Exome analysis of 22,319 individuals links extremely rare CNVs and 22q11.21 dosage to Alzheimer's risk

Olivier Quenez*[1], Catherine Schramm*[2], 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], 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], 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], Gaël Nicolas[1]

*equal contribution

1. Univ Rouen Normandie, Normandie Univ, Inserm U1245 and CHU Rouen, Department of Genetics and CNRMAJ, F-76000 Rouen, France.
2. Univ Rouen Normandie, Normandie Univ, Inserm U1245, F-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, Quebec, 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, GrandRapids, 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, 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 (locationVUmc), 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, Quebec, Canada.
50. Univ Rouen Normandie, Normandie Univ, Inserm U1245 and CHU Rouen, Department of Neurology and CNRMAJ, F-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 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, Quebec, Canada.
88. Univ Rouen Normandie, Normandie Univ, Inserm U1245 and CHU Rouen, Department of Biostatistics and CNRMAJ, F-76000 Rouen, France.

Corresponding author: Gaël Nicolas, MD, PhD - gaelnicolas@hotmail.com

Abstract

Copy number variants (CNVs), defined as deletions or duplications of genomic segments >100 bp, are major contributors to human disease, yet their role in non-monogenic Alzheimer disease (AD) remains poorly characterized. We analyzed rare CNVs from 22,319 exomes, including 4,150 early-onset AD (EOAD), 8,519 late-onset AD and 9,650 unaffected controls. EOAD cases showed increased burdens of rare CNVs, particularly rare deletions within AD-related genes. Loss-of-function analyses implicated deletions in *ABCA1* and *ABCA7* and identified *CTSB* as a candidate *locus*. Exome-wide gene-level dosage analysis highlighted 18 genes across five *loci*, including chr22q11.21, where deletions were restricted to EOAD and duplications enriched in controls. Replication in 33,992 cases and 362,305 controls confirmed AD-risk reduction in 22q11.21 duplication carriers (mega-analysis dosage OR(SCARF2)=0.34, $p=5.52x10^{-7}$). *SCARF2* overexpression increased amyloid-β uptake *in vitro*, supporting a functional link. These findings highlight CNVs as contributors to AD risk and identify 22q11.21 dosage as a strong genetic determinant.

Keywords: Alzheimer disease, deletion, 22q11, duplication, association study, burden, copy number variant

INTRODUCTION

The etiology of Alzheimer disease (AD) includes autosomal dominant AD (ADAD), caused by pathogenic variants in *PSEN1*, *PSEN2* or *APP* and representing <1% of all cases, and non-monogenic (or complex AD), combining genetic and non-genetic factors. Genetic factors play a prominent role in non-monogenic AD, and the etiology of non-monogenic early-onset AD (EOAD, onset ≤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 ones, common variants all show modest odds ratios (ORs) below 1.5[2], except the *APOE-ε4* allele with ORs ranging from ~3 to ~14 depending on ancestry and zygosity, despite a high frequency (9-23%)[1]. Multiple rare (allele frequency<1%) coding variants in five genes (*SORL1*, *TREM2*, *ABCA7*, *ABCA1* and *ATP8B4*), mostly identified by exome sequencing, are associated with higher odds ratios (ORs), ranging from 1.5 to over 20 and thus contribute significantly to the risk of AD in carriers[1, 3]. Recent findings from exome or genome sequencing studies show that (i) the rarer the variants, the higher the ORs, and (ii) rare variants with strong effects on AD risk are significantly enriched in EOAD cases compared to late-onset (LOAD) patients[3, 4]. In contrast with AD risk-increasing variants, only a handful of variants have been significantly associated with strong effects on AD risk reduction (OR<0.7), so far.

Copy number variants (CNVs, deletions or duplications of >100-bp genomic DNA regions) represent a variant type with a major role in multiple diseases[5-7]. In AD, CNVs have been less studied than SNVs and indels. *APP* locus duplications on chromosome 21 are typical examples of CNVs causing ADAD[8, 9], accounting for ~9% of all ADAD cases (update of ref.[10]). Importantly, they also underscore why Trisomy-21 patients develop AD with full penetrance by age 70[11]. 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[12], while a few common CNVs, detected by DNA chips, have emerged as candidate risk factors[13, 14]. 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 patients,[15, 16] 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[17].

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 cases (average age at onset: 74.6 years, mainly LOAD patients) and 6,855 controls from the Alzheimer Disease Sequencing Project (ADSP) consortium[18, 19]. 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 in the respective carriers, including one partial deletion of *SORL1* and five partial deletions of *ABCA7*, each identified in one patient. 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[17]. This makes it necessary to analyze larger datasets and/or enrich them in patients with higher suspected genetic component, as EOAD patients. Despite the development of genome sequencing datasets, a large proportion of the data available so far is made of 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[20, 21]. For example, we previously assessed the performances of a CNV-calling workflow based on the CANOES tool[22], 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[23]. This makes large exome datasets particularly well suited for gene-level CNV association analyses

To identify CNVs associated with AD risk, we used one of the largest sequencing datasets available worldwide, i.e. the combination of the European and American consortia into the ADES-ADSP dataset (Alzheimer Disease European Sequencing – Alzheimer Disease Sequencing Project) as it is enriched in EOAD patients (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 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 CNV carriers, then performed burden analyses at the protein-coding genome level and after restriction to AD gene lists, and finally performed a genome-wide case-control association study at the gene level,

different from any previous study 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.

RESULTS

We included all 22,319 exomes from the combined ADES-ADSP dataset as described in ref.[3], made of European and American cohorts, and proceeded to harmonized CNV detection, sample and variant quality control and filtration, followed by variant interpretation in a Mendelian setting and then, case-control association analyses (see online methods and detailed methods in supplementary information.

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

Out of 22,319 exomes initially included, 22,110 passed QC (4093 EOAD cases, 8458 LOAD cases (late-onset AD) and, 9559 unaffected controls) (Table 1, Figure 1, supplementary information, Tables S1-S3), totaling 261,190 CNVs. The first step consisted in searching for pathogenic CNVs in monogenic dementia genes. We identified 17 carriers of such a pathogenic CNV, all but one with EOAD. Nine cases were novel, as 8 were previously reported[10, 12, 24-26]. Overall, we detected 10 *APP* duplications (mean age at onset (AAO) = 49.6 years, min-max: [43-58]), one in-frame *PSEN1* deletion (AAO=56 years), and six *MAPT* duplications, leading to a primary tauopathy differential diagnosis (mean AAO = 58.2, min-max: [45-87]) (Table 2). Carriers of 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 if EOAD patients exhibit a higher burden of rare CNVs overall, as compared to controls, at two levels (Table S4). First, we performed an exome-wide burden CNV analysis across all protein-coding genes and showed that EOAD patients have a higher likelihood of carrying at least one deleted (set A, non-frequently deleted genes, see methods) or one duplicated (set B, non-frequently duplicated genes) gene than controls (deletions burden: OR=1.15 [1.06; 1.23], $p=3\times10^{-4}$; complete duplications burden: OR=1.10 [1.02; 1.19],

p=0.0125). Second, restricting to AD-GWAS genes[2], the difference was amplified for deletions (OR=2.67 [1.51; 4.74], $p=8\times10^{-4}$) and reversed for duplications (OR=0.58 [0.30; 1.04], p-value=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[27] indicated enrichment in deletions among EOAD cases (OR=1.44 [1.09; 1.90], p=0.0114, Table S4).

Then, to further assess the deletions-related signals, we performed a gene-based integrated loss-of-function (LOF) analysis on a list of known AD-associated genes, gathering short truncating variants (nonsense, frameshift, canonical splice site SNVs/indels) with CNV-deletions (complete and partial) (see online methods and supplementary information). We identified a significant enrichment of *ABCA1* (OR_{EOAD}=5.77 [2.25 ; 17.06], $p=2x10^{-4}$) and *ABCA7* (OR_{EOAD}=2.29 [1.44 ; 3.65], $p=6x10^{-6}$) LOF alleles with AD, now adding CNV-deletions to the known *ABCA7* association[28], while the *CTSB* gene, mapping to a GWAS locus, emerged as a novel candidate (OR_{EOAD}=5.03 [1.50 ; 20.71], $p=8.9x10^{-3}$) (Tables S5, S6, additional results in supplementary information).

