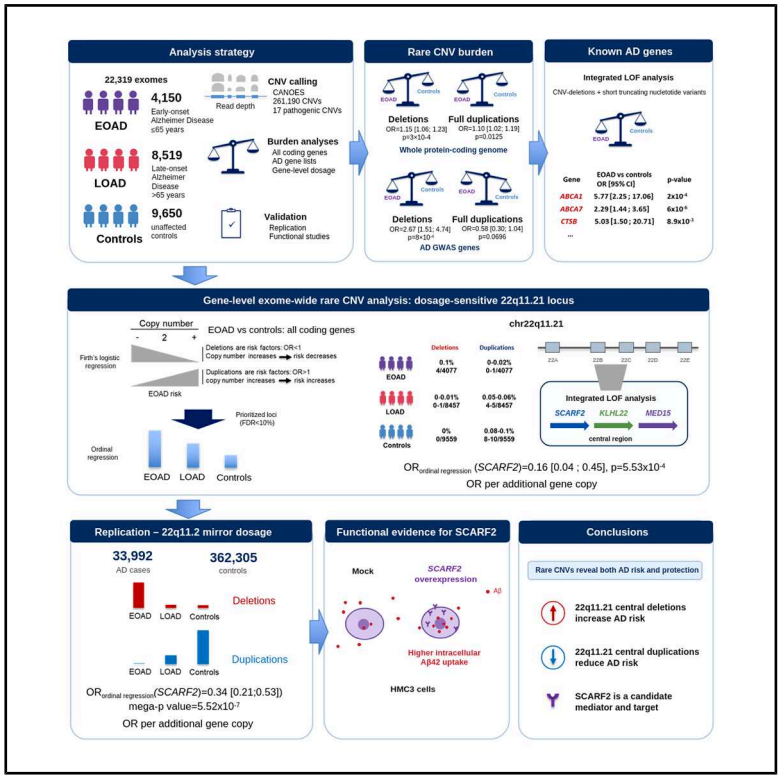
Exome analysis of 22,319 individuals links extremely rare copy-number variants and 22q11.21 dosage to Alzheimer risk

Graphical abstract



Authors

Olivier Quenez,
Catherine Schramm,
Kévin Cassinari, ...,
Sven J. van der Lee,
Camille Charbonnier, Gaël Nicolas

Correspondence

gaelnicolas@hotmail.com

Systematic analysis of rare exonic copy-number variants in Alzheimer disease identifies association with both known genes and a dosage-sensitive locus at 22q11.21, where deletions increase risk and duplications are protective. Convergent genetic and functional evidence prioritizes *SCARF2* as a candidate mediator, with overexpression enhancing amyloid-β uptake.

Quenez et al., 2026, The American Journal of Human Genetics *113*, 2099–2117
October 1, 2026

© 2026 The Authors. Published by Elsevier Inc. on behalf of American Society of Human Genetics.
https://doi.org/10.1016/j.ajhg.2026.07.013

Exome analysis of 22,319 individuals links extremely rare copy-number variants and 22q11.21 dosage to Alzheimer risk

Olivier Quenez,[1,89] Catherine Schramm,[2,89] Kévin Cassinari,[1] Aude Nicolas,[3] Joan Groeneveld,[4,5,6] Guillaume Huguet,[7] Benjamin Grenier-Boley,[3] Marc Hulsman,[4,5,6,8] G. Bragi Walters,[9] Itziar de Rojas,[10,11,12] Anne Rovelet-Lecrux,[1] Sébastien Feuillette,[1] Laetitia Miguel,[1] Anne-Claire Richard,[1] Stéphane Rousseau,[1] Shahzad Ahmad,[13,14] Najaf Amin,[13,15] Philippe Amouyel,[3] Olivia Belbin,[16,17] Céline Bellenguez,[3] Claudine Berr,[18] Paola Bossù,[19] Femke Bouwman,[6,8] Jose Bras,[20,21] Jordi Clarimon,[16,17] Antonio Daniele,[22] Jean-François Dartigues,[23] Stéphanie Debette,[23,24] Jean-François Deleuze,[25] Nicola Denning,[26] Oriol Dols-Icardo,[16,17] Cornelia M. van Duijn,[13,15] Juan Fortea,[16] Nick C. Fox,[27] Ruth Frikke-Schmidt,[28,29] Daniela Galimberti,[30,31] Roberta Ghidoni,[32] Vilmantas Giedraitis,[33] Johan J.P. Gille,[34] Detelina Grozeva,[35]

(Author list continued on next page)

Summary

Copy-number variants (CNVs) are major contributors to human disease. In Alzheimer disease (AD), *APP* duplications cause autosomal-dominant forms, but the role of CNVs in non-monogenic AD remains poorly characterized. We analyzed rare CNVs (frequency <1%) from 22,319 exomes (4,150 early-onset AD [EOAD, ≤65 years], 8,519 late-onset AD [LOAD], 9,650 unaffected control subjects) using harmonized calling and quality control. After identifying 17 individuals with a pathogenic CNV, we performed exome-wide and gene-set burden analyses. EOAD-affected individuals showed increased burdens of rare CNVs affecting coding genes, particularly deletions in AD-related genes. Integrated loss-of-function (LoF) analysis gathering short truncating variants with deletions showed that *ABCA1* (odds ratio [OR] = 5.77 [95% confidence interval 2.25; 17.06], $p = 0.0002$) and *ABCA7* deletions contribute to this deletion burden (OR = 2.29 [1.44; 3.65], $p = 0.0006$), while *CTSB* LoF alleles appear as candidates (OR = 5.03 [1.50; 20.71], $p = 0.0089$). We then performed exome-wide gene-level dosage analysis and highlighted 18 genes across five loci with a false discovery rate of <10%, including the 22q11.21 central region, where deletions were restricted to EOAD (including one *de novo* event) and duplications were enriched in control individuals, with intermediate frequencies in LOAD. We narrowed this locus to the *SCARF2-KLHL22-MED15* region after integrating short truncating variants. Replication in 33,977 affected individuals and 362,322 control subjects confirmed association for 22q11.21 dosage with exome-wide significance ($OR_{SCARF2} = 0.34$ [0.21; 0.53]; mega-p value $= 5.52 \times 10^{-7}$). *SCARF2* overexpression significantly increased amyloid-β uptake, congruent with duplication-associated decreased AD risk. We conclude that rare coding CNVs in a proportion of AD-associated genes and 22q11.21 deletions, including some found in DiGeorge syndrome, increase AD risk. Conversely, we identify 22q11.21 duplication as a strong AD-risk-decreasing factor.

Introduction

The etiology of Alzheimer disease (AD) includes autosomal-dominant AD (ADAD), caused by pathogenic variants in *PSEN1* (MIM: 104311), *PSEN2* (MIM: 600759), or *APP* (MIM: 104760) and representing <1% of all cases, and non-monogenic (or complex) AD, combining genetic and non-genetic factors. Genetic factors play a prominent

[1]Univ Rouen Normandie, Normandie Univ, Inserm U1245 and CHU Rouen, Department of Genetics and CNRMAJ, 76000 Rouen, France;[2]Univ Rouen Normandie, Normandie Univ, Inserm U1245, 76000 Rouen, 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, Lille, France;[4]Genomics of Neurodegenerative Diseases and Aging, Human Genetics, Vrije Universiteit Amsterdam, Amsterdam UMC location VUmc, Amsterdam, the Netherlands;[5]Alzheimer Center Amsterdam, Neurology, Vrije Universiteit Amsterdam, Amsterdam UMC location VUmc, Amsterdam, the Netherlands;[6]Amsterdam Neuroscience, Neurodegeneration, Amsterdam, the Netherlands;[7]Centre Hospitalier Universitaire Sainte-Justine Research Center, Montreal, QC, Canada;[8]Delft Bioinformatics Lab, Delft University of Technology, Delft, the Netherlands;[9]Amgen-deCODE Genetics, Sturlugata 8, 102, Reykjavik, Iceland;[10]Ace Alzheimer Center Barcelona, Universitat Internacional de Catalunya, Barcelona, Spain;[11]CIBERNED, Network Center for Biomedical Research in Neurodegenerative Diseases, National Institute of Health Carlos III, Madrid, Spain;[12]Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Esch-sur-Alzette, Luxembourg;[13]Department of Epidemiology, Erasmus Medical Centre, Rotterdam, the Netherlands;[14]Leiden Academic Centre for Drug Research, Leiden, the Netherlands;[15]Nuffield Department of Population Health, Oxford University, Oxford, UK;[16]Department of Neurology, II B Sant Pau, Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona, Barcelona, Spain;[17]Biomedical Research Networking Center on Neurodegenerative Diseases, National Institute of Health Carlos III, Madrid, Spain;[18]Université Montpellier, INSERM, Institute for Neurosciences of Montpellier, Montpellier, France;[19]Experimental Neuro-psychobiology Laboratory, Department of Clinical and Behavioral Neurology, Istituto di Ricovero e Cura a Carattere Scientifico Santa Lucia Foundation, Rome, Italy;[20]Department of Neurodegenerative Science, Van Andel Institute, Grand Rapids, MI, USA;[21]Division of Psychiatry and Behavioral Medicine, Michigan State University College of Human Medicine, Grand Rapids, MI, USA;[22]Department of Neuroscience, Catholic University of Sacred Heart,

(Affiliations continued on next page)

© 2026 The Authors. Published by Elsevier Inc. on behalf of American Society of Human Genetics.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Rita Guerreiro,[20,21] Edna Grünblatt,[36,37,38] John Hardy,[39] Steffi G. Riedel-Heller,[40] Mikko Hiltunen,[41] Clive Holmes,[42] Jakub Hort,[43] Holger Hummerich,[44] M. Arfan Ikram,[13] M. Kamran Ikram,[13] Martin Ingelsson,[33] Iris E. Jansen,[5,6,45] Amit Kawalia,[46] Robert Kraaij,[47] Patrick G. Kehoe,[48] Marc Lathrop,[49] Morgane Lacour,[50] Afina W. Lemstra,[5,6] Alberto Lleó,[16,17] Lauren Luckcuck,[35] Marcel M.A.M. Mannens,[51] Rachel Marshall,[35] Carlo Masullo,[52] Simon Mead,[44] Patrizia Mecocci,[53,54] Alexandre de Mendonça,[55] Alun Meggy,[26] Shima Mehrabian,[56] Merel O. Mol,[57] Kevin Morgan,[58] Alexandre Morin,[50] Benedetta Nacmias,[59,60] Penny J. Norsworthy,[44] Robert Olaso,[25] Florence Pasquier,[61] Pau Pastor,[62,63] Fabrizio Piras,[64] Julius Popp,[65] Alfredo Ramirez,[46,66,67,68,69] Rachel Raybould,[26] Richard Redon,[70] Marcel J.T. Reinders,[8] Fernando Rivadeneira,[47] Jeroen G.J. van Rooij,[47,71] Natalie S. Ryan,[27] Salha Saad,[35] Pascual Sanchez-Juan,[17,72] Nikolaos Scarmeas,[73,74] Philip Scheltens,[5,6]

(Author list continued on next page)

role in non-monogenic AD, and the etiology of non-monogenic early-onset AD (EOAD, onset at ≤65 years) shows an even greater genetic component.[1]

To date, most identified AD genetic determinants are single-nucleotide variants (SNVs) and short insertions/deletions (indels), from ADAD-causing pathogenic variants to a large diversity of risk factors in non-monogenic AD. Among the latter, common variants all show modest odds ratios (ORs) below 1.5,[2] except the *APOE-ε4* allele (MIM: 107741) with ORs ranging from ∼3 to ∼14 depending on ancestry and zygosity, despite a high frequency (9%–23%).[3] Multiple rare (allele frequency<1%) coding variants in five genes (*SORL1* [MIM: 602005], *TREM2* [MIM: 605086], *ABCA7* [MIM: 605414], *ABCA1*

Fondazione Policlinico Universitario A. Gemelli Istituto di Ricovero e Cura a Carattere Scientifico, Rome, Italy;[23]Université Bordeaux, INSERM, Bordeaux Population Health Research Center, Bordeaux, France;[24]Department of Neurology, Bordeaux University Hospital, Bordeaux, France;[25]Université Paris-Saclay, CEA, Centre National de Recherche en Génomique Humaine (CNRGH), 91057 Evry, France;[26]UKDRI Cardiff, School of Medicine, Cardiff University, Cardiff, UK;[27]Dementia Research Centre, University College London Queen Square Institute of Neurology, London, UK;[28]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark;[29]Department of Clinical Biochemistry, Rigshospitalet, Copenhagen, Denmark;[30]Fondazione Istituto di Ricovero e Cura a Carattere Scientifico C' Granda, Ospedale Policlinico, Milan, Italy;[31]University of Milan, Milan, Italy;[32]Molecular Markers Laboratory, IRCCS Istituto Centro San Giovanni di Dio Fatebenefratelli, Brescia, Italy;[33]Department of Public Health and Carins Sciences/Geriatrics, Uppsala University, Uppsala, Sweden;[34]Genome Diagnostics, Department of Human Genetics, VU University, AmsterdamUMC (location-VUmc), Amsterdam, the Netherlands;[35]Medical Research Council Centre for Neuropsychiatric Genetics and Genomics, Division of Psychological Medicine and Clinical Neuroscience, School of Medicine, Cardiff University, Cardiff, UK;[36]Department of Child and Adolescent Psychiatry and Psychotherapy, Psychiatric University Hospital Zurich, University of Zurich, Zurich, Switzerland;[37]Neuroscience Center Zurich, University of Zurich and ETH Zurich, Zurich, Switzerland;[38]Zurich Center for Integrative Human Physiology, University of Zurich, Zurich, Switzerland;[39]Reta Lila Weston Research Laboratories, Department of Molecular Neuroscience, University College London Institute of Neurology, London, UK;[40]Institute of Social Medicine, Occupational Health and Public Health, University of Leipzig, Leipzig, Germany;[41]Institute of Biomedicine, University of Eastern Finland, Kuopio, Finland;[42]-Clinical and Experimental Science, Faculty of Medicine, University of Southampton, Southampton, UK;[43]Memory Clinic, Department of Neurology, Second Faculty of Medicine, Charles University, Motol University Hospital, Prague, Czech Republic;[44]Medical Research Council Prion Unit at University College London, University College London Institute of Prion Diseases, London, UK;[45]Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, Vrije University, Amsterdam, the Netherlands;[46]Division of Neurogenetics and Molecular Psychiatry, Department of Psychiatry and Psychotherapy, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany;[47]Department of Internal Medicine, Erasmus Medical Centre, Rotterdam, the Netherlands;[48]Translational Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK;[49]McGill University and Genome Quebec Innovation Centre, Montreal, QC, Canada;[50]Univ Rouen Normandie, Normandie Univ, Inserm U1245 and CHU Rouen, Department of Neurology and CNRMAJ, 76000 Rouen, France;[51]Department of Human Genetics, Amsterdam UMC, University of Amsterdam, Amsterdam Reproduction and Development Research Institute, Amsterdam, the Netherlands;[52]Institute of Neurology, Catholic University of the Sacred Heart, Rome, Italy;[53]Division of Gerontology and Geriatrics, Department of Medicine and Surgery, University of Perugia, Perugia, Italy;[54]Division of Clinical Geriatrics, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden;[55]Faculty of Medicine, University of Lisbon, Lisbon, Portugal;[56]Department of Neurology, University Hospital "Alexandrovska," 1431 Sofia, Bulgaria;[57]Department of Clinical Genetics, Erasmus Medical Centre, Rotterdam, the Netherlands;[58]Human Genetics, School of Life Sciences, University of Nottingham, Nottingham, UK;[59]Department of Neuroscience, Psychology, Drug Research and Child Health, University of Florence, Florence, Italy;[60]IRCCS Fondazione Don Carlo Gnocchi, Florence, Italy;[61]Université Lille, INSERM, Centre Hospitalier Universitaire Lille, UMR1172, Resources and Research Memory Center (MRRC) of Distalz, Licend, Lille, France;[62]Fundació Docència i Recerca MútuaTerrassa and Movement Disorders Unit, Department of Neurology, University Hospital MútuaTerrassa, Barcelona, Spain;[63]Memory Disorders Unit, Department of Neurology, Hospital Universitari Mutua de Terrassa, Barcelona, Spain;[64]Laboratory of Neuropsychiatry, Department of Clinical and Behavioral Neurology, Istituto di Ricovero e Cura a Carattere Scientifico, Santa Lucia Foundation, Rome, Italy;[65]University Hospital of Psychiatry Zürich, University of Zürich, Zürich, Switzerland;[66]Department of Psychiatry and Glenn Biggs Institute for Alzheimer's and Neurodegenerative Diseases, San Antonio, TX, USA;[67]Department of Neurodegenerative Diseases and Geriatric Psychiatry, University Hospital Bonn, Medical Faculty, Bonn, Germany;[68]German Center for Neurodegenerative Diseases, Bonn, Germany;[69]Cluster of Excellence Cellular Stress Responses in Aging-Associated Diseases, University of Cologne, Cologne, Germany;[70]Université de Nantes, Centre Hospitalier Universitaire Nantes, Centre National de la Recherche Scientifique, INSERM, l'institut du Thorax, Nantes, France;[71]Department of Neurology, Erasmus Medical Centre, Rotterdam, the Netherlands;[72]Neurology Service, Marqués de Valdecilla University Hospital (University of Cantabria and IDIVAL), Santander, Spain;[73]Department of Neurology, Aiginition Hospital, National and Kapodistrian University of Athens, 11528 Athens, Greece;[74]Department of Neurology, Columbia University, New York, NY 10032, USA;[75]Laboratory for Advanced Hematological Diagnostics, Department of Hematology and Stem Cell Transplant, Lecce, Italy;[76]Centre for Neuropsychiatric Genetics and Genomics, Division of Psychological Medicine and Clinical Neuroscience, School of Medicine, Cardiff University, Cardiff, UK;[77]Department of Human Genetics and Amsterdam Reproduction and Development Research

(Affiliations continued on next page)

Jonathan M. Schott,[27] Davide Seripa,[75] Daoud Sie,[34] Rebecca Sims,[76] Erik A. Sistermans,[77] Sandro Sorbi,[59,60] Kristel Sleegers,[78,79,80] Resie van Spaendonk,[34] John C. van Swieten,[71] Niccolo' Tesi,[4,5,6,8] Betty M. Tijms,[5] Magda Tsolaki,[81,82] André G. Uitterlinden,[47] Jort Vijverberg,[5,6] Pieter Jelle Visser,[5] Michael Wagner,[67,68] Julie Williams,[83] Aline Zarea,[50] EADB Consortium, Emmanuelle Génin,[84] Henne Holstege,[4,5,6,8] Daniel F. Gudbjartsson,[9,85] David Wallon,[50] Magalie Lecourtois,[1] Maria Victoria Fernandez,[86] Hreinn Stefansson,[9] Sébastien Jacquemont,[7,87] Jean-Charles Lambert,[3] Sven J. van der Lee,[4,5,6] Camille Charbonnier,[88] and Gaël Nicolas[1,*]

[MIM: 600046], and *ATP8B4* [MIM: 609123]), mostly identified by exome sequencing, are associated with higher ORs, ranging from 1.5 to >20, and thus contribute significantly to the risk of AD in heterozygote individuals.[3,4] Recent findings from exome or genome-sequencing studies show that (1) the rarer the variants, the higher the ORs, and (2) rare variants with strong effects on AD risk are significantly enriched in EOAD compared to late-onset AD (LOAD).[4,5] In contrast with AD-risk-increasing variants, only a handful of variants have thus far been significantly associated with strong effects on AD risk reduction (OR < 0.7).

Copy-number variants (CNVs, deletions or duplications of >100-bp genomic DNA regions) represent a variant type with a major role in multiple diseases.[6–8] In AD, CNVs have been less studied than SNVs and indels. *APP* locus duplications on chromosome 21 are typical examples of CNVs causing ADAD,[9,10] accounting for ∼9% of all ADAD cases (update of Lanoiselée et al.[11]). Importantly, they also underscore why trisomy-21 individuals develop AD with full penetrance by age 70 years.[12] In addition, they demonstrate that increased expression of the amyloid-β (Aβ) precursor protein (APP) is sufficient to cause AD, further supporting the amyloid cascade stating that Aβ aggregation is a mandatory triggering factor of AD. Ultra-rare in-frame exon9/exon9–10 deletions of *PSEN1* also cause ADAD,[13] while a few common CNVs, detected by DNA chips, have emerged as candidate risk factors.[14,15] However, none of them reached genome-wide significance. Regarding rare CNVs, a handful of studies using DNA chips were published, on a limited number of EOAD-affected individuals,[16,17] before the advent of next-generation sequencing. DNA chips can only detect CNVs larger than 50–100 kb, thus missing smaller CNVs, which show a remarkable diversity of rare events.[18]

Bioinformatics advances now enable the detection of rare CNVs from sequencing data. Recently, structural variants—including CNVs—were detected from genome-sequencing data of 6,525 AD individuals (average age at onset 74.6 years, mainly LOAD) and 6,855 control individuals from the Alzheimer Disease Sequencing Project (ADSP) consortium.[19,20] A moderate burden of (coding and non-coding) ultra-rare CNVs was associated with AD status overall. These structural variants included a few examples of individual CNVs with a potential impact, including one partial deletion of *SORL1* and five partial deletions of *ABCA7*, each identified in AD-affected individuals. However, no gene-level association was assessed, and no CNV was significantly associated with AD, likely due to limited recurrence of CNVs and also possibly because this cohort was not enriched in EOAD cases, limiting power. This power limitation is intrinsic to rare-variant association studies, and this is even more pronounced for CNVs: whereas each individual carries hundreds of rare coding SNVs or indels, only one rare coding CNV is typically observed per genome, on average.[18] This makes it necessary to analyze larger datasets and/or enrich them in individuals with a higher suspected genetic component, such as EOAD-affected individuals. Despite the development of genome-sequencing datasets, a large proportion of the data available so far comprises exomes. While multiple algorithms are often combined for genome sequencing, exome-based approaches relying on specific read-depth tools have been shown to achieve high accuracy for rare coding CNVs without requiring multi-caller strategies.[21,22] For example, we previously assessed the performances of a CNV-calling workflow based on the CANOES tool,[23] a clinical-grade validated CNV workflow that showed good sensitivity (87.25%) and positive predictive value (86.4%–90.2%) for rare CNVs from exome-sequencing data.[24]

To identify CNVs associated with AD risk, we used one large sequencing dataset, i.e., the combination of the European and American consortia into the Alzheimer Disease European Sequencing-Alzheimer Disease Sequencing Project (ADES-ADSP) dataset, as it is rich in EOAD-affected

Institute, Amsterdam UMC, location Vrije Universiteit Amsterdam, Amsterdam, the Netherlands;[78]Complex Genetics of Alzheimer's Disease Group, VIB Center for Molecular Neurology, VIB, Antwerp, Belgium;[79]Laboratory of Neurogenetics, Institute Born-Bunge, Antwerp, Belgium;[80]Department of Biomedical Sciences, University of Antwerp, Antwerp, Belgium;[81]First Department of Neurology, Medical School, Aristotle University of Thessaloniki, Thessaloniki, Greece;[82]Alzheimer Hellas, Thessaloniki, Greece;[83]UK Dementia Research Institute, Cardiff University, Cardiff, UK;[84]Université Brest, INSERM, Etablissement Français du Sang, Centre Hospitalier Universitaire Brest, Unité Mixte de Recherche 1078, GGB, Brest, France;[85]School of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland;[86]Memory Clinic of Fundaciò ACE, Catalan Institute of Applied Neurosciences, 08028 Barcelona, Spain;[87]Department of Pediatrics, Université de Montréal, Montreal, QC, Canada;[88]Univ Rouen Normandie, Normandie Univ, Inserm U1245 and CHU Rouen, Department of Biostatistics and CNRMAJ, 76000 Rouen, France
[89]These authors contributed equally
*Correspondence: gaelnicolas@hotmail.com
https://doi.org/10.1016/j.ajhg.2026.07.013.

individuals (32.7%), thus enhancing power for detecting CNVs with a moderate-to-strong effect. Importantly, this dataset contains a large proportion of exome-sequencing data (68.6% after quality control [QC]), allowing the use of a single caller with a high level of accuracy. We performed an AD case-control study using CNV calling from 22,319 exomes and focused on rare CNVs. We first searched for pathogenic CNVs, then performed burden analyses at the protein-coding genome level and after restriction to AD-related gene lists, and finally performed a genome-wide case-control association study at the gene level, different from previous studies in the field. Prioritized loci were further assessed by performing an integrated loss-of-function (LoF) analysis combining short truncating SNV/indels with CNV deletions, a multi-ethnic replication analysis, and *in vitro* cellular assays.

Material and methods

Detailed methods are provided in supplemental information.

Exome dataset, CNV calling, QC, and pathogenic CNV identification

We included all 22,319 exomes available from Holstege et al.[4] collected as part of ADES (comprising European studies) and ADSP (comprising studies from the United States) (Table S2, with cohort details in Holstege et al.[4]). All samples were selected following the QC detailed in Holstege et al.,[4] except individuals with a pathogenic CNV, who were not excluded from our starting dataset.

Only samples passing the QC described in detail in Holstege et al.[4] were selected for our analysis. In brief, for the individual QC based on short sequence variants, all sequencing (Fastq) files were processed using the same BWA-GATK-based pipeline[25,26] on the Cartesius supercomputer embedded in the Dutch national e-infrastructure. Samples with either a high level of variant missingness, a high suspicion of DNA contamination, a discordant genetic sex annotation, a non-European ancestry, a high number of previously unidentified variants (compared to dbSNP v.150), a deviation from standard heterozygous/homozygous or transition/transversion ratios, and relatives up to the third degree were excluded. As affected subjects, we considered all individuals with a diagnosis of definite or probable AD (using the NIAA or NINCDS-ADRA criteria depending on the date of diagnosis[27,28]), based on clinical and paraclinical information including cerebrospinal fluid (CSF) AD biomarkers, when available. Individuals with unclear diagnosis were excluded (e.g., Braak stage I–II in affected individuals or Braak stage V–VI in control subjects). In addition, individuals with a pathogenic SNV or indel in a list of Mendelian dementia-associated genes were already excluded from stage 1 data in this dataset. We now applied the same analysis, leading to the exclusion of 22 additional individuals with a pathogenic variant among case and control groups from stage 2 so that the megasample from Holstege et al.[4] (stages 1 + 2) fulfilled the same criteria (Table S1). Detailed methods for individual QC are available as a supplemental information file in Holstege et al.[4] To homogenize EOAD/LOAD categorization across cohorts, we defined as EOAD all individuals with an age at onset (AAO) <66 years or age at last visit <66 years if AAO was missing. Individuals without numerical information for age were considered as LOAD.

Overall, our starting dataset contained 12,669 exomes from affected individuals (including 4,150 EOAD) and 9,650 exomes from control individuals (Figure 1).

All participants provided informed written consent, as previously described,[4] and according to the declaration of Helsinki. This study, based on the retrospective analysis of existing data, was approved by the CERDE ethics committee of the Rouen University Hospital (CERNI/CERDE notifications 2017-015 and 2019-055). All the identifiers used in this document are known only to the laboratory members who addressed them.

CNV calling

Bioinformatics tools detect CNVs from sequencing data through various approaches, including the relative distribution of read depth, relative positions of paired reads from each other, or multiple and partial alignments of reads.[29,30] Such approaches applied to genome-sequencing data show diverse recall and precision rates and require multiple combinations of tools at high computational costs. When working from sequencing data obtained following capture, as for exome sequencing, using the read-depth approach shows high levels of accuracy at the exome level for rare CNVs,[12,21] not requiring combinations of tools. Exome-specific tools do not rely on intra-sample read-depth variation along the sequence, but on region-per-region inter-sample comparison. We previously validated a CNV-calling workflow based on the CANOES tool,[23] which demonstrated good sensitivity (87.25%) and positive predictive value (86.4%–90.2%) for rare CNVs from exomes.[24] We adapted this workflow for a large dataset and applied it to BAM files (Figure S12 and supplemental methods). Samples with excessive calls were excluded, as were individuals with excessively large CNVs, which are unlikely germline, to avoid biases from clonal hematopoiesis associated with large blood-specific, age-related CNVs (Table S3 and Figure S1). We then assessed the presence of pathogenic Mendelian CNVs. CNV calling and QC are detailed in supplemental information.