Gene-level case-control discovery analysis

Then, we performed a gene-level rare CNV association study on the 22,093 samples remaining after exclusion of pathogenic CNV carriers. By Firth's logistic regression, we compared the gene dosage (partial/complete deletions = decreased dosage, complete duplications = increased dosage) between EOAD cases and controls. In dosage analysis, ORs should be interpreted for an increase of gene copy by one, i.e. an OR >1 (resp. OR <1) means that duplications (resp. deletions) are associated with a higher disease risk (Figure 1, Figure 2, Table 3). Loci with False Discovery Rate (FDR) below 10% were then assessed in the whole cohort using an ordinal regression including EOAD, LOAD and controls (Table S7). Additional analyses included deletion-only (Table S8) or full duplication-only (Table S9) analyses, *APOE4*-adjusted (Table S10) and population structure-adjusted (Table S11) analyses. All analyses are transcript based, and results are represented per gene when several transcripts show identical results in a given gene.

A total of 4,989 transcripts of 1,716 genes fulfilled our criteria for dosage association analysis, i.e. belonging to non-frequently deleted or duplicated genes, showing at least 4 rare CNVs in the dataset, including at least one (partial or complete) deletion and one complete duplication.

Of them, dosage analysis comparing EOAD patients to unaffected controls 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-KLC3* locus, 0.0271<FDR<0.0412), and three genes were excluded (Table 3, supplementary information).

Three single-gene loci showed an FDR≤10%, namely *MBL2* (FDR=0.0271), *FADS6* (FDR=0.0271) and *ADI1* (FDR=0.0916). The *MBL2* signal did not remain after adjusting for population structure and same adjustment decreased the significance for *FADS6* while it did not affect the other loci (Figure S14, Figure 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 chr19q13.32 (TableS 12, S10, 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) labelled LCR 22A to 22E (Figure 2). The region where all transcripts with FDR≤10% clustered is located between LCR 22B and 22D (labelled as the "central region"), which do not encompass the critical region for DiGeorge (velo-cardio-facial) syndrome (22A-22B). We observed 22q11.21 central region deletions in four EOAD cases (0.1%), and this region was also prioritized in the deletion-only analysis (Table S8; 0 control carrier, 0 to 1 (0.01%) LOAD carrier depending on the CNV coordinates). One of the EOAD cases 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 first signs of cognitive decline at age 43, without any signs of Parkinsonism. The AD diagnosis was confirmed with CSF biomarkers (Figure 2). 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 to 22E regions, were observed in 8 to 10 controls (0.08 to 0.1%, depending on the CNV coordinates), in 4 to 5 LOAD cases (0.05 to 0.06% Figure 2), and in 0 to 1 EOAD case (0 to 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 patients using an ordinal regression, the association signal remained nominally significant (OR=0.16 [0.04 ;

0.45] ($p=3.5\times10^{-4}$) (Table S7) as LOAD patients showed frequencies of deletions and duplications in between EOAD cases and controls.

Overall, DNA from 27 carriers (12 *ERCC2-KLC3*, two 22q11.2, one *MBL2*, three *ADI1* and nine *FADS6*) was available and CNVs were confirmed in 26/27 (one *FADS6* false positive (Table S14) - after removing the latter one, EOAD versus controls dosage analysis remained similar: OR=0.20 [0.07; 0.49], p-value=$4x10^{-4}$), 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[23].

Integrated loss-of-function analysis at the prioritized loci

Then, we performed a case-control integrated LOF analysis gathering CNV-deletions (complete and partial) with short truncating variants (nonsense, frameshift, canonical splice site SNVs/indels) among the genes belonging to the prioritized loci in the dosage analysis (Tables S6, S7). Although this integrated LOF analysis did not substantially modify the interpretation of the dosage analysis results, it narrowed the signal to the *SCARF2-MED15-KLHL22 locus* in the 22q11.21 central region. Indeed, we observed one LOF short variant in *SCARF2* in an EOAD case and 2 in *MED15* in 2 LOAD cases, whereas no control carried such LOF variants, as for CNV-deletions, and all the other genes at this locus appeared as less significant now.

Replication of suggestive loci in large external cohorts

We aimed at replicating the *SCARF2-MED15-KLHL22* locus dosage, *ADI1* duplications and *FADS6* deletions cases enrichment, using chip data from 29,763 AD cases (3411 EOAD) and 359,928 unaffected controls, and exomes from African American AD cases (n=1462 including 182 EOAD cases) and unaffected controls (n=2393). As the *FADS6* signal was driven by a small recurrent deletion in the discovery (<10kb, Figure S6), which is not detectable in DNA chip data, we were not able to perform an unbiased mega-analysis. Regarding *ADI1,* replication results were in the same direction as the discovery analysis (Figure 3, Table S15) but significance was not achieved (Tables S15, S16). However, we succeeded in replicating the *SCARF2-KLHL22-MED15* gene dosage association at the 22q11.21 locus, as ordinal regression confirmed an EOAD<LOAD<control association (Tables S15, S16 Figure 3, replication

OR_{SCARF2}=0.40 [0.23;0.66], p=1.50×10^{-4}), with a mega-analysis OR_{SCARF2}=0.34 [0.21;0.53] (p=5.52×10^{-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 22q11.21 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 the *SCARF2* gene, 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* and *MEGF10*, the latter even being a known receptor for Aβ[29, 30]. We therefore tested whether SCARF2 may influence Aβ cellular uptake by overexpressing *SCARF2* in HEK293 cells, and observed a significantly increased uptake of fluorescent Aβ42 peptides in SCARF2-transfected cells compared to mock and TDP-43-transfected cells, used as controls (Figure 4).

DISCUSSION

Here, we identified an association with gene dosage at the central region of chr 22q11.2. Overall, the 22q11.21 locus showed enrichment of deletions in EOAD cases and mirroring duplications in controls, while LOAD cases remained in between. This region was not known to be associated with AD risk so far, despite a suggestive association in a family-based analysis[31]. In our gene dosage analysis, the association was exome-wide significant (p<2.75x10^{-6}) and we are providing additional levels of evidence, with one of the patients with unaffected parents showing a *de novo* deletion, which is a strong genetic argument at the individual level, and a first biological argument on potential mechanism, as expression of *SCARF2* 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 controls. Notably, very few controls 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 non-carriers of CNV in this region. However, rarity of deletion carriers prevented us to perform 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 that such a deletion/duplication being driven by NAHR, they can be found as *de novo* events, hence limiting the interpretation of family history in this database for this locus. Beyond the fact that the penetrance of AD may not be full by the age of 65 in deletion carriers, most of those from the UK biobank are still rather young, so that some might still develop AD.

Interestingly, while our discovery was enriched in EOAD cases and highlighted at-risk deletions with mirroring duplications in controls, our replication was enriched in controls 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 A673T with a very strong effect (OR~0.2), and three other AD risk-decreasing factors with ORs~0.7: *APOE2*[32], a rare missense variant in *PLCG2*[33], and the very recent identification of a protective 19-bp deletion of an *APOE* enhancer in *APOE4*-positive African Americans, although carriers still show an increased AD risk compared to *APOE3* homozygozity[34]. Duplications of the 22q11.21 central region thus represent the second strongest AD-risk decreasing factor, after the A673T variant, which is both ultrarare and population specific, contrary to the 22q11.2, 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 controls, we also identified a few duplications in LOAD duplication carriers. Of the four LOAD duplication carriers in the discovery, 2 (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* (NM_019112.4:c.4764+5G>A), predicted to result in exon skipping and thus a probable AD risk factor with a moderate effect. As a comparison, 3/10 (30%) unaffected controls carrying a 22q11.21 duplication carried a similar AD risk factor (2 were *APOE4* heterozygous and one carried the *ATP8B4*:c.1183G>A, p.(Gly395Ser) classified as a modest risk factor with OR around 1.5)[35], 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[36].