CNV and transcript filtering

For case-control analyses, we excluded CNVs cross-overlapping by ≥70% with common CNVs (frequency >1% in Database of Genomic Variants gold standard[31] and non-gnomAD, non-neuro, non-Finnish Europeans, v.2.1[18]), CNVs with ≥50% overlap with segmental duplications, gonosomal CNVs, and those overlapping only one target (corresponding to monoexonic CNVs), to reduce false positives.

To address capture kit diversity, we worked at the transcript level (Figure S2). A transcript was considered as fully deleted (complete deletion) or duplicated (complete duplication) if all associated capture targets were affected in the respective capture kit, and it was considered partially deleted or partially duplicated if a subset of the capture targets were affected in the respective capture kit. We then built a copy-number matrix for each autosomal protein-coding transcript, with five possible states per sample and transcript post filtration: (1) missing information, (2) no CNV (copy number = 2), (3) complete or (4) partial duplication, or (5) complete or partial deletion. We worked under the assumption that both complete and partial deletions likely result in haploinsufficiency, i.e., that deletions removing part of the gene coding sequence or the whole coding sequence result in

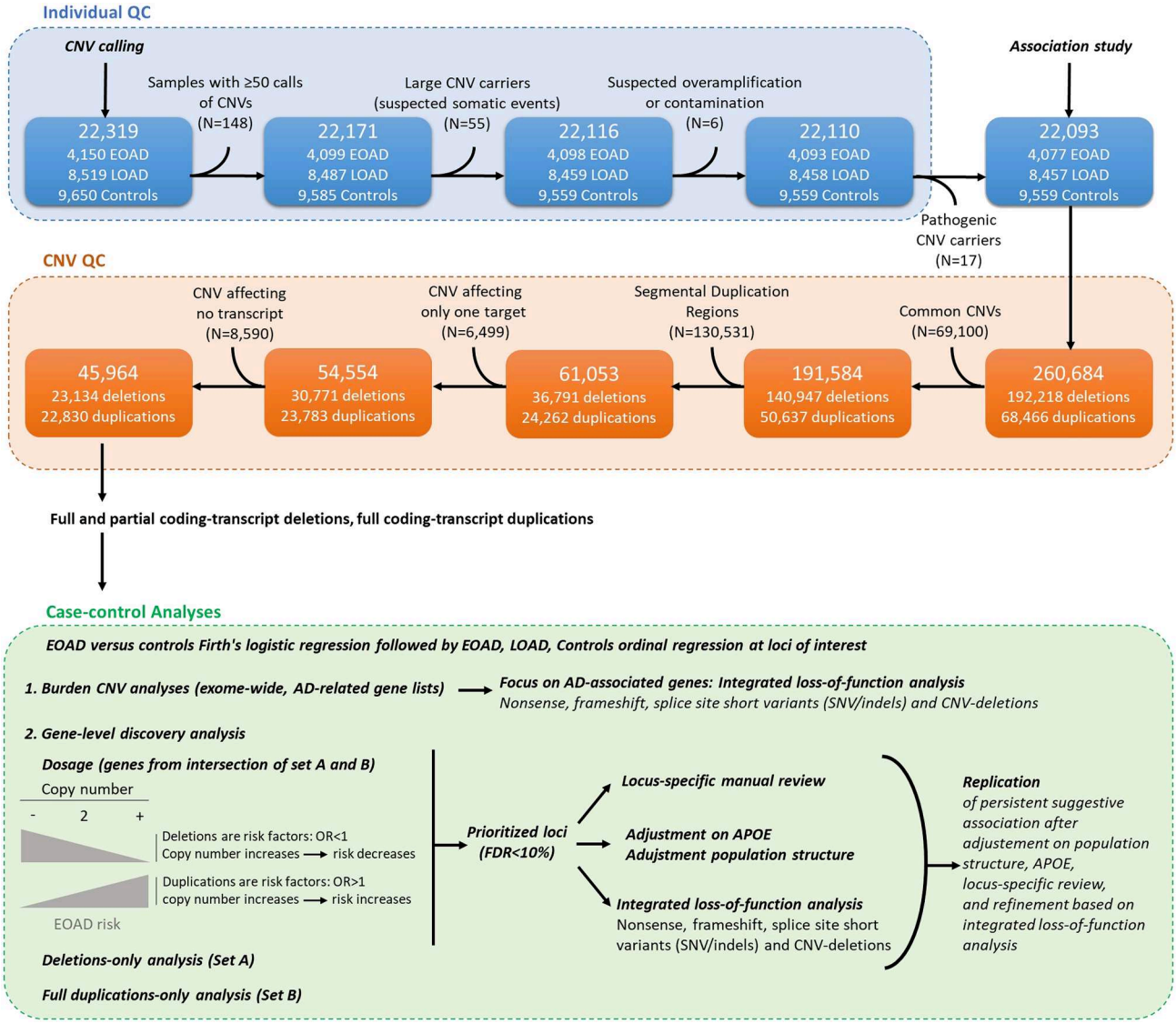


Figure 1. Summary of the study
This study included exome-sequencing samples from the ADES-ADSP described in Holstege et al.[4] restricted to individuals passing their quality control and being included in a CNV-calling-batch of 50+ individuals. QC, quality control; CNV, copy-number variation; EOAD, early-onset Alzheimer disease; LOAD, late-onset Alzheimer disease; OR, odds ratio. Set A, non-frequently deleted genes; set B, non-frequently duplicated genes.

haploinsufficiency. In dosage analyses, we converted states (2), (3), and (5) into transcript copy-number information.

To avoid the analysis of rare CNVs affecting transcripts frequently deleted or duplicated, we then built two sets of transcripts: (1) transcripts with <1% cumulative deletion frequency in public databases (set A, non-frequently deleted transcripts) and (2) transcripts with <1% complete duplication frequency (set B, non-frequently duplicated transcripts).

Statistical analyses

First, we performed burden CNV analyses at the exome level, i.e., summing the total number of genes encompassed by a deletion or a complete duplication (affecting coding transcripts, restricted to set A for the deletion analysis and set B for the duplication analysis). Next, we performed similar burden analysis, now focusing on two gene lists: (1) genes associated with AD in genome-wide association study (GWAS)[2] and (2) genes having a function in the Aβ network.[32] We then focused on AD-associated genes by performing gene-based analyses of LoF variants including both deletion CNVs (complete and partial) and truncating SNVs/indels in genes previously associated with AD. Burden CNV analyses consisted in (1) EOAD-control association analyses using Firth's logistic regression and (2) ordinal regression using the whole dataset (EOAD, LOAD, and control), assuming an EOAD < LOAD < control effect (or conversely), as previously performed for burdens of rare SNV/indels at the gene level.[4]

Second, we performed a gene-level exome-wide discovery analysis, consisting of a systematic gene-level association study of all protein-coding genes. This analysis was transcript based to allow the putative identification of burdens of rare CNVs affecting only a proportion of transcripts of a given gene. Based on the stronger

effect of rare short variants on EOAD risk and the expected extreme rarity of CNVs, limiting power, we compared EOAD-affected subjects with control subjects and assessed gene dosage (dosage = 1 for deletions, 2 for normal copy number, and 3 for duplications) using complete/partial deletions and complete duplications on the fusion of transcript sets A and B. Because of limited power due to rarity of variants, we considered as suggestive all signals with a false discovery rate (FDR) below the permissive threshold of 10%. Suggestive signals were further assessed by ordinal regression using the whole dataset (EOAD, LOAD, and control).

Complementary analyses consisted in (1) gene-level, transcript-based analyses on deletions only (all protein coding genes from set A) and complete duplications only (all protein coding genes set B); and (2) gene-based analysis of LoF variants (deletion CNVs plus truncating SNVs/indels) in genes prioritized by the dosage analysis (integrated LoF analysis).

As our primary analyses were not adjusted for covariates, we further performed the following sensitivity analyses: adjusted analyses based on ancestry (Firth's logistic model including the first ten principal components), based on *APOE4* allele dosage, or after exclusion of affected individuals with missing information regarding AAO ($n = 373$).

All statistical tests were two-sided.

CNV confirmation

CNVs of interest were all manually reviewed. Targeted validation of CNVs was performed when DNA was available, using either quantitative multiplex PCR of short fluorescent fragments (QMPSF) or droplet digital PCR (ddPCR).

Replication

Loci with FDR < 10% in the dosage discovery analysis, and remaining after adjustments and refinement by the integrated LoF analysis, were evaluated in a replication analysis using genotyping data from the European Alzheimer & Dementia Biobank (EADB), the UK Biobank, deCODE, GR@ACE, and exomes of African Americans from ADSP. Detailed sample selection, calling, and QC procedures are reported in supplemental information. Due to heterogeneous numbers of EOAD-affected individuals in the replication cohorts, we pooled them into a single replication dataset. To leverage the full replication dataset in a single replication test, we performed an ordinal regression. Discovery and replication datasets were then pooled in a mega-analysis. Nominal replication significance threshold was set at 0.05, study-wide significance was set at 2.91×10^{-5} corresponding to 1,716 gene-level tests fulfilling the aforementioned criteria out of 18,050 genes resulting from the fusion of sets A and B, with an exome-wide significance set at 2.77×10^{-6} (Bonferroni correction). Hits were considered as replicated if they reached nominal significance threshold in the replication and study-wide significance in the mega-analysis.

In vitro cellular assays

To investigate the involvement of SCARF2 in Aβ uptake, we overexpressed the *SCARF2* cDNA in human microglial clone 3 (HMC3) cells and compared this with a mock condition. Cells were treated with a fluorescent-labeled Aβ peptide (Aβ42-HiLyteTM Fluor 647 peptide, Anaspec), and Aβ internalization was quantified by flow cytometry. To evaluate the specific involvement of SCARF2 on Aβ uptake, mock- and SCARF2-transfected cells were similarly treated with fluorescent-labeled latex beads and analyzed by flow cytometry. SCARF2 protein staining was confirmed by western blotting.

Results

Detection of CNVs from exome-sequencing data of 22,319 individuals

Out of 22,319 exomes initially included, 22,110 passed QC (4,093 EOAD individuals, 8,458 LOAD individuals, and 9,559 unaffected control subjects) (Table 1; Figure 1, supplemental information; Tables S1–S3), totaling 261,190 CNVs. The first step consisted in searching for pathogenic CNVs in monogenic dementia genes. We identified 17 individual heterozygotes for such a pathogenic CNV, all but one with EOAD, including eight previously reported.[11,13,33–35] Overall, we detected ten *APP* duplications (mean AAO = 49.2 years, min–max 43–56), one in-frame *PSEN1* deletion (AAO = 56 years), and six *MAPT* (MIM:157140) duplications, leading to a primary tauopathy differential diagnosis (mean AAO = 58.2, min–max 45–87) (Table 2). Individuals with pathogenic variants were excluded from further case-control analyses.

Burden CNV analyses in the protein-coding genome and focus on AD-related genes

We assessed whether EOAD-affected individuals exhibit a higher burden of rare CNVs overall, as compared to control subjects, at two levels (Table S4). First, we performed an exome-wide burden CNV analysis across all protein-coding genes and showed that EOAD-affected individuals have a higher likelihood of carrying at least one deleted (set A, non-frequently deleted genes, see material and methods) or one duplicated (set B, non-frequently duplicated genes) gene than control individuals (deletions burden: OR = 1.15 [95% confidence interval 1.06; 1.23], $p = 3 \times 10^{-4}$; complete duplications burden: OR = 1.10 [1.02; 1.19], $p = 0.0125$; OR should be interpreted as the odds of carrying at least one deleted [or, respectively, duplicated] gene vs. carrying none). Second, restricting to AD-GWAS genes,[2] the difference was amplified for deletions (OR = 2.67 [1.51; 4.74], $p = 8 \times 10^{-4}$) and reversed for duplications (OR = 0.58 [0.30; 1.04], $p = 0.0696$) suggesting that some deletions in GWAS genes increase EOAD risk while some complete duplications could be protective (Table S4), and at-risk duplications in other genes may remain to be identified. Analysis of the Aβ-network list of genes[32] indicated enrichment in deletions among EOAD-affected individuals (OR = 1.44 [1.09; 1.90], $p = 0.0114$, Table S4).