In theory, two hypotheses might explain the association with gene dosage at the 22q11.21 central region. First, the association could be nonspecific to AD mechanisms but rather a modifier of AAO in individuals prone to develop AD. Indeed, carriers of the central 22q11.21 deletion have a higher risk of neurodevelopmental disorders[37], as do the carriers of DiGeorge-syndrome associated deletions (22A-22B or larger) with higher penetrance, and lower average cognitive performances[38]. Whether or not it is clinically significant prior to AD onset, putatively reduced cognitive reserve might contribute to earlier AD onset ages[39, 40]. However, this hypothesis would require that duplication carriers 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 duplication carriers.

From a medical point of view, our results may be of interest for DiGeorge patients, 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 Parkinson disease risk[41]. However, not much is known about the aging of these patients, as most reports in adults involve individuals in their mid-30s on average and limited survival until late adulthood[42]. At least one patient (aged 56) showed AD-related neuropathology[41], but neuropathological reports are scarce. Our results suggest that, as care of DiGeorge syndrome is improving, enabling patients 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 patients[8, 43, 44]. However, the impact of the latter may be much stronger, with full penetrance by the age of 65-70[11, 45], probably because of the straightforward link between *APP* overexpression and Aβ production, while Aβ clearance involves multiple players, *SCARF2* now becoming a strong candidate Aβ scavenger.

Beyond the 22q11.21 locus, our study also highlighted CNVs in known AD genes directly contributing to the disease in carriers. 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[43, 46] while *APP* duplications account for ~9% autosomal dominant EOAD families, and *MAPT* duplications represent a very rare cause of AD-mimicking primary tauopathy[26], as found in some patients 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 *CSTB* LOF with AD risk. *CTSB*, a GWAS-defined gene, encodes the Cathepsin B, an anti-amyloidogenic enzyme with Aβ peptide cleavage activity[47], which is consistent with LOF alleles in cases and duplications in controls (see also supplementary information).

Here, we used a large existing dataset of exome sequencing data. We propose a novel 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 (i) the identification of high-quality CNV calls in known AD genes (Mendelian and risk factor genes) and (ii) 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 association 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, as the recurrent *FASD6* deletions. Genome sequencing may also be used, although at a high computational cost and with still limited sample sizes, especially for in terms of EOAD case data available[18, 19]. Novel long read sequencing technologies may also offer new opportunities, enabling the analysis of CNVs along with other complex variants[48, 49], 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[50].

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* loss-of-function as 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 late-onset AD patients showing intermediate frequencies of deletions and duplications between EOAD and unaffected controls, a large replication with a multiethnic component, exome-wide statistical significance, one *de novo* deletion in an EOAD case, and a first mechanistic argument, all converging on the same interpretation. This result may be important for the follow up of some patients with DiGeorge syndrome who may present increased risk of AD. In addition, we identified the second strongest AD risk decreasing factor, 22q11.21 duplications. The *SCARF2* gene, located in the prioritized region at the chr22q11.21 locus, may be a therapeutic target, as overexpression increased Aβ uptake, hence most likely Aβ clearance, which is congruent with reduced AD risk in duplication carriers. While additional CNVs may contribute to AD risk, as suggested by burden CNV analyses at the exome and AD gene lists levels, such CNVs remain extremely rare events, limiting power.

ONLINE METHODS

Detailed methods are provided in supplementary information.

Exome dataset, CNV Calling, quality control, and pathogenic CNVs identification

All samples were selected following the quality control detailed in ref.[3], except pathogenic CNV carriers, who were not excluded from our starting dataset.

Only samples passing the quality control (QC) described in detail in ref.[3] were selected for our analysis. Briefly, for the individual QC based on short sequence variants, all sequencing (Fastq) files were processed using the same BWA-GATK-based pipeline 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 novel variants (compared to dbSNP v150), a deviation from standard heterozygous/homozygous or transition/transversion ratios, and relatives up to the 3rd degree were excluded. As cases, 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[51, 52]), 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 cases, or Braak stages of V-VI in controls). In addition, pathogenic SNV and indel variant carriers in a list of Mendelian dementia genes were already excluded from stage-1 data in this dataset. We now applied the same analysis leading to the exclusion of 22 additional pathogenic variant carriers among cases and controls from stage-2 so that the megasample from ref.[3] (stages 1+2) fulfils the same criteria (Table S1). Detailed methods for individual QC are available as a supplementary information file in ref.[3]. In order to homogenize EOAD/LOAD categorization across cohorts, we defined as EOAD, all cases with an age at onset (AAO) <66 years or age at last visit <66 if AAO is missing. Cases without numerical information for age were considered as LOAD.

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

All participants provided informed written consent, as previously described[3], and according to the declaration of Helsinki. This study, based on the retrospective analysis of existing data, was approved by the CERDE ethics committee from 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[53, 54]. 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[20, 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[22], which demonstrated good sensitivity (87.25 %) and positive predictive value (86.4-90.2%) for rare CNVs from exomes[23]. We adapted this workflow for a large dataset and applied it to BAM files. Samples with excessive calls were excluded, as were carriers of excessively large CNVs, which are unlikely germline, to avoid biases from clonal hematopoiesis associated with large

blood-specific, age-related CNVs. We then assessed the presence of pathogenic Mendelian CNVs. CNV calling and quality control are detailed in supplementary information.

CNV and transcripts filtering

For case-control analyses, we excluded CNVs cross-overlapping by ≥70% with common CNVs (frequency >1% in DGV-gold Standard and non gnomAD-non neuro non-Finnish Europeans, v2.1), 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 or duplicated if all associated capture targets were affected in the respective capture kit. We then built a copy number matrix for each autosomal protein-coding transcript, with 5 possible states per sample and transcript post-filtration: (i) missing information, (ii) no CNV (copy number=2), (iii) complete or (iv) partial duplication, or (v) complete or partial deletion. We worked under the assumption that both complete and partial deletions likely result in haploinsufficiency. In dosage analyses, we converted states (ii), (iii) and (v) into a transcript copy number information.

Then, to avoid the analysis of rare CNVs affecting transcripts frequently deleted or duplicated, we built two sets of transcripts: (i) transcripts with <1% cumulative deletion frequency in public databases (set A, non-frequently deleted transcripts) and (ii) 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 full duplication (affecting coding transcripts, restricted to set A for the deletion analysis and to a set B for the duplication analysis). Then, we performed similar burden analysis, now focusing on two gene lists: (i) genes associated with AD in GWAS[2] and (ii) genes having function in Aβ network[27]. Then, we focused on AD-associated genes by performing gene-based analyses of loss-of-function (LOF) variants including both deletion-CNVs and truncating SNVs/indels in genes previously associated with AD. Burden CNV analyses consisted in (i) EOAD-control association analyses using Firth's logistic regression and (ii) ordinal regression using the whole dataset (EOAD, LOAD, controls),

assuming an EOAD<LOAD<control effect (or conversely), as previously performed for burdens of rare SNV/indels at the gene level[3].

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 cases to controls, and assessed gene dosage using complete/partial deletions (decreasing dosage) and complete duplications (increasing dosage), on the fusion of transcript sets A and B. Signals with FDR<10% were considered suggestive and were further assessed by ordinal regression using the whole dataset (EOAD, LOAD, controls).

Complementary analyses consisted in: (i) gene-level, transcript-based analyses on deletions only (all protein coding genes from set A), complete duplications only (all protein coding genes set B) and (ii) gene-based analysis of loss-of-function (LOF) variants (deletion-CNVs plus truncating SNVs/indels) in genes prioritized by the dosage analysis (integrated LOF analysis). In addition, we further performed adjusted analyses based on ancestry (Firth's logistic model including the first ten principal components) and on *APOE4* allele dosage, as sensitivity analyses.

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 (UKBB), deCODE, GR@ACE and exomes of African Americans from ADSP. Detailed sample selection, calling and QC procedures are reported in the supplementary information. Due to heterogeneous numbers of EOAD cases 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. Then, discovery and replication datasets were 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 1,717 gene-level tests fulfilling the above-mentioned 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 both significance thresholds.

In vitro cellular assays

To investigate *SCARF2* involvement in Aβ uptake, we overexpressed the *SCARF2* cDNA in HEK293 cells and used, as controls, the *TARDP* cDNA (TDP-43 protein), which is not known to influence Aβ uptake or relate to Aβ-associated pathways, and a Mock control. Cells were then incubated with a pH-sensitive fluorescent Aβ conjugate ($A\beta^{pH}$, Protonex GreenTM). Aβ internalization was quantified by flow cytometry through mean fluorescence intensity (MFI) per cell. Transfection efficiency was assessed by FLAG immunolabelling followed by flow cytometry analysis.

Acknowledgements

We thank the ADSP study for providing exome data from cases and controls (ADSP umbrella NG00067.v3). ADSP data were prepared, archived, and distributed by the National Institute on Aging Alzheimer's 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's 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.

We thank the high-performance computing service at the University of Lille.

This research has been conducted using the UK Biobank Resource under application number 63653 (to G.Nicolas) and 40980 (to S. Jacquemont).

Ace Alzheimer Center Barcelona acknowledges the support Fundación bancaria "La Caixa", Fundación ADEY, Fundación Echevarne and Grífols SA (GR@ACE project). Also, the Spanish National R+D+I and Instituto de Salud Carlos III (ISCIII) for grants PI19/01301, PI19/01240, P22/I01403 within the European Plan *FEDER "Una manera de hacer Europa"*, and grants PMP22/00022 within the plan *Proyectos de Investigación de Medicina Personalizada de Precisión*, within the European plan Next Generation EU. The support Joint Program for Neurodegenerative Diseases (JPND) towards PREADAPT project (grant NºAC19/00097) and AD-PriOMICS project (AC23/00038), and from German Federal Ministry of Education and Research (BMBF) for the DESCARTES project (BMBF grant Nº01EK2102B and 01EK2102A), German Research Foundation (DFG). IdR was supported by the ISCIII under the grant FI20/00215. Ace Alzheimer Center Barcelona research receives support from Roche, Janssen, Life Molecular Imaging, Araclon Biotech, Alkahest, Laboratorio de Análisis Echevarne, and IrsiCaixa. Some control samples and data from patients included in this study were provided in part by the National DNA Bank Carlos III (www.bancoadn.org, University of Salamanca, Spain) and Hospital Universitario Virgen de Valme (Sevilla, Spain).

Part of the work was carried out on the Cartesius supercomputer, which is embedded in the Dutch national e-infrastructure with the support of SURF Cooperative. Computing hours were

granted to Henne Holstege by the Dutch Research Council ('100plus': project# vuh15226, 15318, 17232, 2020.030; 'Role of VNTRs in AD'; project# 2022.028, 'Alzheimer's Genetics Hub' project# 2022.031 and 2024:036).

Funding

CS was supported by the Fondation pour la Recherche Medicale (ARF201909009263) and the Rouen metropole through the "post-doctorants métropole" funding.

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; ADSP. Full consortium acknowledgements and funding sources are listed in the supplementary information document.

S.L. was funded by ZonMw (PROMO-GENODE: a PROspective study of MOnoGEnic causes Of Dementia #733050512 and The YOD-INCLUDED project #10510032120002 and is part of the Dutch Dementia Research Programme) and received a substantial donation by Edwin Bouw Fonds, Dioraphte and Stichting TeamAlzheimer. S.L. further received funding for the GeneMINDS consortium, which is powered by Health~Holland, Top Sector Life Sciences & Health. GeneMINDS receives co-financing from Brain Research Center, Prevail Therapeutics and Vigil Neurosciences. S.L. is recipiens of ABOARD, which is a public-private partnership receiving funding from ZonMW (#73305095007) and Health~Holland, Topsector Life Sciences & Health (PPP-allowance; #LSHM20106). More than 30 partners participate in ABOARD. ABOARD also receives funding from de Hersenstichting, Edwin Bouw Fonds and Gieskes-Strijbisfonds.

Competing interests

We declare no competing interests related to the study

Supplementary material

Contains 16 supplementary tables, 18 supplementary figures, and a supplementary information document

Data Availability

Summary statistics are available upon request. CNV counts (deletions and full duplications) in all genes are available as a supplement. Access to deidentified raw data of the ADES study can be requested using the Alzheimer Genetics Hub. 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/

REFERENCES

1. Nicolas G. Lessons from genetic studies in Alzheimer disease. *Revue neurologique* 2024;180(5):368-77.
2. Bellenguez C, Kucukali F, Jansen IE, Kleineidam L, Moreno-Grau S, Amin N, et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nature genetics* 2022;54(4):412-36.
3. Holstege H, Hulsman M, Charbonnier C, Grenier-Boley B, Quenez O, Grozeva D, et al. Exome sequencing identifies rare damaging variants in ATP8B4 and ABCA1 as risk factors for Alzheimer's disease. *Nature genetics* 2022;54(12):1786-94.
4. Campion D, Charbonnier C, Nicolas G. SORL1 genetic variants and Alzheimer disease risk: a literature review and meta-analysis of sequencing data. *Acta neuropathologica* 2019;138(2):173-86.
5. Coe BP, Witherspoon K, Rosenfeld JA, van Bon BW, Vulto-van Silfhout AT, Bosco P, et al. Refining analyses of copy number variation identifies specific genes associated with developmental delay. *Nature genetics* 2014;46(10):1063-71.
6. Niestroj LM, Perez-Palma E, Howrigan DP, Zhou Y, Cheng F, Saarentaus E, et al. Epilepsy subtype-specific copy number burden observed in a genome-wide study of 17 458 subjects. *Brain : a journal of neurology* 2020;143(7):2106-18.
7. Auwerx C, Joeloo M, Sadler MC, Tesio N, Ojavee S, Clark CJ, et al. Rare copy-number variants as modulators of common disease susceptibility. *Genome medicine* 2024;16(1):5.
8. Rovelet-Lecrux A, Hannequin D, Raux G, Le Meur N, Laquerriere A, Vital A, et al. APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. *Nature genetics* 2006;38(1):24-6.
9. Cabrejo L, Guyant-Marechal L, Laquerriere A, Vercelletto M, De la Fourniere F, Thomas-Anterion C, et al. Phenotype associated with APP duplication in five families. *Brain : a journal of neurology* 2006;129(Pt 11):2966-76.
10. Lanoiselee HM, Nicolas G, Wallon D, Rovelet-Lecrux A, Lacour M, Rousseau S, et al. APP, PSEN1, and PSEN2 mutations in early-onset Alzheimer disease: A genetic screening study of familial and sporadic cases. *PLoS medicine* 2017;14(3):e1002270.
11. Fortea J, Vilaplana E, Carmona-Iragui M, Benejam B, Videla L, Barroeta I, et al. Clinical and biomarker changes of Alzheimer's disease in adults with Down syndrome: a cross-sectional study. *Lancet* 2020;395(10242):1988-97.
12. Le Guennec K, Veugelen S, Quenez O, Szaruga M, Rousseau S, Nicolas G, et al. Deletion of exons 9 and 10 of the Presenilin 1 gene in a patient with Early-onset Alzheimer Disease generates longer amyloid seeds. *Neurobiology of disease* 2017;104:97-103.
13. Brouwers N, Van Cauwenberghe C, Engelborghs S, Lambert JC, Bettens K, Le Bastard N, et al. Alzheimer risk associated with a copy number variation in the complement receptor 1 increasing C3b/C4b binding sites. *Molecular psychiatry* 2012;17(2):223-33.
14. Wang H, Wang LS, Schellenberg G, Lee WP. The role of structural variations in Alzheimer's disease and other neurodegenerative diseases. *Frontiers in aging neuroscience* 2022;14:1073905.
15. Hooli BV, Kovacs-Vajna ZM, Mullin K, Blumenthal MA, Mattheisen M, Zhang C, et al. Rare autosomal copy number variations in early-onset familial Alzheimer's disease. *Molecular psychiatry* 2014;19(6):676-81.
16. Rovelet-Lecrux A, Legallic S, Wallon D, Flaman JM, Martinaud O, Bombois S, et al. A genome-wide study reveals rare CNVs exclusive to extreme phenotypes of Alzheimer disease. *European journal of human genetics : EJHG* 2012;20(6):613-7.
17. Collins RL, Brand H, Karczewski KJ, Zhao X, Alfoldi J, Francioli LC, et al. A structural variation reference for medical and population genetics. *Nature* 2020;581(7809):444-51.