Next, to further assess the deletions-related signals, we performed a gene-based integrated LoF analysis on a list of known AD-associated genes, gathering short truncating variants (nonsense, frameshift, and canonical splice site SNVs/indels) with CNV deletions (complete and partial) (see supplemental information). We identified a

Table 1. Description of individuals included in case-control analyses

	EOAD	LOAD	All AD	Controls
N	4,077	8,457	12,534	9,559
Age, mean (SD)[a]	58.43 years (5.09)	77.90 years (7.48)	71.38 years (11.42)	81.36 years (12.28)
Sex, *N* (%)				
Males	1,814 (44.5%)	3,307 (39.1%)	5,121 (40.9%)	4,147 (43.4%)
Females	2,263 (55.5%)	5,150 (60.9%)	7,413 (59.1%)	5,412 (56.6%)
APOE, *N* (%)[b]				
E4-non-carriers	1,704 (41.9%)	4,640 (54.9%)	6,344 (50.7%)	7,535 (79.6%)
E4-heterozygous	1,699 (41.7%)	3,417 (40.5%)	5,116 (40.9%)	1,830 (19.3%)
E4-homozygous	668 (16.4%)	390 (4.6%)	1,058 (8.4%)	101 (1.1%)
No. of deletions, mean ± SD (range)	1.18 ± 1.75 (0–35)	1.00 ± 1.38 (0–35)	1.06 ± 1.51 (0–35)	1.03 ± 1.41 (0–32)
Deletion size in bp, mean (range)	36,788 (187–2,860,014)	37,750 (129–7,103,184)	37,401 (129–7,103,184)	35,299 (146–8,273,974)
No. of duplications, mean ± SD (range)	1.09 ± 1.25 (0–23)	1.02 ± 1.37 (0–36)	1.04 ± 1.33 (0–36)	1.02 ± 1.37 (0–33)
Duplication size in bp, mean (range)	91,540 (129–4,847,347)	89,421 (75–4,809,978)	90,141 (75–4,847,347)	89,974 (75–4,840,909)

N, number; SD, standard deviation; bp, base pairs; EOAD, early-onset Alzheimer disease; LOAD, late-onset Alzheimer disease.
[a]Age at onset for affected individuals and age at last exam/inclusion for control subjects. Age is missing for 373 LOAD individuals and 256 control subjects.
[b]APOE genotype is missing for 6 EOAD individuals, 10 LOAD individuals, and 93 control subjects. CNV statistics are presented after filtration.

significant enrichment of *ABCA1* (OR_{EOAD} = 5.77 [2.25; 17.06], $p = 2 \times 10^{-4}$; OR should be interpreted as the odds of carrying at least one LoF variant in the gene vs. carrying none) and *ABCA7* LoF alleles with AD (OR_{EOAD} = 2.29 [1.44; 3.65], $p = 6 \times 10^{-4}$, Figure S8), now adding CNV deletions to the known *ABCA7* association,[36] while *CTSB* (MIM: 116810), mapping to a GWAS locus, emerged as a candidate (OR_{EOAD} = 5.03 [1.50; 20.71], $p = 8.9 \times 10^{-3}$) (Tables S5 and S6 and Figures S7–S11; for additional results see supplemental information).

Gene-level case-control discovery analysis

We next performed a gene-level rare CNV association study on the 22,093 samples remaining after exclusion of individuals with pathogenic CNVs. We compared the gene dosage between EOAD and control groups (Figure S13). In dosage analysis, ORs should be interpreted for an increase of 1 in gene copy number, i.e., OR > 1 (respectively, OR < 1) means that duplications (respectively, deletions) are associated with a higher disease risk (Figures 1 and 2; Table 3). Loci with FDR < 10% were then assessed in the whole cohort using an ordinal regression including EOAD, LOAD, and control groups (Table S7). Additional analyses included deletion-only (Table S8) or complete duplication-only (Table S9) analyses, APOE4-adjusted analysis (Table S10), and population structure-adjusted analysis (Table S11). Removing samples with missing AAO information (set as LOAD in the primary analysis) did not affect the results (not shown). In deletion-only (respectively, complete duplication-only) analyses, ORs should be interpreted for a decrease (respectively, increase) of 1 in gene copy number, i.e., OR > 1 means that deletions (respectively, duplications) are associated with a higher disease risk, whereas OR < 1 means that deletions (respectively, duplications) are associated with a lower disease risk. All analyses are transcript based, and results are represented per gene when several transcripts show identical results in a given gene.

From 4,989 transcripts of 1,716 genes fulfilling our criteria for dosage association analysis, dosage analysis comparing EOAD-affected individuals to unaffected control subjects prioritized 44 transcripts (FDR ≤ 10%, Table 3) from 18 genes, at five loci, as 13 genes mapped to chromosome 22q11.21 (0.0271 < FDR < 0.0386), two other contiguous genes mapped to 19q13.32 (*ERCC2* [MIM 126340]-*KLC3* [MIM: 601334] locus, 0.0271 < FDR < 0.0412), and three genes were excluded (Table 3 and supplemental information).

Three single-gene loci showed an FDR ≤10%, namely *MBL2* (MIM: 154545; FDR = 0.0271, Figure S4), *ADI1* (MIM: 613400; FDR = 0.0916, Figure S5), and *FADS6* (MIM: 620577; FDR = 0.0271, Figure S6). The *MBL2* signal did not remain after adjusting for population structure, and the same adjustment decreased the significance for *FADS6* while not affecting the other loci (Figures S14 and S15; Table S11). The *ERCC2-KLC3* signal was driven by complete duplications (Table S9), and we showed that these rare duplications occurred on an *APOE4*-linked haplotype (Fisher's exact test *p* value = 4.26×10^{-13}) on

Table 2. Mendelian pathogenic CNVs

Sample	Cohort	AAO	Sex	CNV coordinate	HGVS nomenclature	Gene	Event
ALZ-0596-001[b]	ADES-FR	50	male	chr17:43679169–44145099	g.43585921_44159772dup (GenBank: NC_000017.10)	*MAPT*	complete duplication
ALZ-0441-005[b]	ADES-FR	45	male	chr17:43717946–44364336	g.43679455_44363695dup (GenBank: NC_000017.10)	*MAPT*	complete duplication
EXT-1114-001[b]	ADES-FR	54	female	chr17:43679164–44364341	g.43585926_44408484dup (GenBank: NC_000017.10)	*MAPT*	complete duplication
ROU-1373-001[b]	ADES-FR	54	male	chr17:43717946–44159945	g.43679455_44172015dup (GenBank: NC_000017.10)	*MAPT*	complete duplication
SC45A00007	PERADES	59	female	chr17:43679204–44159910	g.43611648_44171922dup (GenBank: NC_000017.10)	*MAPT*	complete duplication[a]
ID1	ADSP	87	female	chr17:43898720–44101538	g.43893949_44108840dup (GenBank: NC_000017.10)	*MAPT*	complete duplication[a]
EXT-0814-001[d]	ADES-FR	54	female	chr21:22370838–28338704	g.20058403_30248689dup (GenBank: NC_000021.8)	*APP*	complete duplication
EXT-0773-001[d,e]	ADES-FR	44	male	chr21:26946242–31066399	g.22910384_31233900dup (GenBank: NC_000021.8)	*APP*	complete duplication
EXT-0857-001[d]	ADES-FR	56	male	chr21:26946246–28338704	g.22910379_30248689dup (GenBank: NC_000021.8)	*APP*	complete duplication
FR16640974	Netherlands Brain Bank	47	male	chr21:22370838–28338827	g.20230358_30248689dup (GenBank: NC_000021.8)	*APP*	complete duplication[a]
TE17A00020a_DS1	PERADES	45	male	chr21:27056123–28338719	g.27012218_30248703dup (GenBank: NC_000021.8)	*APP*	complete duplication[a]
20732	PERADES	53	female	chr21:27062175–27484514	g.27056277_27512460dup (GenBank: NC_000021.8)	*APP*	complete duplication[a]
CH32A00027	PERADES	54	female	chr21:27066065–28338719	g.27062287_30248703dup (GenBank: NC_000021.8)	*APP*	complete duplication[a]
09D04764	ADC-Amsterdam	50	male	chr21:26541838–29645932	g.25725527_30248655dup (GenBank: NC_000021.8)	*APP*	complete duplication[a]
10D05060	ADC-Amsterdam	43	female	chr21:27253167–27542996	g.27142399_27840733dup (GenBank: NC_000021.8)	*APP*	complete duplication[a]
SID14617	AgeCoDe-UKBonn	46	female	chr21:26946221–27542992	g.25725527_27840779dup (GenBank: NC_000021.8)	*APP*	complete duplication[a]
EXT-0313-001[c]	ADES-FR	56	male	chr14:73672981–73678665	g.73664857_73683810del (GenBank: NC_000014.8)	*PSEN1*	deletion exon 9–10

AAO, age at onset (in years); CNV, copy-number variant; HGVS, Human Genome Variation Society.
[a]Previously unreported.
[b]Reported in Le Guennec et al.[33]
[c]Reported in Le Guennec et al.[13]
[d]Reported in Lanoiselée et al.[11]
[e]Reported in Rovelet-Lecrux et al.[34]

19q13.32 (Tables S10 and S12; Figure S3). The *ERCC2-KLC3* association signal did not remain significant after adjusting for *APOE*-ε4 dosage (OR = 1.58 [0.85; 2.99], $p = 0.148$), while other loci remained unaffected (Table S10).

The 22q11.21 region is known for rare recurrent deletions and duplications driven by non-allelic homologous recombination (NAHR) through low copy repeats (LCR) labeled LCR 22A to 22E (Figure 2). The region where all transcripts with FDR $\leq$ 10% clustered is located between LCR 22B and 22D (labeled as the "central region"), which do not encompass the critical region for DiGeorge (velo-cardio-facial, MIM: 188400) syndrome (LCR 22A–22B). We observed 22q11.21 central region deletions in four EOAD-affected individuals (0.1%), and this region was also prioritized in the deletion-only analysis (Table S8; 0 control, 0 to 1 [0.01%] LOAD depending on the CNV coordinates). One of the EOAD individuals carried a larger LCR 22A–22D deletion, typical of DiGeorge syndrome (Figure 2). He had no specific neurodevelopmental history, but he was born with congenital heart disease and presented the first signs of cognitive decline at age 43, without any signs of Parkinsonism. The AD diagnosis was confirmed with CSF biomarkers (Table 4). After targeted analysis in both unaffected parents, we showed that the deletion occurred *de novo*. Interestingly, duplications mirroring the LCR 22A–22D or the 22B–22D deletions, or affecting the 22C–22D or 22C–22E regions, were observed in 8–10 control individuals (0.08%–0.1%, depending on the CNV coordinates), in 4–5 LOAD individuals

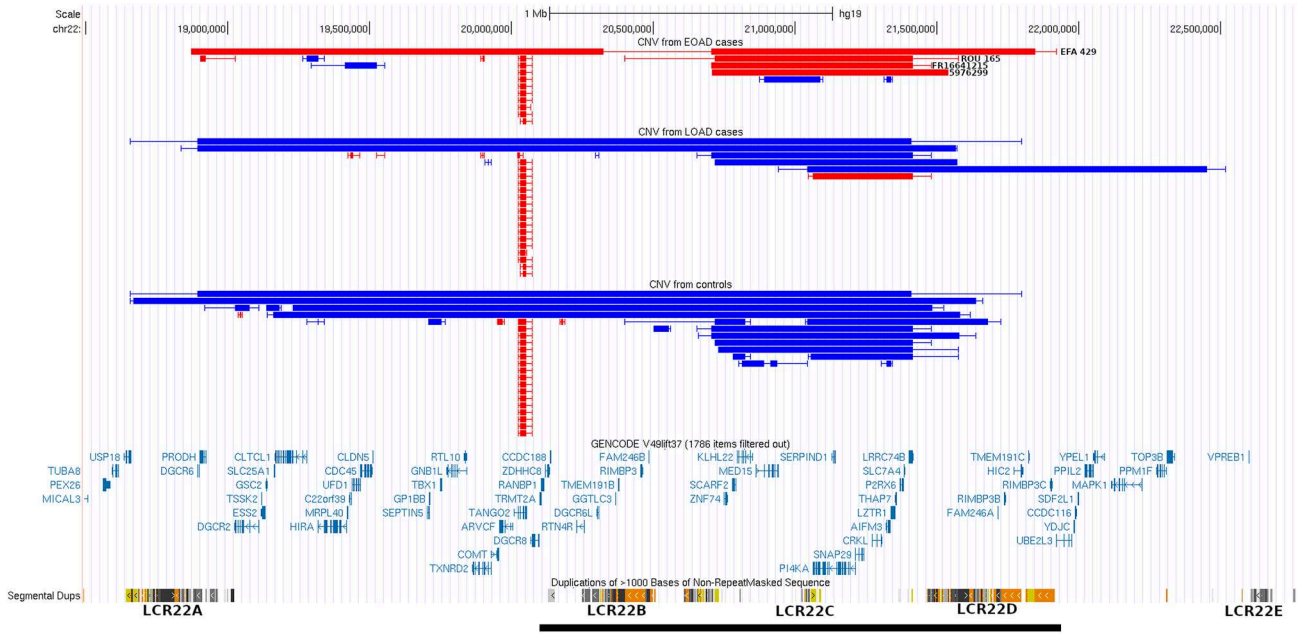


Figure 2. Focus on the 22q11.21 locus

UCSC genome browser view of CNVs identified in the main analysis. Red, deletions; blue, duplications. Large areas represent coordinates of CNVs as detected by CANOES. Thinner areas show breakpoint uncertainties, i.e., regions between two targets of the sample-specific capture kit. Genes are indicated below, and the regions of segmental duplications appear on the bottom part. Low copy repeats (LCRs) underlying recurrent rearrangements by non-allelic homologous recombination (NAHR) are indicated below. A black line underlines the locus prioritized here. Note that not all duplications are taken into account in the statistical analysis; only those encompassing a whole transcript are included.

(0.05%–0.06%, Figure 2), and in 0–1 EOAD individual (0%–0.02%) (Table S13). Overall, this suggests a dosage effect with deletions increasing AD risk and duplications decreasing the risk or delaying ages at onset. Performing the dosage analysis at the level of the whole cohort including LOAD using an ordinal regression, the association signal remained nominally significant (OR = 0.16 [0.04; 0.45], $p = 3.5 \times 10^{-4}$) (Table S7) as LOAD-affected individuals showed frequencies of deletions and duplications in between those in EOAD and control individuals.

Overall, DNA from 27 individuals (12 *ERCC2-KLC3*, two 22q11.21, one *MBL2*, three *ADI1*, and nine *FADS6*) was available, and CNVs were confirmed in 26 of 27 (one *FADS6* false positive [Table S14]—after removing the latter, EOAD vs. control dosage analysis remained similar: OR = 0.20 [0.07; 0.49], $p = 4 \times 10^{-4}$), in addition to two 22q11.21 deletions confirmed in independent chip data. This further supports the accuracy of the calling method, which is consistent with our previous results of validation based on a large number of CNVs and comparing exome-DNA chip pairs.[24]

Integrated LoF analysis at the prioritized loci

We then performed a case-control integrated LoF analysis gathering CNV deletions (complete and partial) with short truncating variants (nonsense, frameshift, and canonical splice site SNVs/indels) among the genes belonging to the prioritized loci in the dosage analysis (Tables S6 and S7). Although this integrated LoF analysis did not substantially modify the interpretation of the dosage analysis results, it narrowed the signal to the *SCARF2* (MIM: 613619)-*MED15* (MIM: 607372)-*KLHL22* (MIM: 618020) locus in the 22q11.21 central region. Indeed, we observed one LoF short variant in *SCARF2* in an EOAD individual and two in *MED15* in two LOAD individuals, whereas no control individual carried such LoF variants, as for CNV deletions, and all the other genes at this locus now appeared as less significant.