18. Lee WP, Tucci AA, Conery M, Leung YY, Kuzma AB, Valladares O, et al. Copy Number Variation Identification on 3,800 Alzheimer's Disease Whole Genome Sequencing Data from the Alzheimer's Disease Sequencing Project. *Frontiers in genetics* 2021;12:752390.
19. Wang H, Dombroski BA, Cheng PL, Tucci A, Si YQ, Farrell JJ, et al. Structural variation detection and association analysis of whole-genome-sequence data from 16,543 Alzheimer's disease sequencing project subjects. *Alzheimer's & dementia : the journal of the Alzheimer's Association* 2025;21(6):e70277.
20. Gabrielaite M, Torp MH, Rasmussen MS, Andreu-Sanchez S, Vieira FG, Pedersen CB, et al. A Comparison of Tools for Copy-Number Variation Detection in Germline Whole Exome and Whole Genome Sequencing Data. *Cancers* 2021;13(24).
21. Hehir-Kwa JY, Pfundt R, Veltman JA. Exome sequencing and whole genome sequencing for the detection of copy number variation. *Expert review of molecular diagnostics* 2015;15(8):1023-32.
22. Backenroth D, Homsy J, Murillo LR, Glessner J, Lin E, Brueckner M, et al. CANOES: detecting rare copy number variants from whole exome sequencing data. *Nucleic acids research* 2014;42(12):e97.
23. Quenez O, Cassinari K, Coutant S, Lecoquierre F, Le Guennec K, Rousseau S, et al. Detection of copy-number variations from NGS data using read depth information: a diagnostic performance evaluation. *European journal of human genetics : EJHG* 2021;29(1):99-109.
24. Le Guennec K, Quenez O, Nicolas G, Wallon D, Rousseau S, Richard AC, et al. 17q21.31 duplication causes prominent tau-related dementia with increased MAPT expression. *Molecular psychiatry* 2017;22(8):1119-25.
25. Rovelet-Lecrux A, Charbonnier C, Wallon D, Nicolas G, Seaman MN, Pottier C, et al. De novo deleterious genetic variations target a biological network centered on Abeta peptide in early-onset Alzheimer disease. *Molecular psychiatry* 2015;20(9):1046-56.
26. Wallon D, Boluda S, Rovelet-Lecrux A, Thierry M, Lagarde J, Miguel L, et al. Clinical and neuropathological diversity of tauopathy in MAPT duplication carriers. *Acta neuropathologica* 2021;142(2):259-78.
27. Campion D, Pottier C, Nicolas G, Le Guennec K, Rovelet-Lecrux A. Alzheimer disease: modeling an Abeta-centered biological network. *Molecular psychiatry* 2016;21(7):861-71.
28. Steinberg S, Stefansson H, Jonsson T, Johannsdottir H, Ingason A, Helgason H, et al. Loss-of-function variants in ABCA7 confer risk of Alzheimer's disease. *Nature genetics* 2015;47(5):445-7.
29. Wilkinson K, El Khoury J. Microglial scavenger receptors and their roles in the pathogenesis of Alzheimer's disease. *International journal of Alzheimer's disease* 2012;2012:489456.
30. Singh TD, Park SY, Bae JS, Yun Y, Bae YC, Park RW, et al. MEGF10 functions as a receptor for the uptake of amyloid-beta. *FEBS letters* 2010;584(18):3936-42.
31. Lee JH, Cheng R, Graff-Radford N, Foroud T, Mayeux R, National Institute on Aging Late-Onset Alzheimer's Disease Family Study G. Analyses of the National Institute on Aging Late-Onset Alzheimer's Disease Family Study: implication of additional loci. *Archives of neurology* 2008;65(11):1518-26.
32. Belloy ME, Andrews SJ, Le Guen Y, Cuccaro M, Farrer LA, Napolioni V, et al. APOE Genotype and Alzheimer Disease Risk Across Age, Sex, and Population Ancestry. *JAMA neurology* 2023.
33. Sims R, van der Lee SJ, Naj AC, Bellenguez C, Badarinarayan N, Jakobsdottir J, et al. Rare coding variants in PLCG2, ABI3, and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease. *Nature genetics* 2017;49(9):1373-84.
34. Brutman JN, Busald T, Nizamis E, Kaufman EJ, Lin E, Hendricks NE, et al. A common 19 bp APOE enhancer deletion is protective against Alzheimer's disease in African Americans. *Nature communications* 2026;17(1).

35. Nicolas G, Zarea A, Lacour M, Quenez O, Rousseau S, Richard AC, et al. Assessment of Mendelian and risk-factor genes in Alzheimer disease: A prospective nationwide clinical utility study and recommendations for genetic screening. *Genetics in medicine : official journal of the American College of Medical Genetics* 2024;26(5):101082.
36. Alver M, Mancini V, Lall K, Schneider M, Romano L, Estonian Biobank Research T, et al. Contribution of schizophrenia polygenic burden to longitudinal phenotypic variance in 22q11.2 deletion syndrome. *Molecular psychiatry* 2022;27(10):4191-200.
37. Rump P, de Leeuw N, van Essen AJ, Verschuuren-Bemelmans CC, Veenstra-Knol HE, Swinkels ME, et al. Central 22q11.2 deletions. *American journal of medical genetics Part A* 2014;164A(11):2707-23.
38. Huguet G, Schramm C, Douard E, Jiang L, Labbe A, Tihy F, et al. Measuring and Estimating the Effect Sizes of Copy Number Variants on General Intelligence in Community-Based Samples. *JAMA psychiatry* 2018;75(5):447-57.
39. Stern Y. Cognitive reserve in ageing and Alzheimer's disease. *The Lancet Neurology* 2012;11(11):1006-12.
40. Nelson ME, Jester DJ, Petkus AJ, Andel R. Cognitive Reserve, Alzheimer's Neuropathology, and Risk of Dementia: A Systematic Review and Meta-Analysis. *Neuropsychology review* 2021;31(2):233-50.
41. Butcher NJ, Kiehl TR, Hazrati LN, Chow EW, Rogaeva E, Lang AE, et al. 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 neurology* 2013;70(11):1359-66.
42. Boot E, Oskarsdottir S, Loo JCY, Crowley TB, Orchanian-Cheff A, Andrade DM, et al. Updated clinical practice recommendations for managing adults with 22q11.2 deletion syndrome. *Genetics in medicine : official journal of the American College of Medical Genetics* 2023;25(3):100344.
43. Grangeon L, Cassinari K, Rousseau S, Croisile B, Formaglio M, Moreaud O, et al. Early-Onset Cerebral Amyloid Angiopathy and Alzheimer Disease Related to an APP Locus Triplication. *Neurology Genetics* 2021;7(5):e609.
44. Sleegers K, Brouwers N, Gijselinck I, Theuns J, Goossens D, Wauters J, et al. APP duplication is sufficient to cause early onset Alzheimer's dementia with cerebral amyloid angiopathy. *Brain : a journal of neurology* 2006;129(Pt 11):2977-83.
45. Grangeon L, Charbonnier C, Zarea A, Rousseau S, Rovelet-Lecrux A, Bendetowicz D, et al. Phenotype and imaging features associated with APP duplications. *Alzheimer's research & therapy* 2023;15(1):93.
46. Ehn E, Eisfeldt J, Laffita-Mesa JM, Thonberg H, Schoumans J, Portaankorva AM, et al. A de novo, mosaic and complex chromosome 21 rearrangement causes APP triplication and familial autosomal dominant early onset Alzheimer disease. *Scientific reports* 2025;15(1):2912.
47. Mueller-Steiner S, Zhou Y, Arai H, Roberson ED, Sun B, Chen J, et al. Antiamyloidogenic and neuroprotective functions of cathepsin B: implications for Alzheimer's disease. *Neuron* 2006;51(6):703-14.
48. Logsdon GA, Vollger MR, Eichler EE. Long-read human genome sequencing and its applications. *Nature reviews Genetics* 2020;21(10):597-614.
49. Miller DE, Sulovari A, Wang T, Loucks H, Hoekzema K, Munson KM, et al. Targeted long-read sequencing identifies missing disease-causing variation. *American journal of human genetics* 2021;108(8):1436-49.
50. Le Borgne J, Gomez L, Heikkinen S, Amin N, Ahmad S, Choi SH, et al. X-chromosome-wide association study for Alzheimer's disease. *Molecular psychiatry* 2025;30(6):2335-46.
51. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* 1984;34(7):939-44.

52. McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Jr., Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & dementia : the journal of the Alzheimer's Association* 2011;7(3):263-9.
53. Teo SM, Pawitan Y, Ku CS, Chia KS, Salim A. Statistical challenges associated with detecting copy number variations with next-generation sequencing. *Bioinformatics* 2012;28(21):2711-8.
54. Zhao M, Wang Q, Wang Q, Jia P, Zhao Z. Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. *BMC bioinformatics* 2013;14 Suppl 11(Suppl 11):S1.

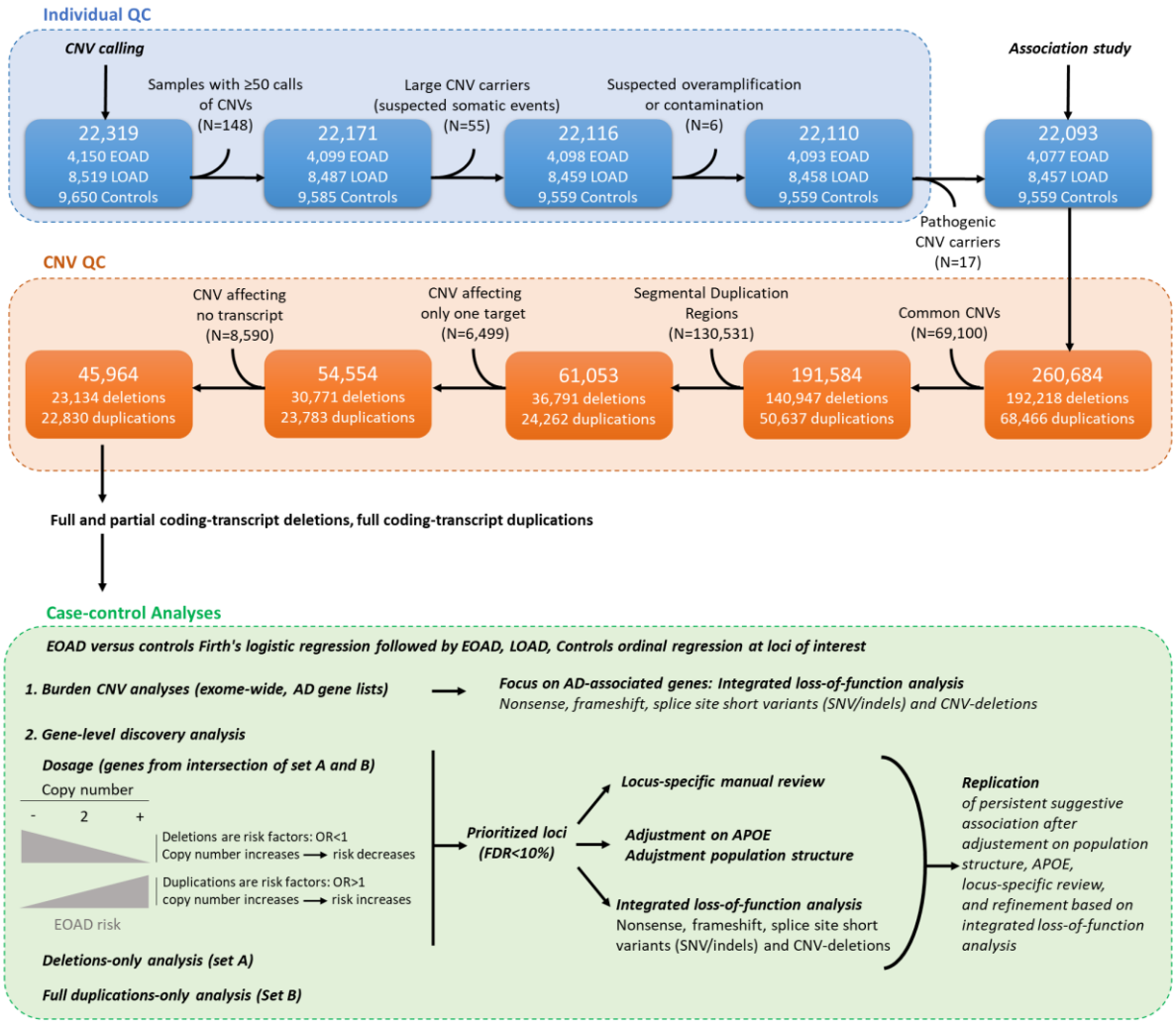


Figure 1: Summary of the study

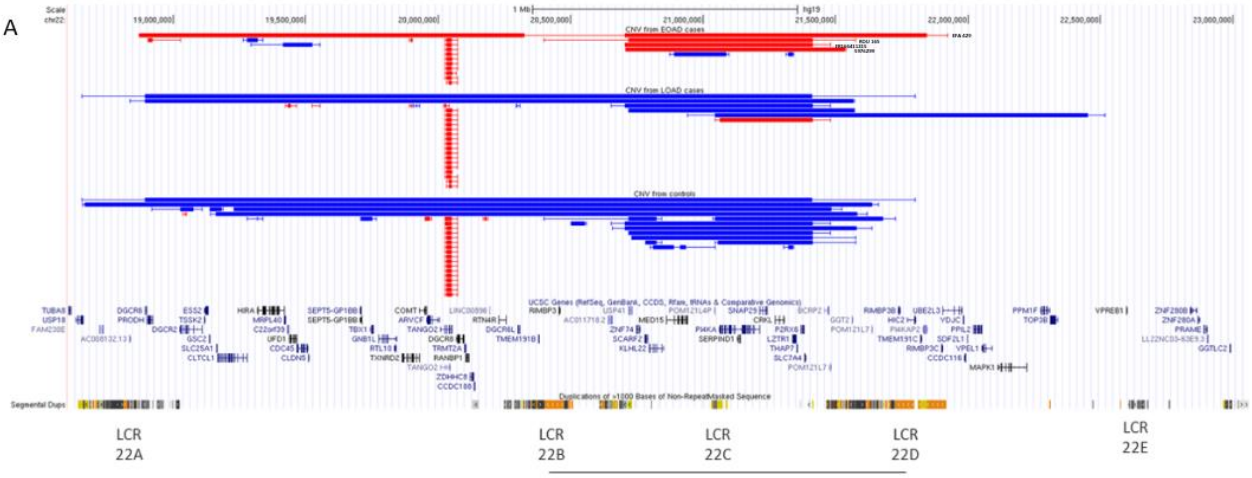
This study included exome sequencing samples from the ADES-ADSP described in Holstege et al. 2022, restricted to individuals passing their quality control and being included in a CNV-calling-batch of 50+ individuals.