Replication of suggestive loci in large external cohorts

We aimed at replicating the *SCARF2-MED15-KLHL22* locus dosage, *ADI1* duplications, and *FADS6* deletion cases enrichment, using chip data from 32,535 AD-affected individuals (3,411 EOAD) and 359,929 unaffected control subjects, and exomes from African American AD individuals ($n = 1{,}442$ including 182 EOAD individuals) and unaffected control subjects ($n = 2{,}393$). As the *FADS6* signal was driven by a small recurrent deletion in the discovery (<10 kb, Figure S6), which is not detectable in DNA chip data, we were not able to perform an unbiased mega-analysis (Figure S18). Regarding *ADI1*, replication results were in the same direction as the discovery analysis (Figures 3A and 3B; Tables S15 and S16; Figure S17), but significance was not achieved. However, we succeeded in replicating the *SCARF2-KLHL22-MED15* gene-dosage association at the 22q11.2 locus, as ordinal regression confirmed an EOAD < LOAD < control association (Tables S15 and

Table 3. Results from EOAD vs. control dosage analysis and subsequent analyses on deletions and duplications

	N total			No. of individuals with CNVs (del/dup)			EOAD vs. control—dosage analysis			EOAD vs. control—deletion analysis		EOAD vs. control—duplication analysis			
Gene	**EOAD**	**LOAD**	**Control**	**EOAD**	**LOAD**	**Control**	**OR [95% CI]**[a]	***p* value**[b]	**FDR**	**OR [95% CI]**	***p* value**	**OR [95% CI]**	***p* value**	**Transcripts**[e]	**Chromosomal band**
PRAMEF26[c]	2,911	7,302	6,461	8 (8/0)	2 (2/0)	2 (1/1)	0.08 [0.01; 0.34]	2.88E−04	3.29E−02	12.61 [2.84;118.46]	3.96E−04	0.74 [0.005;13.87]	8.50E−01	GenBank: NM_001306072.3	1p36.21
ADI1	4,077	8,457	9,559	5 (0/5)	1 (0/1)	1 (1/0)	26.58 [3.06; 3482.37]	1.05E−03	9.16E−02	0.78 [0.005; 14.65]	8.78E−01	25.82 [2.93; 3389.88]	1.40E−03	GenBank: NM_001306077.2, NM_018269.4	2p25.3
MTUS1[d]	4,077	8,457	9,559	54 (54/0)	38 (37/1)	57 (57/0)	0.47 [0.33;0.66]	1.95E−05	2.71E−02	related transcripts are not in set A		no duplication in EOAD nor in Controls		GenBank: NM_001001924.3, NM_001363057.2, NM_001363058.2, NM_001363059.2, NM_001363061.2	8p22
MBL2	4,077	8,457	9,559	2 (2/0)	25 (25/0)	45 (44/1)	5.76 [2.22; 21.16]	5.44E−05	2.71E−02	0.13 [0.02;0.39]	2.02E−05	0.78 [0.005;14.65]	8.78E−01	GenBank: NM_000242.3, NM_001378373.1, NM_001378374.1	10q21.1
FADS6	4,077	8,457	9,559	16 (15/1)	11 (11/0)	5 (5/0)	0.19 [0.07; 0.45]	1.65E−04	2.71E−02	6.63 [2.65;19.34]	3.34E−05	7.04 [0.38; 1026.65]	1.89E−01	GenBank: NM_178128.6	17q25.1
PLIN4[d]	4,077	8,457	9,559	20 (19/1)	10 (10/0)	12 (12/0)	0.31 [0.15; 0.61]	8.38E−04	7.66E−02	related transcripts are not in set A		3.67 [1.82; 7.67]	2.83E−04	GenBank: NM_001393888.1, NM_001393889.1, NM_001393890.1, NM_001393891.1, NM_001367868.2	19p13.3
KLC3	4,077	8,457	9,559	24 (1/23)	33 (1/32)	16 (0/16)	3.06 [1.65; 5.79]	4.28E−04	4.12E−02	7.04 [0.38;1026.65]	1.89E−01	3.35 [1.8;6.41]	1.57E−04	GenBank: NM_177417.3	19q13.32
ERCC2	4,077	8,457	9,559	29 (1/28)	43 (2/41)	20 (0/20)	3.04 [1.74; 5.4]	1.09E−04	2.71E−02	7.04 [0.38;1026.65]	1.89E−01	3.28 [1.86;5.86]	4.09E−05	GenBank: NM_001130867.2	
	4,077	8,457	9,559	29 (1/28)	42 (2/40)	20 (0/20)								GenBank: NM_000400.4	
ZNF74	4,077	8,457	9,559	4 (4/0)	4 (0/4)	9 (0/9)	0.04 [0.0003; 0.29]	2.71E−04	3.29E−02	21.12 [2.25;2799.03]	4.75E−03	0.12 [0.001;0.97]	4.58E−02	GenBank: NM_001256523.2, NM_001256525.2, NM_003426.4, NM_001256524.2	22q11.21
SCARF2	4,077	8,457	9,559	4 (4/0)	4 (0/4)	9 (0/9)	0.04 [0.0003; 0.29]	2.71E−04	3.29E−02	21.12 [2.25;2799.03]	4.75E−03	0.12 [0.001;0.97]	4.58E−02	GenBank: NM_153334.7, NM_182895.5	
KLHL22	4,077	8,457	9,559	4 (4/0)	4 (0/4)	8 (0/8)	0.04 [0.0003; 0.30]	3.80E−04	3.86E−02	21.12 [2.25;2799.03]	4.75E−03	0.14 [0.001;1.10]	6.57E−02	GenBank: NM_032775.4	
MED15	4,077	8,457	9,559	4 (4/0)	4 (0/4)	9 (0/9)	0.04 [0.0003; 0.29]	2.71E−04	3.29E−02	21.12 [2.25;2799.03]	4.75E−03	0.12 [0.001;0.97]	4.58E−02	GenBank: NM_001003891.3, NM_001293234.2, NM_001293235.2, NM_001293236.2, NM_001293237.2, NM_015889.5	

(*Continued on next page*)

Table 3. ***Continued***

	***N* total**			**No. of individuals with CNVs (del/dup)**			**EOAD vs. control—dosage analysis**			**EOAD vs. control—deletion analysis**		**EOAD vs. control—duplication analysis**			
Gene	**EOAD**	**LOAD**	**Control**	**EOAD**	**LOAD**	**Control**	**OR [95% CI]**[a]	***p* value**[b]	**FDR**	**OR [95% CI]**	***p* value**	**OR [95% CI]**	***p* value**	**Transcripts**[e]	**Chromosomal band**
PI4KA	4,077	8,457	9,559	4 (4/0)	6 (1/5)	10 (0/10)	0.03 [0.0003; 0.28]	1.92E−04	2.71E−02	21.12 [2.25;2799.03]	4.75E−03	0.11 [0.001;0.87]	3.19E−02	GenBank: NM_001362862.2, NM_001362863.2, NM_058004.4	
SERPIND1	4,077	8,457	9,559	4 (4/0)	6 (1/5)	10 (0/10)	0.03 [0.0003; 0.28]	1.92E−04	2.71E−02	21.12 [2.25;2799.03]	4.75E−03	0.11 [0.001;0.87]	3.19E−02	GenBank: NM_000185.4	
SNAP29	4,077	8,457	9,559	4 (4/0)	6 (1/5)	10 (0/10)	0.03 [0.0003; 0.28]	1.92E−04	2.71E−02	21.12 [2.25;2799.03]	4.75E−03	0.11 [0.001;0.87]	3.19E−02	GenBank: NM_004782.4	
CRKL	4,077	8,457	9,559	4 (4/0)	6 (1/5)	10 (0/10)	0.03 [0.0003; 0.28]	1.92E−04	2.71E−02	21.12 [2.25;2799.03]	4.75E−03	0.11 [0.001;0.87]	3.19E−02	GenBank: NM_005207.4	
LZTR1	4,077	8,457	9,559	4 (4/0)	6 (1/5)	10 (0/10)	0.03 [0.0003; 0.28]	1.92E−04	2.71E−02	21.12 [2.25;2799.03]	4.75E−03	0.11 [0.001;0.87]	3.19E−02	GenBank: NM_006767.4	
THAP7	4,077	8,457	9,559	4 (4/0)	6 (1/5)	10 (0/10)	0.03 [0.0003; 0.28]	1.92E−04	2.71E−02	21.12 [2.25;2799.03]	4.75E−03	0.11 [0.001;0.87]	3.19E−02	GenBank: NM_030573.3, NM_001008695.1	
P2RX6	4,077	8,457	9,559	4 (4/0)	6 (1/5)	10 (0/10)	0.03 [0.0003; 0.28]	1.92E−04	2.71E−02	21.12 [2.25;2799.03]	4.75E−03	0.11 [0.001;0.87]	3.19E−02	GenBank: NM_001159554.2, NM_001349874.2, NM_001349875.2, NM_001349876.2, NM_001394691.1, NM_001394692.1, NM_001394693.1, NM_001394694.1, NM_001394695.1, NM_001394696.1, NM_001394697.1, NM_005446.5	
SLC7A4	4,077	8,457	9,559	4 (4/0)	6 (1/5)	10 (0/10)	0.03 [0.0003; 0.28]	1.92E−04	2.71E−02	21.12 [2.25;2799.03]	4.75E−03	0.11 [0.001;0.87]	3.19E−02	GenBank: NM_004173.3	
LRRC74B	3,917	7,799	8,016	4 (4/0)	6 (1/5)	10 (0/10)	0.04 [0.0003; 0.29]	1.86E−04	2.71E−02	18.44 [1.97;2443.22]	7.39E−03	0.10 [0.001;0.75]	2.01E−02	GenBank: NM_001291006.2	

N, number. EOAD, early-onset Alzheimer disease; LOAD, late-onset Alzheimer disease.
[a]OR should be interpreted for an increase of transcript copies by 1. i.e., OR > 1 (respectively, OR < 1) means that duplications (respectively, deletions) are associated with a higher disease risk.
[b]Non-adjusted *p* values, selected based on FDR < 10%.
[c]Gene overlapping with repeats or duplicated gene in the genome (despite CNVs not overlapping >50% with repeats).
[d]Gene for which the signal in dosage analysis is driven by deletion, whereas transcripts are not in set A.
[e]The table gathers transcript results on a single line per gene when several transcripts exhibit the same *p* values. Genes at the same locus are shown on the first column of the table with corresponding chromosome band.

Table 4. Clinical information of individuals with a 22q11.2 deletion in the main analysis

Individual ID	EFA 429	ROU 165	FR16641215	5976299
Age at sample (years)	46	53	64	53
Sex	male	male	female	male
APOE genotype	3–4	3–4	3–3	3–4
Age at onset (years)	43	51	54	49
Educational level (years)	12	9	N/A	5
First reported symptoms	memory and executive functions	memory	N/A	memory, writing
Medical history of the 22q11 deletion spectrum	congenital heart defect, no history of neurodevelopmental disorders	no history of neurodevelopmental disorders	N/A	no history of neurodevelopmental disorders
Family history of neurodegenerative disease	Parkinson disease in mother at 62 related to a *LRRK2* pathogenic variant	negative	dementia in father with onset around 60 and death at 69 years	negative
Disease duration (years)	6 (to death)	8	10 (to death)	9
Additional arguments supporting AD pathophysiology	• MMSE = 12/30 at 46 years. • severe cortical diffuse atrophy on brain imaging • CSF Aβ42: 471 pg/mL (N > 650) • CSF tau: 732 pg/mL (N < 350) • CSF P-tau: 100 pg/mL (N < 60)	• MMSE = 15/30 at 53 years • severe episodic memory disorder with misorientation and logopenic aphasia • bilateral mediotemporal and parietal cortical atrophy • hypoperfusion in bilateral temporal-parietal cortical regions	• neuropathological analysis: numerous amyloid plaques and neurofibrillary tangles, as well as severe cellular neuronal loss with reactive gliosis matching AD neuropathological diagnosis	• MMSE = 20/30 at 53 years • mediocre recall and recognition, apraxia, visuospatial and visualgnostic disabilities • major parietal atrophy and sparse white matter lesions • CSF Aβ42: 450 pg/mL (N > 650) • CSF tau: 359 pg/mL (N < 350) • CSF P-tau: 181 pg/mL (N < 60) • Diffuse increased tracer uptake at the frontal, parietal, and cingular cortex on PiB PET

AD, Alzheimer disease; EOAD, early-onset AD; LOAD, late-onset AD; MMSE, Mini Mental State Examination; CSF, cerebrospinal fluid; P-tau, phosphorylated tau; PiB PET, Pittsburgh compound B positron emission tomography.

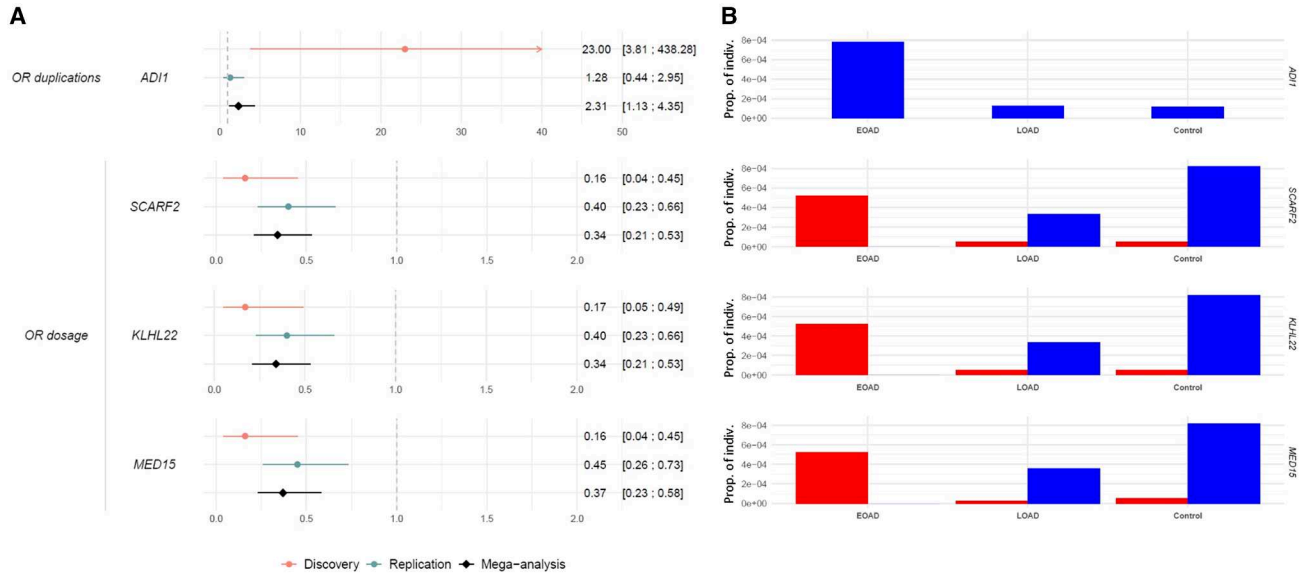


Figure 3. Mega-analysis of results from both main analysis and replication

(A) Forest plot associated with the mega-analysis of results from both discovery and replication. Numbers on right display odds ratio (OR) with 95% confidence interval. For dosage analyses (*SCARF2*, *KLHL22*, and *MED15*), OR should be interpreted for an increase of transcript copies by 1. i.e., OR > 1 means that duplications (and OR < 1 means that deletions) are associated with a higher disease risk (see Figure 1).

(B) Proportion of deletions and duplications by status and gene. Red, deletions; blue, complete duplications. Data include both discovery and replication datasets. Prop. of indiv., proportion of individuals.

S16; Figures 3A and S16; replication $OR_{SCARF2} = 0.40$ [0.23; 0.66], $p = 1.50 \times 10^{-4}$), with a mega-analysis $OR_{SCARF2} = 0.34$ [0.21; 0.53] ($p = 5.52 \times 10^{-7}$), achieving both study-wide and exome-wide significance thresholds (OR to be interpreted per additional gene copy).

Assessing SCARF2 as a potential biological driver of the 22q11.2 association

Finally, we sought to provide an initial step toward understanding a potential biological basis for the *SCARF2-KLHL22-MED15* gene-dosage association. We prioritized *SCARF2*, as it belongs to the class F scavenger receptor family and is expressed in multiple cell types in the brain and in macrophages. Although a scavenger receptor role is yet to be demonstrated for SCARF2, such activity has been confirmed for other family members, including *SCARF1* (MIM: 607873) and *MEGF10* (MIM: 612453), the latter even being a known receptor for Aβ.[37,38] We therefore investigated whether SCARF2 may influence Aβ cellular uptake by overexpressing *SCARF2* in HMC3 cells (Figures 4 and S19). SCARF2-transfected cells exhibited a significantly higher uptake of fluorescent Aβ42 peptides compared to mock control cells. Conversely, the uptake of fluorescent latex beads was not increased, underscoring SCARF2 specificity in Aβ phagocytosis.

Discussion

Here, we identified an association with gene dosage at the central region of 22q11.21. Overall, the 22q11.21 locus showed enrichment of deletions in EOAD-affected individuals and mirroring duplications in control subjects, while LOAD-affected individuals remained in between. This region has not been known to be associated with AD risk so far, despite a suggestive association in a family-based analysis.[39] In our gene-dosage analysis, the association was exome-wide significant ($p < 2.77 \times 10^{-6}$), and we are providing additional levels of evidence, with one of the AD-affected individuals with unaffected parents showing a *de novo* deletion, which is a strong genetic argument at the individual level, and one piece of potential mechanism, as expression of *SCARF2* cDNA led to increased Aβ uptake.

Overall, central 22q11.21 deletions remain extremely rare. Here, no control carried such a deletion in the discovery, while duplications were enriched in unaffected control individuals. Notably, very few control subjects carrying central 22q11.21 deletions or larger ones (B-C or B-D) were identified in the replication (19/362,321, 0.005%), with a frequency that remains compatible with the absence of deletion in the discovery (0/9,559, i.e., frequency <0.01%). Among those having available information on parents in the UK Biobank, 28.57% (2/7) had a parent with AD or related dementia, compared to 15.65% (35,569/227,256) in individuals without a CNV in this region. However, rarity of deletion prevented us from performing a powerful comparison (Fisher exact test p value = 0.3017). It should be noted that 10.36% of participants have an unknown status of family history in this database, and as such deletions/duplications are being driven by NAHR, they can be found as *de novo* events, hence limiting the interpretation of family history in this

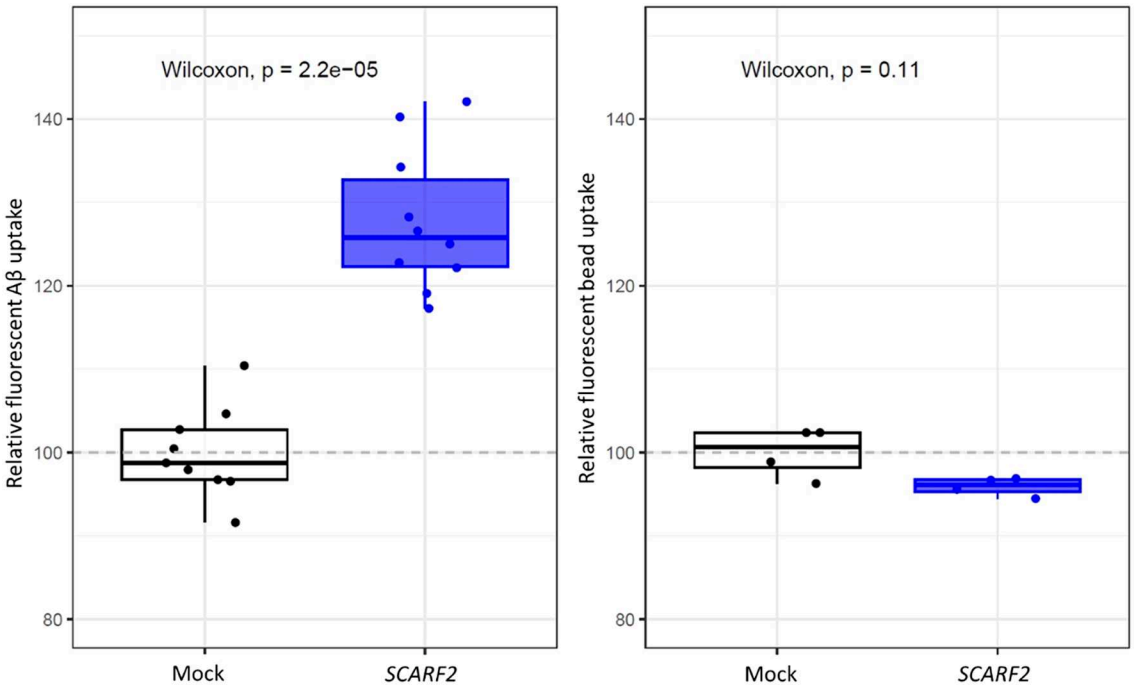


Figure 4. $A\beta_{42}$ uptake in HMC3 cells
Mock-transfected and *SCARF2*-cDNA-transfected HMC3 cells were incubated with fluorescent Aβ42 peptides (left) or with fluorescent latex beads (right). Aβ42 or bead uptake was assessed by flow cytometry. *SCARF2* transfection was associated with a significantly increased uptake of Aβ42 peptides compared to the mock condition (Wilcoxon, $p = 2.2 \times 10^{-5}$), while the uptake of latex beads remained unchanged in both conditions ($p = 0.11$).

database for this locus. Beyond the fact that the penetrance of AD may not be full by the age of 65 years in deletion carriers, most of those from the UK Biobank are still rather young, so some might still develop AD.

Interestingly, while our discovery was enriched in EOAD individuals and highlighted at-risk deletions with mirroring duplications in control subjects, our replication was enriched in control individuals and highlighted the protective role of 22q11.21 central duplications with more power than the discovery alone. With an OR of 0.34 [0.18;0.56] in the mega-analysis, duplications appear as a strong AD-risk-decreasing factor. So far, only a handful of genetic factors have been shown to significantly reduce AD risk, after exclusion of GWAS hits with ORs close to 1, namely the *APP* Icelandic variant with a very strong effect (OR $\sim$ 0.2)[40] and three other AD-risk-decreasing factors with ORs of $\sim$0.7: *APOE2*,[41] a rare missense variant in *PLCG2* (MIM: 600220),[42] and the very recent identification of a protective 19-bp deletion of an *APOE* enhancer in *APOE4*-positive African Americans, though still associated with an increased AD risk compared to *APOE3* homozygosity.[43] Duplications of the 22q11.21 central region thus represent one of the strongest AD-risk-decreasing factors in terms of OR estimates, after the Icelandic *APP* variant,[40] which is both ultra-rare and population specific, contrary to 22q11.21 duplications, which are found in multiple ethnicities.

Of note, 22q11.21 central duplications reduce AD risk in the context of a complex determinism, conferring only relative "protection." As one can identify AD risk factors in unaffected control subjects, we also identified a few duplications in LOAD duplication heterozygotes. Of the four LOAD duplication heterozygotes in the discovery, two (50%) carried at least one AD risk factor with OR > 1.5, both being *APOE4* heterozygous and one of them also carrying a splice region variant in *ABCA7*, c.4764+5G>A (GenBank: NC_000019.10, NM_019112.4) (p.?), predicted to result in exon skipping and thus a probable AD risk factor with moderate effect. As a comparison, 3/10 (30%) unaffected control individuals carrying a 22q11.21 duplication carried a similar AD risk factor (two were *APOE4* heterozygous and one carried the *ATP8B4* c.1183G>A [GenBank: NC_000015.10, NM_024837.4] [p.Gly395Ser] variant classified as a modest risk factor with OR around 1.5),[44] further demonstrating the complex nature of the genetic determinism of AD and the need for integrated models gathering multiple risk and protective factors for an accurate AD risk prediction.[45]

In theory, two hypotheses might explain the association with gene dosage at the 22q11.21 central region. First, the association could be non-specific to AD mechanisms but rather a modifier of AAO in individuals prone to develop AD. Indeed, individuals with a central 22q11.21 deletion have a higher risk of neurodevelopmental disorders,[46] as do those individuals with a DiGeorge-syndrome associated deletion (22A–22B or larger) with higher penetrance, and lower average cognitive performances.[47] Whether or not it is clinically significant

prior to AD onset, putatively reduced cognitive reserve might contribute to earlier AD-onset ages.[48,49] However, this hypothesis would require that duplication heterozygotes have better average cognitive performances, which is not known, 22B–22D duplications being considered as variants of uncertain significance. Second, dosage of certain genes in this interval could be directly influencing the pathophysiology of AD. Given its presumed functions and expression, we hypothesized that *SCARF2* might act as a scavenger receptor for Aβ. We showed that overexpression of *SCARF2* led to increased Aβ uptake. Although these results will require further confirmation in other models and the underlying mechanisms remain to be determined, including a putative function as an actual scavenger receptor, they are consistent with the observed reduced risk in the case of a duplication. In addition, we cannot rule out the hypothesis that other genes in the same region may affect AD-related mechanisms.

From a medical point of view, our results may be of interest for individuals with DiGeorge syndrome, among those with large deletions including *SCARF2-KLHL22-MED15* (22A–22C deletions or larger). DiGeorge syndrome presents with variable penetrance symptoms, including neurodevelopmental disorders ranging from mild intellectual disability, psychosis, and autism spectrum disorder to normal intelligence, congenital heart defects, palatal malformations, velopharyngeal incompetence, phosphocalcic metabolism disorders, immune deficiency, some morphological features, and increased risk of Parkinson disease.[50] However, not much is known about the aging of these individuals, as most reports in adults involve individuals in their mid-30s on average and limited survival until late adulthood.[51] At least one individual with DiGeorge syndrome (aged 56) showed AD-related neuropathology,[50] but neuropathological reports are scarce. Our results suggest that, as care of DiGeorge syndrome is improving, enabling individuals to age, cognitive follow-up should be proposed, especially to those with a large deletion. This finding of a 22q11.21 association with AD is somehow reminiscent of the trisomy-21 and *APP* duplication.[9,52,53] However, the impact of the latter may be much stronger, with full penetrance by the age of 65–70 years,[12,54] probably because of the straightforward link between *APP* overexpression and Aβ production, while Aβ clearance involves multiple players, with *SCARF2* now becoming a strong candidate Aβ scavenger.

Beyond the 22q11.21 locus, our study also highlighted CNVs in known AD-associated genes directly contributing to the disease in heterozygotes. Rare CNVs have long been implicated in diverse Mendelian disorders, including neurodegenerative diseases, in which they nevertheless remain an infrequent mechanism. For instance, only two *APP* triplications have been reported so far,[52,55] while *APP* duplications account for ∼9% of autosomal-dominant EOAD families, and *MAPT* duplications represent a very rare cause of AD-mimicking primary tauopathy,[35] as found in some individuals here. In line with Mendelian dementia genes, we show that rare CNVs also contribute to non-monogenic AD risk, as for *ABCA1* and *ABCA7*, where deletions represented a small proportion (<10%) of LoF alleles. Interestingly, we also identified a suggestive association of *CTSB* LoF with AD risk. *CTSB*, a GWAS-defined gene, encodes cathepsin B, an anti-amyloidogenic enzyme with Aβ peptide cleavage activity,[56] which is consistent with LoF alleles in affected individuals and duplications in control individuals (see also supplemental information).