QC: Quality Control; CNV: Copy Number Variation; EOAD: Early Onset Alzheimer Disease; LOAD: Late Onset Alzheimer Disease; OR: Odd-ratio

Set A: non frequently deleted genes

Set B: non frequently duplicated genes

A



B

Patient ID	EFA-429-001	ROU-165-001	FR16641215	FR16642636
Age at sample (y)	[46-50]	[51-55]	[60-65]	[51-55]
Sex	male	male	female	male
APOE genotype	3-4	3-4	3-3	3-4
Age at onset (y)	[41-45]	[51-55]	[51-55]	[46-50]
Educationnal level (y)	12	9	NA	5
First reported symptoms	memory and executive functions	memory	NA	memory, writing
Medical history of the 22q11 deletion spectrum	Congenital heart defect, no history of neurodeveloppemental disorders	no history of neurodeveloppemental disorders	NA	no history of neurodeveloppemental disorders
Family history of neurodegenerative disease	Parkinson disease in mother at 62 related to *LRRK2* pathogenic variant	negative	Dementia in father with onset around 60 and death at 69	negative
Disease duration (y)	6 (to death)	8	10 (to death)	9
Additionnal arguments supporting AD pathophysiology	- MMSE = 12/30 at 46 y. - 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 - Severe episodic memory disorder with misorientation and logopenic aphasia - bilateral mediotemporal and parietal corrtical 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 - 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 singular cortex on PIB PET

Figure 2. Focus on the 22q11.21 locus

A. UCSC genome browser view of CNVs identified in the main analysis In red: deletions; In blue: duplications. Large areas represent coordinates of CNVs as detected by CANOES. Thinner areas show breakpoints uncertainties, i.e. regions between two targets of the sample-specific capture kit. Genes are indicated below, and the regions of segmental duplications appear at the bottom part. Low copy repeats (LCR) 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.

B. Clinical information of carriers of the deletion in the main analysis Y: years; AD: Alzheimer Disease; EOAD: Early Onset AD; LOAD: Late Onset AD; MMSE: Mini Mental State Examination; CSF: Cerebrospinal Fluid

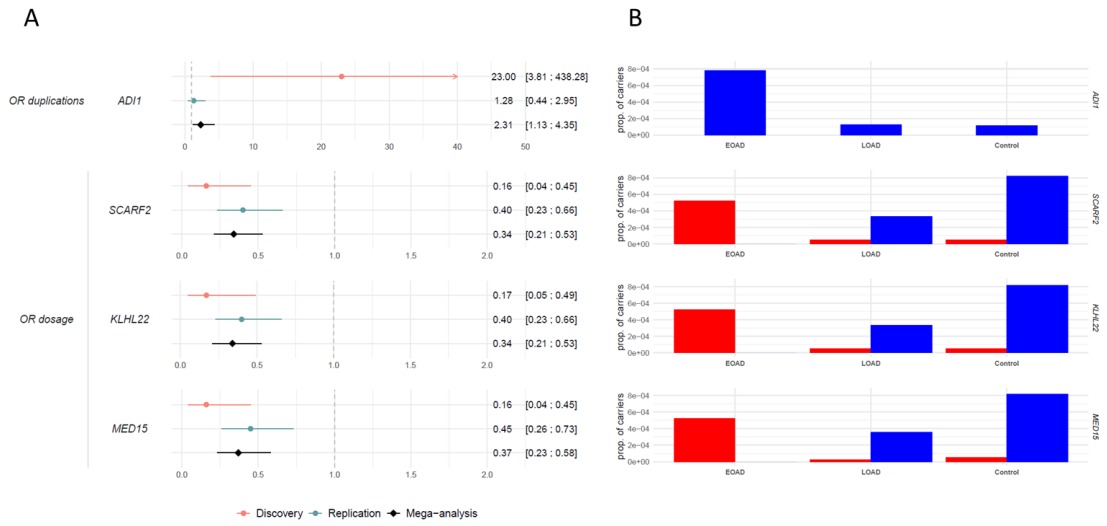


Figure 3. Mega-analysis of results from both main analysis and replication

A. Forest plot associated with the mega-analysis of results from both discovery and replication.

Figure displays OR with 95% Confidence Intervals. For dosage analyses (*SCARF2*, *KLHL22*, *MED15*), OR should be interpreted for an increase of transcript copies by 1. i.e. a OR > 1 (resp. OR <1) means that duplications (resp. deletions) are associated with a higher disease risk (see Figure 1). OR: Odd-ratio

B. Proportion of deletions and duplications carriers by status and gene
In red: deletions; In blue: complete duplications. Data include both discovery and replication datasets.

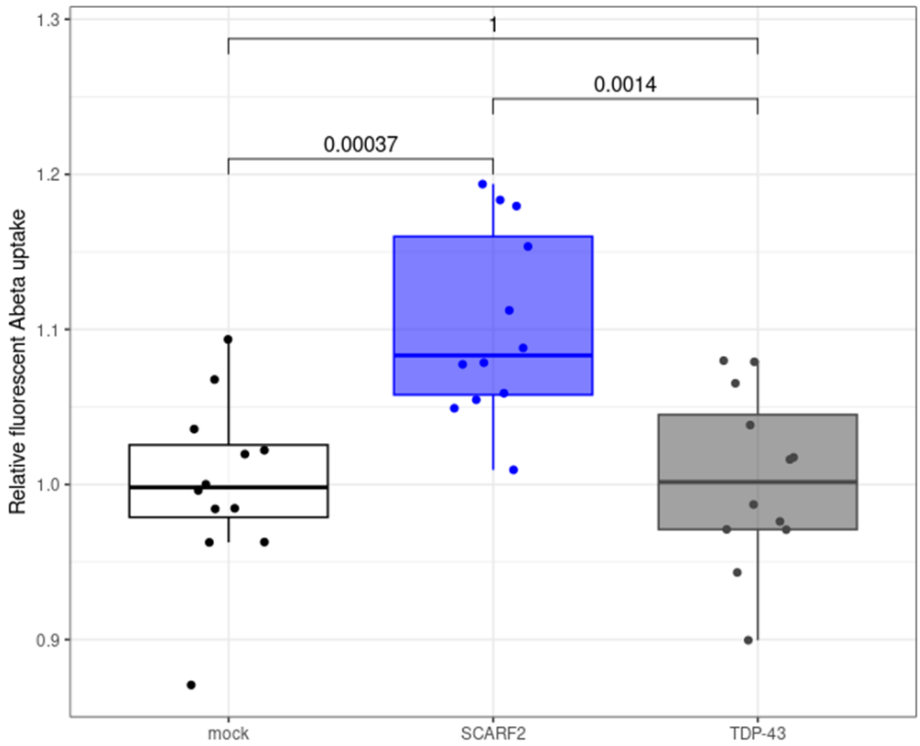


Figure 4. Aβ$_{42}$ uptake in HEK293 cells.

HEK293 cells transfected with an empty vector (mock) or overexpressing the SCARF2 or the TDP43 proteins were incubated with fluorescent Aβ$_{42}$ peptides, and Aβ$_{42}$ uptake was assessed by flow cytometry. SCARF2 overexpression was associated with a significantly increased uptake of Aβ$_{42}$ peptides compared to the mock condition (p=0.00037), while TDP-43 overexpression did not affect Aβ$_{42}$ uptake (p=1).