Here, we used a large existing dataset of exome-sequencing data. We propose a QC and analysis strategy and obtained a dataset harmonized at the transcript level to reduce both the recurrence issue at the CNV level and the heterogeneity among datasets. This allowed (1) the identification of high-quality CNV calls in known AD-associated genes (Mendelian and risk factor genes) and (2) identification of suggestive loci, despite the extreme rarity of the individual CNVs. This design was adapted to the switch from DNA chips to exomes, largely reducing the size of the smaller CNVs under analysis. Our study design was thus adapted to the identification of single gene association. Despite this design, the most significant finding was a multiple-gene recurrent CNV mediated by NAHR. The low number of single-gene associations is probably related to a power issue in finding a burden of ultra-rare CNVs in a given gene, but our study design may be successful in finding such signals in future larger studies. In a replication attempt, we mostly relied on chip data that do not allow the calling of smaller CNVs, such as recurrent *FADS6* deletions. Genome sequencing may also be used, although at a high computational cost and with still limited sample sizes, especially in terms of EOAD case data available.[19,20] Recent long-read sequencing technologies may also offer new opportunities, enabling the analysis of CNVs along with other complex variants,[57,58] but sequencing costs remain high, limiting the sample sizes. In addition, we did not study the X chromosome, due to the complexity of the analysis models. Specific analysis of this chromosome may reveal potential variants associated with AD.[59]

In conclusion, we identified ultra-rare CNVs directly contributing to AD risk, including autosomal-dominant CNVs, deletions in AD-associated *ABCA1* and *ABCA7* contributing to AD risk in non-monogenic AD, *CTSB* LoF as a candidate risk mechanism, and a small region at the 22q11.21 locus, deletions of which increase EOAD risk while duplications decrease AD risk. We are providing multiple layers of evidence, including a clear mirror/dosage signal with LOAD-affected individuals showing intermediate frequencies of deletions and duplications between EOAD-affected individuals and unaffected control individuals, a large replication with a multi-ethnic component, exome-wide statistical significance, one *de novo* deletion in an EOAD individual, and one piece of mechanistic evidence, all converging on the same interpretation. This result may be important for the follow-up of some individuals with DiGeorge syndrome who

may present increased risk of AD. In addition, we identified a strong AD-risk-decreasing factor, 22q11.21 duplications. *SCARF2*, located in the prioritized region at the 22q11.21 locus, may be a therapeutic target, as overexpression increased Aβ uptake, and hence most likely Aβ clearance, which is congruent with reduced AD risk observed in the case of a duplication. While additional CNVs may contribute to AD risk, as suggested by burden CNV analyses at the exome and AD-related gene lists, such CNVs remain extremely rare events, limiting power.

Data and code availability

All computing codes are available at https://github.com/U1245. CNV counts (deletions and full duplications) in all genes are available as Data S2. Access to de-identified raw data of the ADES study can be requested using the Alzheimer Genetics Hub. The authors do not own data from ADSP and UK Biobank, which can be requested through their respective websites following specific procedures. CNV calling data have been deposited on the NIAGADS portal (dataset page: https://dss.niagads.org/datasets/ng00176/; study page: https://dss.niagads.org/studies/sa000081/).

Acknowledgments

This research was conducted using the funding obtained by the following study cohorts: ADES-FR, AgeCoDe-UKBonn, Barcelona SPIN, AC-EMC, ERF and Rotterdam, ADC-Amsterdam, 100-plus study, EMIF-90+, Control Brain Consortium, PERADES, UCL-DRC EOAD, and ADSP. Full consortium acknowledgments and funding sources are listed in the supplemental information.

We thank the ADSP study for providing exome data from affected individuals and control subjects (ADSP umbrella NG00067.v3). ADSP data were prepared, archived, and distributed by the National Institute on Aging Alzheimer Disease Data Storage Site (NIAGADS) at the University of Pennsylvania (U24AG041689), funded by the National Institute on Aging. The ADSP umbrella contains data from the Alzheimer Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in the analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

This research has been conducted using the UK Biobank Resource under application number 63653 (to G.N.) and 40980 (to S.J.). C.S. was supported by the Fondation pour la Recherche Médicale (ARF201909009263) and the Rouen metropole through the "post-doctorants métropole" funding. We thank the CEREBRO cluster at the University of Rouen Normandie, the high-performance computing service at the University of Lille, and the Cartesius supercomputer, which is embedded in the Dutch national e-infrastructure with the support of the SURF Cooperative. Additional funding information and acknowledgments are available in the supplemental information.

Author contributions

G.N. conceptualized and designed the study; O.Q., C.S., K.C., A.N., J.G., G.H., B.G.-B., M.H., G.B.W., I.D.R., A.R.-L., S.F., L.M., A.-C.R., S.R., S.A., N.A., P.A., O.B., C. Bellenguez, C. Berr, P.B., F.B., J.B., J.C., A.D., J.-F. Dartigues, S.D., J.-F. Deleuze, N.D., O.D.-I., C.M.v.D., J.F., N.C.F., R.F.-S., D.G., E.G., R. Ghidoni, V.G., J.J.P.G., D. Grozova, R. Guerreiro, E. Grünblatt, J.H., S.G.R.-H., M. Hiltunen, C.H., J. Hort, H. Hummerich, M.A.I., M.K.I., M.I., I.E.J., A.K., R.K., P.G.K., M. Lathrop, M. Lacour, A.W.L., A.L., L.L., M.M.A.M.M., R.M., C.M., S. Mead, P.M., A.d.M., A. Meggy, S. Mehrabian, M.O.M., K.M., A. Morin, B.N., P.J.N., R.O., F.P., P.P., F. Piras, J.P., A. Ramirez, R. Raybould, R. Redon, M.J.T.R., F.R., J.G.J.v.R., N.S.R., S. Saad, P.S.-J., N.S., P.S., J.M.S., D. Seripa, D. Sie, R.S., E.A.S., S. Sorbi, K.S., R.v.S., J.C.v.S., N.T., B.M.T., M.T., A.G.U., J.V., P.J.V., M.W., J.W., A.Z., H. Holstege, D.F.G., D.W., M. Lecourtois, M.V.F., H.S., S.J., J.-C.L., S.J.v.d.L., C.C., and G.N. had a major role in the acquisition of data; O.Q., C.S., K.C., A.N., J.G., G.H., B.G.-B., G.B.W., I.d.R., A.R.-L., S.F., L.M., M. Lecourtois, C.C., H.S., and G.N. had a major role in data analysis and statistics; G.N., O.Q., and C.S. interpreted the data; G.N., O.Q., and C.S. drafted the manuscript; and O.Q., C.S., J.G., G.H., I.d.R., G.B.W., A.R.-L., M. Lecourtois, D.W., J.-C.L., C.C., E. Génin, S.J.v.d.L., and G.N. revised the manuscript.

Declaration of interests

The authors declare no competing interests.

Web resources

Alzheimer Genetics Hub, https://alzheimergenetics.org/
CANOES, https://github.com/david-a-parry/canoes, https://github.com/U1245/canoes-centered-workflow
Database of Genomic Variants (DGV), https://dgv.tcag.ca/gnomAD, https://gnomad.broadinstitute.org/
GenBank, https://www.ncbi.nlm.nih.gov/genbank/
GitHub repository, https://github.com/U1245
NIAGADS, https://dss.niagads.org/
OMIM, https://www.omim.org/

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2026.07.013.

Received: April 20, 2026
Accepted: July 27, 2026
Published: August 26, 2026

References

1. Nicolas, G. (2024). Recent advances in Alzheimer disease genetics. Curr. Opin. Neurol. *37*, 154–165. https://doi.org/10.1097/WCO.0000000000001242.
2. Bellenguez, C., Küçükali, F., Jansen, I.E., Kleineidam, L., Moreno-Grau, S., Amin, N., Naj, A.C., Campos-Martin, R., Grenier-Boley, B., Andrade, V., et al. (2022). New insights into the genetic etiology of Alzheimer's disease and related dementias. Nat. Genet. *54*, 412–436. https://doi.org/10.1038/s41588-022-01024-z.
3. Nicolas, G. (2024). Lessons from genetic studies in Alzheimer disease. Rev. Neurol. *180*, 368–377. https://doi.org/10.1016/j.neurol.2023.12.006.

4. Holstege, H., Hulsman, M., Charbonnier, C., Grenier-Boley, B., Quenez, O., Grozeva, D., van Rooij, J.G.J., Sims, R., Ahmad, S., Amin, N., et al. (2022). Exome sequencing identifies rare damaging variants in ATP8B4 and ABCA1 as risk factors for Alzheimer's disease. Nat. Genet. *54*, 1786–1794. https://doi.org/10.1038/s41588-022-01208-7.
5. Campion, D., Charbonnier, C., and Nicolas, G. (2019). SORL1 genetic variants and Alzheimer disease risk: a literature review and meta-analysis of sequencing data. Acta Neuropathol. *138*, 173–186. https://doi.org/10.1007/s00401-019-01991-4.
6. Coe, B.P., Witherspoon, K., Rosenfeld, J.A., van Bon, B.W.M., Vulto-van Silfhout, A.T., Bosco, P., Friend, K.L., Baker, C., Buono, S., Vissers, L.E.L.M., et al. (2014). Refining analyses of copy number variation identifies specific genes associated with developmental delay. Nat. Genet. *46*, 1063–1071. https://doi.org/10.1038/ng.3092.
7. Niestroj, L.M., Perez-Palma, E., Howrigan, D.P., Zhou, Y., Cheng, F., Saarentaus, E., Nürnberg, P., Stevelink, R., Daly, M.J., Palotie, A., et al. (2020). Epilepsy subtype-specific copy number burden observed in a genome-wide study of 17 458 subjects. Brain *143*, 2106–2118. https://doi.org/10.1093/brain/awaa171.
8. Auwerx, C., Jõeloo, M., Sadler, M.C., Tesio, N., Ojavee, S., Clark, C.J., Mägi, R., Estonian Biobank Research Team, Reymond, A., and Kutalik, Z. (2024). Rare copy-number variants as modulators of common disease susceptibility. Genome Med. *16*, 5. https://doi.org/10.1186/s13073-023-01265-5.
9. Rovelet-Lecrux, A., Hannequin, D., Raux, G., Le Meur, N., Laquerrière, A., Vital, A., Dumanchin, C., Feuillette, S., Brice, A., Vercelletto, M., et al. (2006). APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. Nat. Genet. *38*, 24–26. https://doi.org/10.1038/ng1718.
10. Cabrejo, L., Guyant-Maréchal, L., Laquerrière, A., Vercelletto, M., De la Fournière, F., Thomas-Antérion, C., Verny, C., Letournel, F., Pasquier, F., Vital, A., et al. (2006). Phenotype associated with APP duplication in five families. Brain *129*, 2966–2976. https://doi.org/10.1093/brain/awl237.
11. Lanoiselée, H.M., Nicolas, G., Wallon, D., Rovelet-Lecrux, A., Lacour, M., Rousseau, S., Richard, A.C., Pasquier, F., Rollin-Sillaire, A., Martinaud, O., et al. (2017). APP, PSEN1, and PSEN2 mutations in early-onset Alzheimer disease: A genetic screening study of familial and sporadic cases. PLoS Med. *14*, e1002270. https://doi.org/10.1371/journal.pmed.1002270.
12. Fortea, J., Vilaplana, E., Carmona-Iragui, M., Benejam, B., Videla, L., Barroeta, I., Fernández, S., Altuna, M., Pegueroles, J., Montal, V., et al. (2020). Clinical and biomarker changes of Alzheimer's disease in adults with Down syndrome: a cross-sectional study. Lancet *395*, 1988–1997. https://doi.org/10.1016/S0140-6736(20)30689-9.
13. Le Guennec, K., Veugelen, S., Quenez, O., Szaruga, M., Rousseau, S., Nicolas, G., Wallon, D., Fluchere, F., Frébourg, T., De Strooper, B., et al. (2017). Deletion of exons 9 and 10 of the Presenilin 1 gene in a patient with Early-onset Alzheimer Disease generates longer amyloid seeds. Neurobiol. Dis. *104*, 97–103. https://doi.org/10.1016/j.nbd.2017.04.020.
14. Brouwers, N., Van Cauwenberghe, C., Engelborghs, S., Lambert, J.C., Bettens, K., Le Bastard, N., Pasquier, F., Montoya, A.G., Peeters, K., Mattheijssens, M., et al. (2012). Alzheimer risk associated with a copy number variation in the complement receptor 1 increasing C3b/C4b binding sites. Mol. Psychiatry *17*, 223–233. https://doi.org/10.1038/mp.2011.24.
15. Wang, H., Wang, L.S., Schellenberg, G., and Lee, W.P. (2022). The role of structural variations in Alzheimer's disease and other neurodegenerative diseases. Front. Aging Neurosci. *14*, 1073905. https://doi.org/10.3389/fnagi.2022.1073905.
16. Hooli, B.V., Kovacs-Vajna, Z.M., Mullin, K., Blumenthal, M.A., Mattheisen, M., Zhang, C., Lange, C., Mohapatra, G., Bertram, L., and Tanzi, R.E. (2014). Rare autosomal copy number variations in early-onset familial Alzheimer's disease. Mol. Psychiatry *19*, 676–681. https://doi.org/10.1038/mp.2013.77.
17. Rovelet-Lecrux, A., Legallic, S., Wallon, D., Flaman, J.M., Martinaud, O., Bombois, S., Rollin-Sillaire, A., Michon, A., Le Ber, I., Pariente, J., et al. (2012). A genome-wide study reveals rare CNVs exclusive to extreme phenotypes of Alzheimer disease. Eur. J. Hum. Genet. *20*, 613–617. https://doi.org/10.1038/ejhg.2011.225.
18. Collins, R.L., Brand, H., Karczewski, K.J., Zhao, X., Alföldi, J., Francioli, L.C., Khera, A.V., Lowther, C., Gauthier, L.D., Wang, H., et al. (2020). A structural variation reference for medical and population genetics. Nature *581*, 444–451. https://doi.org/10.1038/s41586-020-2287-8.
19. Lee, W.P., Tucci, A.A., Conery, M., Leung, Y.Y., Kuzma, A.B., Valladares, O., Chou, Y.F., Lu, W., Wang, L.S., Schellenberg, G.D., and Tzeng, J.Y. (2021). Copy Number Variation Identification on 3,800 Alzheimer's Disease Whole Genome Sequencing Data from the Alzheimer's Disease Sequencing Project. Front. Genet. *12*, 752390. https://doi.org/10.3389/fgene.2021.752390.
20. Wang, H., Dombroski, B.A., Cheng, P.L., Tucci, A., Si, Y.Q., Farrell, J.J., Tzeng, J.Y., Leung, Y.Y., Malamon, J.S., Wang, L., et al. (2025). Structural variation detection and association analysis of whole-genome-sequence data from 16,543 Alzheimer's disease sequencing project subjects. Alzheimer's. &. Dementia *21*, e70277. https://doi.org/10.1002/alz.70277.
21. Hehir-Kwa, J.Y., Pfundt, R., and Veltman, J.A. (2015). Exome sequencing and whole genome sequencing for the detection of copy number variation. Expert Rev. Mol. Diagn. *15*, 1023–1032. https://doi.org/10.1586/14737159.2015.1053467.
22. Gabrielaite, M., Torp, M.H., Rasmussen, M.S., Andreu-Sánchez, S., Vieira, F.G., Pedersen, C.B., Kinalis, S., Madsen, M.B., Kodama, M., Demircan, G.S., et al. (2021). A Comparison of Tools for Copy-Number Variation Detection in Germline Whole Exome and Whole Genome Sequencing Data. Cancers *13*, 6283. https://doi.org/10.3390/cancers13246283.
23. Backenroth, D., Homsy, J., Murillo, L.R., Glessner, J., Lin, E., Brueckner, M., Lifton, R., Goldmuntz, E., Chung, W.K., and Shen, Y. (2014). CANOES: detecting rare copy number variants from whole exome sequencing data. Nucleic Acids Res. *42*, e97. https://doi.org/10.1093/nar/gku345.
24. Quenez, O., Cassinari, K., Coutant, S., Lecoquierre, F., Le Guennec, K., Rousseau, S., Richard, A.C., Vasseur, S., Bouvignies, E., Bou, J., et al. (2021). Detection of copy-number variations from NGS data using read depth information: a diagnostic performance evaluation. Eur. J. Hum. Genet. *29*, 99–109. https://doi.org/10.1038/s41431-020-0672-2.
25. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.
26. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel,