	EOAD	LOAD	All AD	Controls
N	4,077	8,457	12,534	9,559
Age mean (SD)[1]	58.43 y (5.09)	77.90 y (7.48)	71.38 y (11.42)	81.36 y (12.28)
Gender N (%)				
males	1814 (44.5%)	3307 (39.1%)	5121 (40.9%)	4147 (43.4%)
females	2263 (55.5%)	5150 (60.9%)	7413 (59.1%)	5412 (56.6%)
***APOE* N (%)[2]**				
E4-non carriers	1704 (41.9%)	4640 (54.9%)	6344 (50.7%)	7535 (79.6%)
E4-heterozygous	1699 (41.7%)	3417 (40.5%)	5116 (40.9%)	1830 (19.3%)
E4-homozygous	668 (16.4%)	390 (4.6%)	1058 (8.4%)	101 (1.1%)
N. 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)
N. 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)

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

N: number, y: years, SD: standard deviation, bp: base pairs, EOAD: early onset Alzheimer disease, LOAD: late onset Alzheimer disease

[1]AAO for cases and age at last exam/inclusion for controls. Age is missing for 373 LOAD and 256 controls

[2]APOE genotype is missing for 6 EOAD, 10 LOAD and 93 controls

Sample	Cohort	AAO	Gender	CNV coordinate	HGVS nomenclature	Gene	Event
ALZ-0596-001[1]	ADES-FR	49-50	Male	chr17:43,679,169-44,145,099	NC_000017.10:g.43585921_44159772dup	MAPT	complete duplication
ALZ-0441-005[1]	ADES-FR	41-45	Male	chr17:43,717,946-44,364,336	NC_000017.10:g.43679455_44363695dup	MAPT	complete duplication
EXT-1114-001[1]	ADES-FR	51-55	Female	chr17:43,679,164-44,364,341	NC_000017.10:g.43585926_44408484dup	MAPT	complete duplication
ROU-1373-001[1]	ADES-FR	51-55	Male	chr17:43,717,946-44,159,945	NC_000017.10:g.43679455_44172015dup	MAPT	complete duplication
SC45A00007a	**PERADES**	**56-60**	**Female**	**chr17:43,679,204-44,159,910**	**NC_000017.10:g.43611648_44171922dup**	**MAPT**	**complete duplication**
C-FHS-51259	**ADSP**	**86-90**	**Female**	**chr17:43,898,720-44,101,538**	**NC_000017.10:g.43893949_44108840dup**	**MAPT**	**complete duplication**
EXT-0814-001[3]	ADES-FR	51-55	Female	chr21:22,370,838-28,338,704	NC_000021.8:g.20058403_30248689dup	APP	complete duplication
EXT-0773-001[3 4]	ADES-FR	41-45	Male	chr21:26,946,242-31,066,399	NC_000021.8:g.22910384_31233900dup	APP	complete duplication
EXT-0857-001[3]	ADES-FR	56-60	Male	chr21:26,946,246-28,338,704	NC_000021.8:g.22910379_30248689dup	APP	complete duplication
FR16640974	**Netherlands Brain Bank**	**46-50**	**Male**	**chr21:22,370,838-28,338,827**	**NC_000021.8:g.20230358_30248689dup**	**APP**	**complete duplication**
TE17A00020a_DS1	**PERADES**	**41-45**	**Male**	**chr21:27,056,123-28,338,719**	**NC_000021.8:g.27012218_30248703dup**	**APP**	**complete duplication**
20732	**PERADES**	**51-55**	**Female**	**chr21:27,062,175-27,484,514**	**NC_000021.8:g.27056277_27512460dup**	**APP**	**complete duplication**
CH32A00027a	**PERADES**	**51-55**	**Female**	**chr21:27,066,065-28,338,719**	**NC_000021.8:g.27062287-30248703dup**	**APP**	**complete duplication**
09D04764	**ADC-Amsterdam**	**56-50**	**Male**	**chr21:26,541,838-29,645,932**	**NC_000021.8:g25725527_30248655dup**	**APP**	**complete duplication**
10D05060	**ADC-Amsterdam**	**41-45**	**Female**	**chr21:27,253,167-27,542,996**	**NC_000021.8:g.27142399_27840733dup**	**APP**	**complete duplication**
SID14617	**AgeCoDe-UKBonn**	**46-50**	**Female**	**chr21:26,946,221-27,542,992**	**NC_000021.8:g.25725527_27840779dup**	**APP**	**complete duplication**
EXT-0313-001[2]	ADES-FR	56-60	Male	chr14:73,672,981-73,678,665	NC_000014.8:g.73664857_73683810del	PSEN1	deletion exon 9-10

Table 2. Mendelian pathogenic CNVs

AAO: Age at onset (in years), CNV: copy number variant, HGVS: Human Genome Variation Society; in bold: novel

[1] reported in Le Guennec et al., Mol Psychiatry, 2017

[2] reported in Le Guennec et al., Neurobiol Dis, 2017

[3] reported in Lanoiselée et al., PLoS Med, 2017

[4] reported in Rovelet-Lecrux et al., Mol Psychiatry, 2015

Chr	Gene	N total			N carriers of CNVs (DEL/DUP)			EOAD vs Controls - Dosage analysis			EOAD vs Controls - Deletion analysis		EOAD vs Controls - Duplication analysis		Transcripts***
		EOAD	LOAD	Controls	EOAD	LOAD	Controls	OR [95%CI][1]	p-value[2]	FDR	OR [95%CI]	p-value	OR [95%CI]	p-value	
1	*PRAMEF26**	2911	7302	6461	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	NM_001306072.3
2	*ADI1*	4077	8457	9559	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	NM_001306077.2 ; NM_018269.4
8	*MTUS1***	4077	8457	9559	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		NM_001001924.3; NM_001363057.2; NM_001363058.2; NM_001363059.2; NM_001363061.2
10	*MBL2*	4077	8457	9559	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	NM_000242.3; NM_001378373.1; NM_001378374.1
17	*FADS6*	4077	8457	9559	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	NM_178128.6
19	*PLIN4***	4077	8457	9559	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	NM_001393888.1 ; NM_001393889.1 ; NM_001393890.1 ; NM_001393891.1; NM_001367868.2
19	*KLC3*	4077	8457	9559	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	NM_177417.3
19	*ERCC2*	4077	8457	9559	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	NM_001130867.2
		4077	8457	9559	29 (1 / 28)	42 (2 / 40)	20 (0 / 20)								NM_000400.4
22	*ZNF74*	4077	8457	9559	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	NM_001256523.2; NM_001256525.2; NM_003426.4; NM_001256524.2
22	*SCARF2*	4077	8457	9559	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	NM_153334.7; NM_182895.5
22	*KLHL22*	4077	8457	9559	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	NM_032775.4
22	*MED15*	4077	8457	9559	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	NM_001003891.3; NM_001293234.2; NM_001293235.2; NM_001293236.2; NM_001293237.2; NM_015889.5
22	*PI4KA*	4077	8457	9559	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	NM_001362862.2; NM_001362863.2; NM_058004.4
22	*SERPIND1*	4077	8457	9559	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	NM_000185.4
22	*SNAP29*	4077	8457	9559	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	NM_004782.4
22	*CRKL*	4077	8457	9559	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	NM_005207.4
22	*LZTR1*	4077	8457	9559	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	NM_006767.4
22	*THAP7*	4077	8457	9559	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	NM_030573.3; NM_001008695.1

22	*P2RX6*	4077	8457	9559	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	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
22	*SLC7A4*	4077	8457	9559	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	NM_004173.3
22	*LRRC74B*	3917	7799	8016	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	NM_001291006.2

Table 3. Results from EOAD versus Controls dosage analysis and subsequent analyses on deletions and duplications

Chr: chromosome; N: number. EOAD: early onset Alzheimer disease. LOAD: late onset Alzheimer disease

[1]OR should be interpreted for an increase of transcript copies by 1. i.e. a OR > 1 (resp. OR <1) means that duplications (resp. deletions) are associated with a higher disease risk

[2]non adjusted p-values. selected based on FRD < 10%

*gene overlapping with repeats or duplicated gene in the genome (despite CNVs not overlapping >50% with repeats)

**gene for which the signal in dosage analysis is driven by deletion whereas transcripts are not in set A

*** 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 marked by a colored block on the left