S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303. https://doi.org/10.1101/gr.107524.110.

27. McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., and Stadlan, E.M. (1984). Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology *34*, 939–944.
28. McKhann, G.M., Knopman, D.S., Chertkow, H., Hyman, B.T., Jack, C.R., Jr., Kawas, C.H., Klunk, W.E., Koroshetz, W.J., Manly, J.J., Mayeux, R., et al. (2011). 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. Alzheimer's Dement. *7*, 263–269. https://doi.org/10.1016/j.jalz.2011.03.005.
29. Teo, S.M., Pawitan, Y., Ku, C.S., Chia, K.S., and Salim, A. (2012). Statistical challenges associated with detecting copy number variations with next-generation sequencing. Bioinformatics *28*, 2711–2718. https://doi.org/10.1093/bioinformatics/bts535.
30. Zhao, M., Wang, Q., Wang, Q., Jia, P., and Zhao, Z. (2013). Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. BMC Bioinf. *14*, S1. https://doi.org/10.1186/1471-2105-14-S11-S1.
31. MacDonald, J.R., Ziman, R., Yuen, R.K.C., Feuk, L., and Scherer, S.W. (2014). The Database of Genomic Variants: a curated collection of structural variation in the human genome. Nucleic Acids Res. *42*, D986–D992. https://doi.org/10.1093/nar/gkt958.
32. Campion, D., Pottier, C., Nicolas, G., Le Guennec, K., and Rovelet-Lecrux, A. (2016). Alzheimer disease: modeling an Abeta-centered biological network. Mol. Psychiatr. *21*, 861–871. https://doi.org/10.1038/mp.2016.38.
33. Le Guennec, K., Quenez, O., Nicolas, G., Wallon, D., Rousseau, S., Richard, A.C., Alexander, J., Paschou, P., Charbonnier, C., Bellenguez, C., et al. (2017). 17q21.31 duplication causes prominent tau-related dementia with increased MAPT expression. Mol. Psychiatry *22*, 1119–1125. https://doi.org/10.1038/mp.2016.226.
34. Rovelet-Lecrux, A., Charbonnier, C., Wallon, D., Nicolas, G., Seaman, M.N., Pottier, C., Breusegem, S.Y., Mathur, P.P., Jenardhanan, P., Le Guennec, K., et al. (2015). De novo deleterious genetic variations target a biological network centered on Abeta peptide in early-onset Alzheimer disease. Mol. Psychiatr. *20*, 1046–1056. https://doi.org/10.1038/mp.2015.100.
35. Wallon, D., Boluda, S., Rovelet-Lecrux, A., Thierry, M., Lagarde, J., Miguel, L., Lecourtois, M., Bonnevalle, A., Sarazin, M., Bottlaender, M., et al. (2021). Clinical and neuropathological diversity of tauopathy in MAPT duplication carriers. Acta Neuropathol. *142*, 259–278. https://doi.org/10.1007/s00401-021-02320-4.
36. Steinberg, S., Stefansson, H., Jonsson, T., Johannsdottir, H., Ingason, A., Helgason, H., Sulem, P., Magnusson, O.T., Gudjonsson, S.A., Unnsteinsdottir, U., et al. (2015). Loss-of-function variants in ABCA7 confer risk of Alzheimer's disease. Nat. Genet. *47*, 445–447. https://doi.org/10.1038/ng.3246.
37. Wilkinson, K., and El Khoury, J. (2012). Microglial scavenger receptors and their roles in the pathogenesis of Alzheimer's disease. Int. J. Alzheimers Dis. *2012*, 489456. https://doi.org/10.1155/2012/489456.
38. Singh, T.D., Park, S.Y., Bae, J.S., Yun, Y., Bae, Y.C., Park, R.W., and Kim, I.S. (2010). MEGF10 functions as a receptor for the uptake of amyloid-beta. FEBS Lett. *584*, 3936–3942. https://doi.org/10.1016/j.febslet.2010.08.050.
39. Lee, J.H., Cheng, R., Graff-Radford, N., Foroud, T., Mayeux, R.; and National Institute on Aging Late-Onset Alzheimer's Disease Family Study Group (2008). Analyses of the National Institute on Aging Late-Onset Alzheimer's Disease Family Study: implication of additional loci. Arch. Neurol. *65*, 1518–1526. https://doi.org/10.1001/archneur.65.11.1518.
40. Jonsson, T., Atwal, J.K., Steinberg, S., Snaedal, J., Jonsson, P.V., Bjornsson, S., Stefansson, H., Sulem, P., Gudbjartsson, D., Maloney, J., et al. (2012). A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. Nature *488*, 96–99. https://doi.org/10.1038/nature11283.
41. Belloy, M.E., Andrews, S.J., Le Guen, Y., Cuccaro, M., Farrer, L.A., Napolioni, V., and Greicius, M.D. (2023). APOE Genotype and Alzheimer Disease Risk Across Age, Sex, and Population Ancestry. JAMA Neurol. *80*, 1284–1294. https://doi.org/10.1001/jamaneurol.2023.3599.
42. Sims, R., van der Lee, S.J., Naj, A.C., Bellenguez, C., Badarinarayan, N., Jakobsdottir, J., Kunkle, B.W., Boland, A., Raybould, R., Bis, J.C., et al. (2017). Rare coding variants in PLCG2, ABI3, and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease. Nat. Genet. *49*, 1373–1384. https://doi.org/10.1038/ng.3916.
43. Brutman, J.N., Busald, T., Nizamis, E., Kaufman, E.J., Lin, E., Hendricks, N.E., Chintamen, S., Sulovari, A., Smukowski, S.N., Ye, Y., et al. (2026). A common 19 bp APOE enhancer deletion is protective against Alzheimer's disease in African Americans. Nat. Commun. *17*, 2237. https://doi.org/10.1038/s41467-026-68808-3.
44. Nicolas, G., Zarea, A., Lacour, M., Quenez, O., Rousseau, S., Richard, A.C., Bonnevalle, A., Schramm, C., Olaso, R., Sandron, F., et al. (2024). Assessment of Mendelian and risk-factor genes in Alzheimer disease: A prospective nationwide clinical utility study and recommendations for genetic screening. Genet. Med. *26*, 101082. https://doi.org/10.1016/j.gim.2024.101082.
45. Alver, M., Mancini, V., Läll, K., Schneider, M., Romano, L., Estonian Biobank Research Team, Mägi, R., Dermitzakis, E.T., Eliez, S., and Reymond, A. (2022). Contribution of schizophrenia polygenic burden to longitudinal phenotypic variance in 22q11.2 deletion syndrome. Mol. Psychiatry *27*, 4191–4200. https://doi.org/10.1038/s41380-022-01674-9.
46. Rump, P., de Leeuw, N., van Essen, A.J., Verschuuren-Bemelmans, C.C., Veenstra-Knol, H.E., Swinkels, M.E.M., Oostdijk, W., Ruivenkamp, C., Reardon, W., de Munnik, S., et al. (2014). Central 22q11.2 deletions. Am. J. Med. Genet. *164A*, 2707–2723. https://doi.org/10.1002/ajmg.a.36711.
47. Huguet, G., Schramm, C., Douard, E., Jiang, L., Labbe, A., Tihy, F., Mathonnet, G., Nizard, S., Lemyre, E., Mathieu, A., et al. (2018). Measuring and Estimating the Effect Sizes of Copy Number Variants on General Intelligence in Community-Based Samples. JAMA Psychiatry *75*, 447–457. https://doi.org/10.1001/jamapsychiatry.2018.0039.
48. Stern, Y. (2012). Cognitive reserve in ageing and Alzheimer's disease. Lancet Neurol. *11*, 1006–1012. https://doi.org/10.1016/S1474-4422(12)70191-6.

49. Nelson, M.E., Jester, D.J., Petkus, A.J., and Andel, R. (2021). Cognitive Reserve, Alzheimer's Neuropathology, and Risk of Dementia: A Systematic Review and Meta-Analysis. Neuropsychol. Rev. *31*, 233–250. https://doi.org/10.1007/s11065-021-09478-4.
50. Butcher, N.J., Kiehl, T.R., Hazrati, L.N., Chow, E.W.C., Rogaeva, E., Lang, A.E., and Bassett, A.S. (2013). Association between early-onset Parkinson disease and 22q11.2 deletion syndrome: identification of a novel genetic form of Parkinson disease and its clinical implications. JAMA Neurol. *70*, 1359–1366. https://doi.org/10.1001/jamaneurol.2013.3646.
51. Boot, E., Oskarsdottir, S., Loo, J.C.Y., Crowley, T.B., Orchanian-Cheff, A., Andrade, D.M., Arganbright, J.M., Castelein, R.M., Cserti-Gazdewich, C., de Reuver, S., et al. (2023). Updated clinical practice recommendations for managing adults with 22q11.2 deletion syndrome. Genet. Med. *25*, 100344. https://doi.org/10.1016/j.gim.2022.11.012.
52. Grangeon, L., Cassinari, K., Rousseau, S., Croisile, B., Formaglio, M., Moreaud, O., Boutonnat, J., Le Meur, N., Miné, M., Coste, T., et al. (2021). Early-Onset Cerebral Amyloid Angiopathy and Alzheimer Disease Related to an APP Locus Triplication. Neurol. Genet. *7*, e609. https://doi.org/10.1212/NXG.0000000000000609.
53. Sleegers, K., Brouwers, N., Gijselinck, I., Theuns, J., Goossens, D., Wauters, J., Del-Favero, J., Cruts, M., van Duijn, C.M., and Van Broeckhoven, C. (2006). APP duplication is sufficient to cause early onset Alzheimer's dementia with cerebral amyloid angiopathy. Brain *129*, 2977–2983. https://doi.org/10.1093/brain/awl203.
54. Grangeon, L., Charbonnier, C., Zarea, A., Rousseau, S., Rovelet-Lecrux, A., Bendetowicz, D., Lemaitre, M., Malrain, C., Quillard-Muraine, M., Cassinari, K., et al. (2023). Phenotype and imaging features associated with APP duplications. Alz. Res. Therapy *15*, 93. https://doi.org/10.1186/s13195-023-01172-2.
55. Ehn, E., Eisfeldt, J., Laffita-Mesa, J.M., Thonberg, H., Schoumans, J., Portaankorva, A.M., Viitanen, M., Lindstrand, A., Nennesmo, I., and Graff, C. (2025). A de novo, mosaic and complex chromosome 21 rearrangement causes APP triplication and familial autosomal dominant early onset Alzheimer disease. Sci. Rep. *15*, 2912. https://doi.org/10.1038/s41598-025-86645-0.
56. Mueller-Steiner, S., Zhou, Y., Arai, H., Roberson, E.D., Sun, B., Chen, J., Wang, X., Yu, G., Esposito, L., Mucke, L., and Gan, L. (2006). Antiamyloidogenic and neuroprotective functions of cathepsin B: implications for Alzheimer's disease. Neuron *51*, 703–714. https://doi.org/10.1016/j.neuron.2006.07.027.
57. Logsdon, G.A., Vollger, M.R., and Eichler, E.E. (2020). Long-read human genome sequencing and its applications. Nat. Rev. Genet. *21*, 597–614. https://doi.org/10.1038/s41576-020-0236-x.
58. Miller, D.E., Sulovari, A., Wang, T., Loucks, H., Hoekzema, K., Munson, K.M., Lewis, A.P., Fuerte, E.P.A., Paschal, C.R., Walsh, T., et al. (2021). Targeted long-read sequencing identifies missing disease-causing variation. Am. J. Hum. Genet. *108*, 1436–1449. https://doi.org/10.1016/j.ajhg.2021.06.006.
59. Le Borgne, J., Gomez, L., Heikkinen, S., Amin, N., Ahmad, S., Choi, S.H., Bis, J., Grenier-Boley, B., Rodriguez, O.G., Kleineidam, L., et al. (2025). X-chromosome-wide association study for Alzheimer's disease. Mol. Psychiatry *30*, 2335–2346. https://doi.org/10.1038/s41380-024-02838-5.